Application of big data in COVID-19 epidemic

Joseph Bamidele Awotunde[1], Emmanuel Abidemi Adeniyi[2], Paul Oluwatoba Kolawole[1], Roseline Oluwaseun Ogundokun[2]

[1]DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF ILORIN, ILORIN, KWARA, NIGERIA; [2]DEPARTMENT OF COMPUTER SCIENCE, LANDMARK UNIVERSITY, OMU ARAN, KWARA, NIGERIA

1. Introduction

Data science can simply be referred to as a method of formatting data that comprises an emerging approach of acquiring, storing, and analyzing data to produce a meaningful and an efficient way of extracting basic information for decision making [1,2]. Data science cuts across almost all the fields that use scientific approaches, procedure systems, and computer algorithms to obtain knowledgeable information and perceptions from both structured and unstructured data [3—5]. Thus the main goal remains to be the improvement of understanding from various data types [6,7].

The field of data science is associated with the computer science field, but a keen observation of the two fields reveals that they differ slightly with some characteristics. The field of computer science deals with the development of computer programs and procedures for data processing and management, while data science applies to any form of a manual or computerized method of analyzing data. Data science involves gathering, investigating, and producing of data for analysis that is thoroughly associated with the field of mathematics and statistics [8—10].

The advancement in information technology increases the amount of data used by the modern establishment, thereby making data science an essential tool for maintaining data in any organization [10—12]. For instance, an organization with a huge amount of user data needs data science to effectively improve methods to collect, accumulate, and process the data. Various systematic techniques may be used by the company to process and obtain resourceful outcomes about the user data. Data mining and big data are used interrelated with data science. To comprehend and process actual data concepts, data science is of great importance, as it incorporates the field of statistics and machine learning with their various procedures [13,14]. The study of information

science, computer science, mathematics, and statistics with their concepts and methods are all embedded in the field of data science [14,15].

The 21st century is an age of big data affecting every aspect of human life, including biology and medicine [16,17]. The move from paper medical records to electronic health record (EHR) systems has resulted in an unprecedented increase in data [18,19]. Big data thus offers a great opportunity for doctors, epidemiologists, and specialists in health policy to make evidence-driven decisions that will eventually enhance patient care [20,21]. "Big data is not only a modern reality for the biomedical scientist but an imperative that needs to be fully grasped and used in the search for new knowledge" [22−24].

The coronavirus pandemic is a global health issue that threatens the peace of the entire world since the New Year 2020 [25]. The number of confirmed cases of people who contracted the disease increases daily, but the death rate recorded is low compared with the recovery cases, which makes the disease less deadly but more infectious (2,367,959 cases in 213 countries, with over 160,450 deaths as of April 20, 2020) (https://covid19. who.int/). The daily increase of this infectious disease has become a huge challenge to the socioeconomic development of each country across the globe. The Secretary-General of the United Nation has advised that urgent measures should be taken by governments in various countries to contain the outspread of the infectious disease [26−30]. One of the goals of the United Nations Sustainable Development is to discuss economic, social, and environmental concerns between 2015 and 2030 and change toward maintainable growth [31−33]. The UN Sustainable Development Goals (SDGs) include 17 aims and 169 objectives. Three (3) SDG focuses on maintaining lives and encouraging well-being for all ages, including the SDG 3.3 specifically targeting the end of AIDS, tuberculosis, malaria, and abandoned tropical illnesses, as well as the fight against hepatitis, water-borne diseases, and other transmittable infections, by 2030 [31,32]. The coronavirus disease 2019 (COVID-19) epidemic directly challenges the accomplishment of the abovementioned health objectives and also affects the achievement of economic and social development goals.

The transmission features of the COVID-19 epidemic have still not been properly recognized in the sense of global environmental changes [32,34]. Furthermore, the pace of global development, increased population size, more common and nuanced encounters, and lack of medical security in developing countries all contribute to the complexities of COVID-19 prevention and regulation. COVID-19 seems to have a greater impact on the global economy when compared with the severe acute respiratory syndrome (SARS) that occurred in 2013.

The bone of contention is how to fight the spread of infectious diseases and how this one is different from the 2003 epidemic. Most countries depended on the observation of classic steps to monitor diseases and public health to contain the COVID-19 pandemic, similar to those used with SARS in 2003 [35]. They cover from drastic quarantine actions in China (for example, locking down over 60 million people in Hubei province) to extensive contact monitoring with hundreds of contact tracers (for example, Singapore,

Hong Kong, South Korea). Nevertheless, these interventions may not be successful in tackling the COVID-19 scale by 2020. From Table 8.1 and Fig. 8.1, as of May 17, 2020, a total of 4,798,545 confirmed cases and 316,505 deaths were reported across 218 countries, thus showing that there is a need for a better way of prevention of this novel pandemic.

The 12 countries with the highest number of confirmed cases are shown in Table 8.2; the United States of America has 1,409,452 (29%) of the total confirmed cases around the globe, the Russian Federation has 281,751, and the United Kingdom came third with 240,165. Interestingly, six of these countries are from Europe, Asia, and America, with three countries each, while the rest are not among the 12 leading countries.

Table 8.1   The confirmed and death cases by continents.

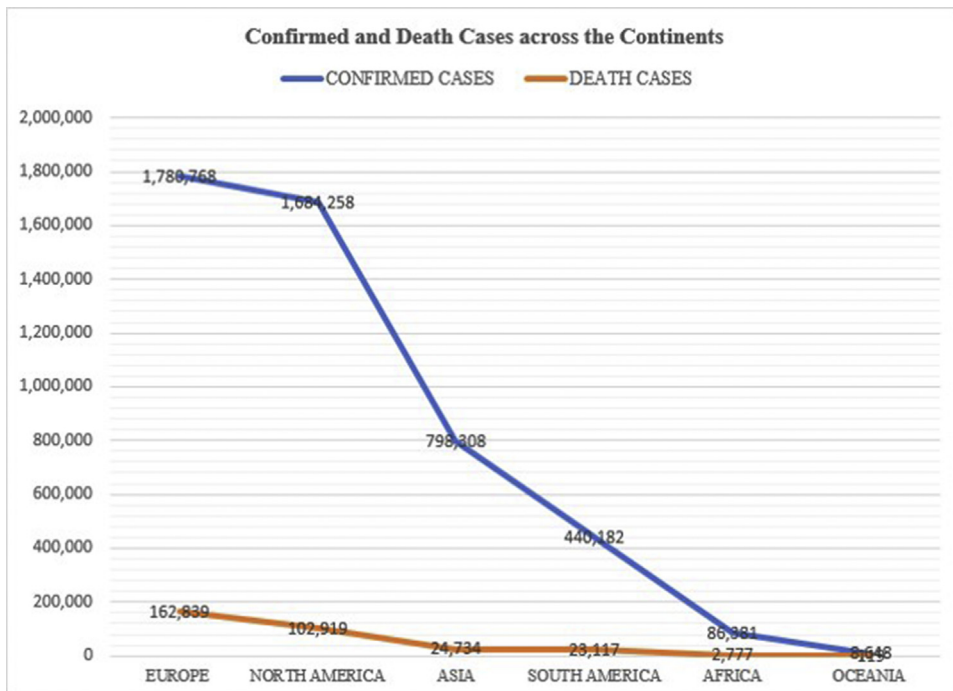| Continents	Confirmed cases	Death cases
Europe	1,780,768	162,839
North America	1,684,258	102,919
Asia	798,308	24,734
South America	440,182	23,117
Africa	86,381	2777
Oceania	8648	119

FIGURE 8.1 The confirmed and death cases across continents.

Table 8.2   The 12 countries with the highest number of confirmed cases in the world.

Country name	Total cases	Percentage (%)
The United States of America	1,409,452	42
Russian Federation	281,751	8
The United Kingdom	240,165	7
Spain	230,698	7
Italy	224,760	7
Brazil	218,223	6
Germany	174,355	5
Turkey	148,067	4
France	140,008	4
Iran	120,198	4
India	90,927	3
Peru	84,495	3
TOTAL	3,363,099	100

This chapter, therefore, discusses the possible application of big data and analytics to improve conventional public health approaches to tackle COVID-19 to control, manage, identify, and avoid COVID-19 and reduce the effect on health implicitly associated with COVID-19. The remaining part of this chapter is organized as follows: Section 2 discusses the growth of data in healthcare, the challenges, and the importance of big data in COVID-19. Section 3 presents the big data privacy and ethical challenges in COVID-19. Section 4 discusses big data analytics in the COVID-19 epidemic. Finally, Section 5 concludes the chapter.

2. The growth of data in healthcare

The mutual collaboration that exists in the healthcare sector results in the generation of a huge amount of data from various sources [36–38]. This has a result of huge figures of varied sections and departments including physicians of different disciplines, ranging from nurses, pathologists, radiologists, and laboratory technologists, which collaborate efforts to achieve unified goals toward bringing medical costs and mistakes down to the barest minimum, as well as to provide excellent and standardized healthcare services. Various stakeholders that work collaboratively in the health sector get data from different sources. This source includes data from patient observation, imaging reports from scans, interview and test results, insurance and bills, summaries of patient discharges, reports from pharmacists, case notes from physicians, admission notes from hospitals, feedback from social media, and journals of medical articles.

Data in the healthcare sector are usually huge and not easy to handle [39,40]. This results from the enormous way by which data grows in the healthcare sector, the rate at which data are produced, and the variety of data in the healthcare system [36,37].

The rate at which data are captured, stored, analyzed, and retrieved in the healthcare sector has changed rapidly from the aged paper-based storage technique to the use of digital techniques and methods [37,41].

On the other hand, the complication of data makes processing and analysis of data by the age-long traditional method very difficult and uneasy to handle. However, the vast volume, as well as the complexity, of these data makes it difficult for the data to be processed and analyzed by traditional approaches and techniques. Therefore the application of advanced technologies, which includes virtualization and cloud computing, allows for huge and effective data processing in the healthcare system, thereby rapidly turning the healthcare system into a big data industry. Nevertheless, in these modern days, the improvement in the information and communication technologies (ICT) brings the advancement of varied data from new sources in the healthcare system (Fig. 8.2).

This source includes the Global Positioning System (GPS), data from gene sequence, file logs, devices that identify radio frequency (RFID), smart meters, and posts from social media. The increasing rate with which data is being produced from various sources brings about an increase in the amount of data in the healthcare system [37,43]. Thus their results give tedious means of storing, processing, and analyzing data with the agelong traditional method of data processing applications [44]. Nevertheless, modern methods and tools, as well as advanced computing technologies, are used to store up, manage, and explore values from large and varying data in the healthcare system in a real-time manner [37,43,44]. Therefore the healthcare system has now become a big data industry. Big data now brings huge opportunities to the healthcare system [45].
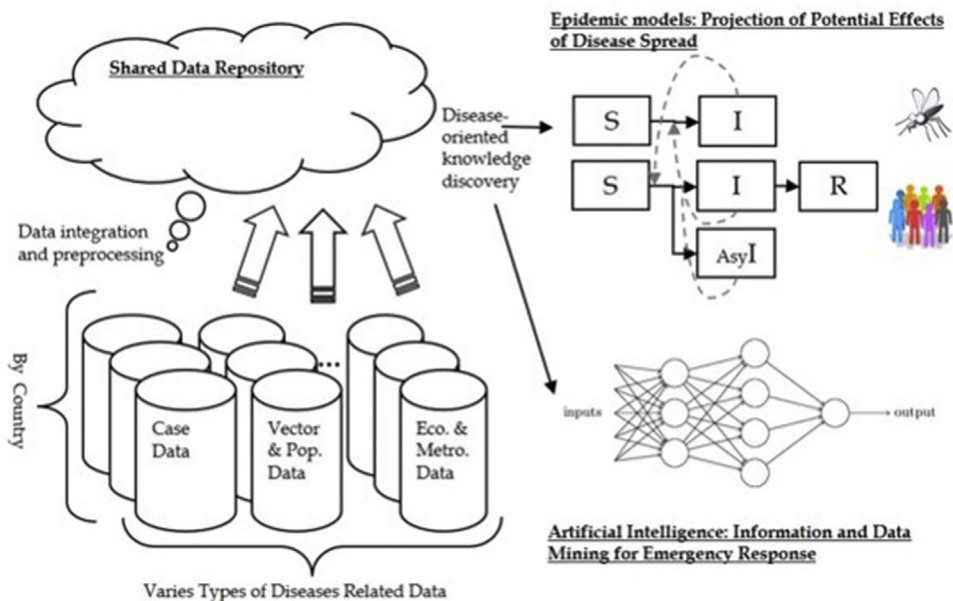

FIGURE 8.2 An integrated big data conceptual model for infectious diseases [42].

Improvement in information technology and data computing has greatly changed population-based research by encouraging easy access to a huge amount of data. Sometimes such database links are referred to as "big data" [46,47]. In order to make efficient use of these data for research in clinical health or public health, the researchers need to widen research further than the traditional surveillance model, as operating with big data differs from focusing on performing narrow analysis, treatment-oriented clinical data. Therefore leveraging on big data to reflect accurately on the heterogeneous population it represents becomes expedient [37,43]. This endeavor needs a swift research environment that can adopt a quick advancement in computing technology to an all-time combined data while making use of new methods to reduce complexity [37,47]. Also, trends and patterns that make it easy to diagnose and treat patients are revealed by big data.

As big data in today's digital world will be crucial to handed the COVID-19 outbreak, the criteria for effective collection of data and analysis on a global scale need to be clear. The chapter claims the use of prediction and surveillance of digitally accessible data and algorithms. For example, it is of vital importance in the battle against the COVID-19 pandemic to recognize people who have traveled to places where the disease has spread or has been identified and isolate the contacts of infected people. However, it is equally important to use these data and algorithms responsibly, in compliance with the data protection regulations and with due regard for privacy and confidentiality. The inability to do so would weaken public confidence, which would cause people less inclined to obey recommendations or guidelines from public health and more likely to have worse health outcomes [48]. Hence the exploitation of big data in COVID-19 will bring about improved care with minimum cost and good satisfaction to patients.

2.1   Challenges of big data in COVID-19

The ability to combine sophisticated epidemiologic models and new big data sources has been discussed [49,50]. For example, the use of smart-phones, social networks, or public health response satellites at different stages, mainly in the situation of disease forecasts, for making a decision [50,51]. These new sources of data include significant, real-time data on tourism visits that spread infection and spatial shifts in vulnerable residents that probably up till now were hard to report on timescales related to a rapidly fast-growing epidemic [50]. With increasing flexibility and growing universal connectivity, this information will be crucial for monitoring and containment planning [52].

By principle, with correct data sharing protocols in place, accurate, up-to-date disease predictions can be generated that are guided using this data stream. The success of big data in fighting epidemics depends on various organizations, both public and private, and the government is needed to be in contact frequently [50,51]. Privacy issues for daily use of new data channels need to be addressed, specifically the most suitable way to vigorously integrate such data streams to confirm individual confidentiality. Yet even if security is tackled, the transformation of new methods into a policymaking context raises additional systemic challenges [50,51].

Big data mining provides many tempting opportunities [43,53]. Nonetheless, when investigating big data sets and mining meaning and expertise from these facts extracted, researchers and professionals face many challenges [43,53−55]. The complications at various levels include data collection, storing, scanning, sharing, reviewing, handling, and watching. Furthermore, security and privacy issues particularly arise in distributed applications that are powered by data. The surge of knowledge and distributed sources also surpass our harnessing capability indeed, although the scale of big data continually grows exponentially, the actual technical ability to accomplish and explore large datasets is only in the comparatively lower rates of data petabytes, exabytes, and zettabytes [43]. In this chapter, we address in detail some technical issues that are still open to research.

While working with big data, data scientists face a lot of obstacles. One obstacle is the compilation, integration, and storage of tremendous datasets generated from disseminated sources, with far less needed hardware and software [56,57]. The management of big data results in another huge problem. Effective big data management is essential to promote the abstraction of accurate information to maximize investment [58,59]. Proper management of data is the basis for big data analytics [60].

Another difficulty is to synchronize with an organization's internal infrastructure's external data sources, and big data dispersed systems (sensors, apps, databases, networks, etc.) [43]. Most of the time, an analysis of the data produced within organizations is not sufficient. We need to go one step further to integrate internal data with external data sources, to gain useful insight and information. External data may include the sources from third parties, market fluctuation details, weather forecasts and traffic situations, social network data, comments from clients, and resident input. This will aid to optimize the power of analytic models used in the analysis, for example.

The computer design and performance are an important question. Indeed, CPU output is known to double every one and half years according to Moore's Law and diskette drive output is also expanding at the same time. The input and output operations, however, do not follow the same form of results (i.e., spontaneous input and output speeds increased moderately, while systematic input and output speeds developed slowly through thickness) [57]. Thus this varied device capacity will delay data access and affect the efficiency and scalability of big data applications. Machine learning is a very active area of study in artificial intelligence (AI) and pattern recognition nowadays. This plays a significant part in applications of probabilistic learning, such as computer vision, voice recognition, and the interpretation of natural languages [61]. Currently, the ability of computer-learning techniques and functionality engineering algorithms to process raw natural data is limited [62]. Deep learning is, on the contrary, more effective in solving data processing and learning complications contained in large datasets. This helps extract a huge size of unsupervised and unstructured raw data automatically. Dynamic policies learning consists of incremental and group learning that are important approaches to learn from a large stream of data with meaningful perception [63].

Big data and data stream uses the concept of collaborative and incremental learning. They deal with numerous challenges that arise during the learning of data, such as data accessibility and restricted resources. They make use of user profiling and stock forecasting applications. The application of incremental learning to data produces a faster means of prediction when dealing with the new data stream. Hence, using the incremental algorithm is highly recommended when the drift concept is not available. On the other hand, it is highly recommended to use ensemble algorithms to obtain accurate outcomes with large drift concept. Also, when dealing with a modest data stream or real-time processing, the incremental algorithm should be considered for use. Conversely, in the case of a complex data stream, an ensemble algorithm is considered the best.

Another big problem is analogic reasoning: fundamentally, disease forecasting is unpredictable. In the narrative on big data and AI, there is often an underlying assumption about the statistical covariates or cell phone data and difficult models of simulation that avoid the need to gather basic epidemiologic details. Nevertheless, for the evolving COVID-19 epidemic, emphasizing this opinion still lacks a reliable fact on situation counts and biotic processes that drive an outbreak of the disease, let only the behavioral reactions of infected individuals, and make it difficult to rapidly adjust or understand precise intricate prototypes on spatiotemporal scales applicable in making a verdict. Hopefully, the best effective systems would continue to be simple [64,65] equally out of necessity for versatile modules that provide quick responses, given the controversy surrounding emergency epidemiologic studies and because simple models are more easily available to interpret and communicate [66]. Clear knowledge of both the importance and weaknesses of typical outputs is a precondition for their successful implementation but is often absent. Because policymakers have not yet reached an in-depth modeling experience, the lack of consistent message threatens to lead to two destructive results: (1) assume that the method would be misinformed without skepticism and choices or (2) ignore off-hand modeling and neglect the proof that we must control epidemics most efficiently.

During epidemics, verdicts have to be taken rapidly, on a patchy basis, and inaccurate data models could be effective tools for guiding them. Considering the abovementioned obstacles, computing capacity, novel methods, and novel data are steadily advancing sources that give authentic courage for improved monitoring and development of valuable predicting systems. The novel database available to us comprises not only inactive experimental large data streams from cell phones but also comprehensive ecological data and local sensor information from embedded devices, internet search information, pathogenic genomic data that can be produced quickly informing the response during an outbreak, [67] and crowd-sourced methods to track emergencies changing rapidly [68]. Information sharing systems and structured aggregation strategies are being established to secure personal data privacy [69], and internet access enables rapid data transmission and collaboration between geographically distant responding teams. Methodologically, effective modeling methods that combine multiple

predictions to reduce uncertainty are being built [65,70]. In reaction to the outbreak of COVID-19, an innovative, collaborative approach has unfolded between academic organizations, for example, between Twitter and other channels to exchange, evaluate, and publicly debate the effects of new research as it arises (Ironically, the control of misinformation that is now proliferating on social networks during a crisis is likely to become one of the most significant concerns for containing the disease.).

Unless the abovementioned problems are resolved, these technologies will remain dislocated and impracticable. It is promising that all these problems may be strengthened by transferring most of the funding and knowledge focus to those communities that require most help during epidemics. The unpredictability of the COVID-19 outbreak and the progressive technologic mechanisms of these methods will require agile, dispersed groups of people covering the systematic and functioning dimensions of the epidemic response using novel big data methods to supplement and explain the limited spread, sociogeographic data that are necessary for the progress of important predictions. Provincial or small groups with investigative experience and current connections with government and industry stakeholders can contribute to scalable modeling strategies that exploit broad curated and international scientific database skills, building on partnerships formed if no new or ongoing outbreaks occur. This strategy may likewise be of help to mitigate some policy concerns related to data sharing, particularly through sensitive information not being shared publicly.

To achieve this idea, significant continuous investment is required to finance individuals to fill what currently represents a significant void in the analytics process, especially in some countries with low and middle wages. Big data can be created from a lot of sources, such as social network sites, smartphones, IoT sensors, and publicly available data [71], in different formats such as text or video. Big data applies to patient care details in the sense of COVID-19, such as medical records, X-ray reports, case history, list of doctors and nurses, and information about the epidemic location. Big data has proved its capability to support fighting infectious diseases like COVID-19 [72,73].

2.2   Importance of big data in COVID-19

The possibility of fast-spreading diseases that cause uncontrolled death and a bad influence on the economies and sustainability of countries in the world has stretched the importance and needs for developing a quantitative framework to provide support for making nearly real-time decisions in the public health system. According to a case study, the 2003 SARS outbreak that originated in China and spread to 29 countries became a healthcare-acquired infection in various regions by August 2003 [74,75]. The 2009 pandemic known as influenza (H1N1) originated in Mexico and spread rapidly all over the world through airline network and infected 20 countries, with maximum hit on the travelers coming from Mexico within a little time of the disease outbreak [76]. The socioeconomic implication and outcome of the outbreak of a disease related to 2009 (influenza A [H1N1]) were projected to be a huge amount ranging between $360 billion

and $4 trillion [77], which was the calculation for the outbreak in the first year. Significantly, the SARS outbreak at Wuhan, China, in 2019 was calculated and projected to be 2,034,802 confirmed cases and 123,150 mortalities across the globe. Ever since the outbreak of COVID-19 in China, people who traveled globally from Wuhan transferred the disease to various countries of the world.

The outbreak of SARS had significant effects on the people and economy all over the world. The effects include restrictions and bans on traveling, total closure of shops and markets resulting in a huge loss on earnings and proceeds, fear of contracting the disease from the infected person, and unpleasant side effects on world tourism at large, as many flights were canceled or suspended. Express and huge damage on government finances in various countries as agricultural activities were greatly hampered resulting in food shortage, and so many other effects resulted from the outbreak. Therefore it becomes imperative to make an immediate response from all angles to fight and contain the virus efficiently. Big data entails the innovative processing of both new and existing data to proffer meaningful solutions that are of huge business benefits [78–80]. However, processing large data, or varied data, will only be a mere technologic solution until it is associated with business objectives and goals.

Big data promises more lights to be shed on complicated facts and information on the dynamics of how infectious diseases are transmitted and how to develop new modeling and analytic tools that have far been hidden due to lack of smooth and workable data. More progress will be made in the area of disease forecast when a huge amount of information on epidemics required for modeling is available. Big data will be the best in accomplishing this goal. Improved consistency in reporting diseases and case definition is therefore expedient within time and space. Data obtained from news media sources can give a perfect estimation and evaluation of how the disease is transmitted. Thus it is of high importance, especially when comprehensive and well-examined data are not available. In a country with low and middle income, and where studies on disease transmission details are scarce, resources based on internet surveys will provide an opportunity for a detailed study and correct assessment on how the disease is transmitted, for instance, in the crisis of an infectious disease like COVID-19 when real-time assessment is expedient.

A well-calibrated computational model tool that is useful in assessing and analyzing the epidemic spread signifies a potent tool to sustain the process of decision-making at a time of emergency epidemic outbreaks. Epidemic models are progressively used to generate predictions of the spatial-temporal progression of a disease outbreak at several spatial scales and for easy access to possible impacts of the various strategies of the intervention [81]. The ability to produce innovative facts or information is greatly expanded by big data. The costs applicable to answer various questions retrospectively and prospectively by gathering ordered and structured data are exorbitant. Acquisition of finer data in a computerized format is achieved by analyzing unstructured data that are within EHR systems by using various computational techniques such as traditional language processors that extract clinical concepts from free documents. Analyzing the

unstructured data contained within her systems using computational techniques (e.g., natural language processing to extract medical concepts from free-text documents) permits finer data acquisition in an automated fashion. For example, computerized detection within her systems with the use of natural language processing is better used to detect complications that arise after the operation when compared to patient well-being indicators based on discharge coding [82]. Big data provides the possibility to generate an observable indication base for clinically related questions, which may not be achievable but will be helpful with generalizability issues. Generalizability issues restrict the application of conclusions that are derived from random trials carried out on a very narrow scale of participants to patients that show dissimilar characteristics. Big data helps disseminate knowledge. Many healthcare practitioners strive to be updated with current evidential proof and support that guide clinical practices.

The computerization of medical journalism to a great extent has enhanced access; conversely, with great numbers of studies, translation of knowledge becomes difficult. If at all a physician has enough relevant data and guiding principles, organizing information to formulate a realistic approach for the treatment of patients with numerous chronic illnesses becomes extremely difficult. Analyzing existing EHR systems to generate a template to guide in making clinical decisions is the only way by which this problem may be solved [83]. This technique is exploited in the alliance involving Memorial Sloan Kettering Cancer Center and IBM's Watson supercomputer to aid in the diagnosis and proposed treatment for cancer patients. The difference between the big data method and traditional decision support tools is that with the big data technique, suggestions are obtained through analysis of patients' data in real time instead of depending solely on the use of rule-based decision trees. For instance, longitudinal analytic data have been proven to be efficient for forecasting future diagnostic risks and domestic abuse in patients [83,84].

With data-driven clinical decision support tools, the cost is greatly minimized and proper care standardization is guaranteed. Physicians can get information that guides them on the diagnostic and therapy options that are provided by esteemed colleagues encountering related patient profiles with the use of big data on the cloud. The healthcare system can be transformed by getting information directly across to patients and equipping them to take part in more active roles. In the future, clinical reports can also reside with the patients, compared to the existing model where health records of patients are stored with healthcare professionals, thereby making the patients be in a passive situation. Big data provides the opportunity to progress the clinical records by connecting the conventional health-related data such as family history and medication list to other private data located on other sites such as education, income, neighborhood, military service [83,85], exercise regimens, diet habits, and forms of entertainment, all that can be easily accessed without the need to interview the patient with a comprehensive list of questions. Thus big data provides an opportunity to collaborate with the conventional medical model with social determinants of health in a patient-directed fashion. The initiatives of public health to reduce obesity and smoking can be

efficiently delivered by focusing the message on the most appropriate set of people based on the profiles on their social media. With the opportunity to use analytic capabilities, system biology can be incorporated.

Big data assist in translating personalized medicine initiatives into clinical practice (e.g., genomics) with data [83,86]. The Genomics Network and Electronic Medical Records achieve this with the use of natural language processing to phenotype patients, in an attempt to simplify genomics research. With big data, patients are opportune to access correct and updated information. This helps patients to comprehend their choices, take decisions about their care, and also better their lifestyle to protect them from chronic diseases. Big data makes it easy to detect and identify/diagnose diseases at an early stage. Thus the right decisions on how to treat a certain disease efficiently and on time are guaranteed. Hence death rate and the rate of contracting diseases are greatly reduced. Big data aids the quality level of care given to patients by making sure that decisions are focused on a huge amount of related and updated data.

Big data is promising in the area of identifying public health intervention targeted through the analysis of large volume and varied data and improved subsequent interventions by the use of large-velocity feedback mechanisms [87]. The quantity of data generated from the existence of the human race to 2003 can now be obtained in a few minutes [25]. Also, improved computational models, such as machine learning-based models, have revealed enormous possibilities in tracing the source or forecasting the spread of a novel disease in the nearest future [88,89]. Therefore leveraging big data and intelligent analytics becomes expedient in order to enjoy their use in COVID-19 outbreak and public health.

3. Big data privacy and ethical challenges in COVID-19

The latest outbreak of COVID-19 has brought the difficulty of securing personal data to a head in a transnational sense [90]; this is because COVID-19 spreads fast with the international travel of people [91]. Many countries require international travelers to disclose their personal information such as the name, gender, date of birth, travel history, purpose of travel, and residence, among others, and impose quarantine requirements accordingly [92]. Using a genuine case where the Chinese media secretly published the sensitive information of a foreign traveler, the article describes that multiple patterns for LEX cause emerged at each point of dispute-of-law analysis: (1) the European Union, the United States, and China vary in characterizing the right to personal data; (2) the expanding centralized approach to relevant legislation lies in the fact that all three territories either find the law on personal information privacy to be a contractual law or follow linking factors leading to the law of the forum; and (3) actively support the de-Americanization of meaningful data privacy legislation. The patterns and their mechanisms have important consequences in the application of regulations for transnational information [90].

Evolving contagious diseases such as HIV/AIDS, SARS, and pandemic influenza, and the 2001 anthrax attacks, have proven that we appear vulnerable to health risks from contagious diseases [93−95]. The key expert recommendation for two decades has been the value of improving global public health surveillance to provide an early warning [94,95]. Based on digital information sources, including data from smartphones and other digital equipment, the outbreaks triggered by new findings are of special importance in infectious diseases [25,93,95], for which the survey data and accurate predictions are indeed limited. Recent research shows the likelihood of predicting the spreading of COVID-19 by integrating Government Aviation Guide data with WeChat App data on human accessibility and other digital platforms run by the Chinese technology giant Tencent [96].

Smartphone data also demonstrated possibilities for modeling regional cholera outbreak in Haiti during the 2010 outbreak [97], while using big data analysis demonstrated efficacy during the Ebola crisis in the year 2014−16 [98]. Moreover, the big data gathering of cellular data from users around the world, particularly mobile data archives and social network accounts, also raised questions about security and data integrity during those epidemics. In 2014, the GSM organization (a sector body that serves the benefits of major mobile network operators) urged privacy issues to issue privacy rules for the use of cell phone data to respond to an Ebola outbreak [25,99].

In the information-intensive world of 2020, these issues can easily be compounded by pervasive data points and automated surveillance devices. China is the area hardest hit with COVID-19 and has reportedly used omnipresent transmitter data and safety monitoring software to avoid disease spread [25,100]. The *New York Times* reports [101] that how these data are reviewed and recycled for monitoring reasons is not clear. For instance, the document said that Alipay Health Code, an Alibaba-backed government-run app that helps decision on who is to be isolated for COVID-19, also seemed to share details with the police [25,101]. The European Republic in Italy with the highest figure of COVID-19 incidents, the remote data-privacy agency was advised to issue a statement [102] on March 2, 2020, to explain the requirements of permissible data usage for containment and prevention. In its declaration, the expert cautioned against data gathering and analysis by a noninstitutional body that violates privacy [102,103].

A couple of days later, the European Data Security Committee published a press release on the idea of securing people information if used to counter COVID-19 and highlighted relevant research of the General Data Protection Regulation, which include legal reasons for the handling of people information in the form of outbreaks [90]. For instance, the article enables the compilation of people information "for purposes of general interest in the field of health research, including guarding toward severe cross-border challenges to health," given that such analysis is directly proportional to the purpose sought, reverences the principle of the right to data privacy, and safety the rights and liberties of the person concerned. Big data is recognized as important tools for COVID-19 pandemic management in this era of technologic advancement; there must be specific criteria for effective gathering and analysis of data on a global scale.

They claim that the usage of technologically accessible data and techniques for fore-casting and tracking, e.g., describing people who have traveled to places where the infection has transmitted or has been identified and quarantining the people contacted, is of vital importance in combating the COVID-19 outbreak. However, it is equally necessary to use these data and technologies responsibly by data privacy legislation and with proper regard for safety and security. Failure to do so will weaken public confidence, causing people less likely to obey advice or guidelines on global health and more likely to have worse patient outcomes [48].

Practices of conscientious data management can guide both collection of data and information processing. The concept of proportionality will apply to the sharing of information from the persons affected, which means the collection of data must be proportionate to the severity of the hazard to public health, be partial to what is required to accomplish particular public health goals, and be systematically proven. For example, obtaining access to personal contact device data tracking purposes may be acceptable if it happens inside defined limits; if it has a valid reason, such as alert and quarantine those who have had contact with the infected person or the virus itself; and if no less-intrusive option, such as using anonymized mobile tracking data, is appropriate for that reason. Also, health monitoring "do for yourself," as the Italian data protection agency has called it, must be resisted.

Data product quality standards are required at the data analysis stage [25,104]. Deficiencies in data protection, which are normal when data from digital personal devices are used, can lead to minor errors in one or more factors [105,106], which in turn may have an enormous impact on predictive large-scale models. Also, privacy violations, inadequate or unsuccessful deidentification, and prejudices in databases can become major causes of public health mistrust [107,108]. Data protection issues are not only of a technologic aspect but also of a legal and political nature [109,110]. For purposes tracing the index or infected cases, demanding or guaranteeing access to personal devices may be more successful than simply exploiting anonymous mobile tracking data.

Taiwanese studies display a compelling method to exploit big data analysis to react to COVID-19 without fostering public distrust. To improve classifying incidents, Taiwanese agencies have combined their national health insurance database with customs database travel history. Certain tools have also been used for surveillance purposes, such as QR code scanning and electronic monitoring. Such interventions were paired with techniques for official communications including regular health checks and support for those under lockdown [111]. More nations are planning to use big data and emerging technology in the battle against the evolving COVID-19 outbreak, so if used appropriately, data and algorithms are one of the main resources in our bow [25].

4. Big data analytics in COVID-19 epidemic

Big data became a big dilemma a few years ago. In the 2000s, big data overwhelmed the technologies of storage and CPU systems by the abundance of terabytes of data required to be stored. The IT industries face the problem of data scalability as the need to store a

large volume of data skyrockets [112,113]. The rapid rate at which health information technologies advance leads to the need for big data in the hospital environments and the healthcare system at large. However, the extraction of meaningful information from the highly loaded datasets is still inadequate and very limited. The decision-support tools that are required for efficient handling of rich and large-scale healthcare records, optimizing operational dynamics in the healthcare environment, and extracting relevant knowledge and information with regard to the health conditions of patients and availability of additional and efficient healthcare services are seriously lacking in the practice of general medicine. The healthcare system grows from not being able to handle big data to wasting huge budgets on its collection and analysis, respectively. As of now, the healthcare system is leveraging big data to extract essential information that was not acknowledged in the past.

The healthcare system can now study big data to have a better understanding of the current situation on health and still be able to track evolving aspects such as patient behavior by using advanced and sophisticated analytic tools. To gain more insight into better healthcare intelligence and facts, it is helpful to tap into data that has never been tapped before. Most of this untapped information will be very new to you as they come from devices, sensors, third parties, social media, and web applications. Data are obtained incessantly in a real-time manner from some big data sources. When put together, one can easily detect that the big data not only is about a huge volume of data but also focuses on astonishingly diverse data types that can be delivered at varying frequencies and speeds. Hence it is noteworthy that we now have the coming together of two technical entities.

The first entity is big data that has been used for enormous and detailed information. Second, the advanced and more sophisticated analytics, which comprises of a collection of various types of tools, that includes tools that are based on natural language processing, data mining, AI, statistics, predictive analytics, and many more. When all of this is put together, you get what is called the big data analytics, which is the hottest new practice today in Heath Intelligence [113−115]. The availability of massive data presents unlimited opportunities to visualize, manage, analyze, and extract meaningful information from huge, varied, dispersed, and heterogeneous datasets that enable making better medical decisions and improved performance in the healthcare system. Across all corners of the healthcare system, big data is inspiring a reflective transformation. The manner with which biomedical research has been carried out and how the delivery of healthcare services has been managed across the globe are the new transformations in informatics and analytics research. The analysis of big data in COVID-19 involves several tools and sophisticated technologies as illustrated in Fig. 8.3. Big data can provide healthcare intelligence for the COVID-19 outbreak, thereby being useful for governments, organizations, and healthcare policymakers for smart and quick response during this pandemic. The big data if not analyzed become a liability and unrealistic to the owners and users. For intelligent prediction, forecasts, and decision-making, big data analytics involves mining and extracting useful knowledge. Scalability, visualization, and
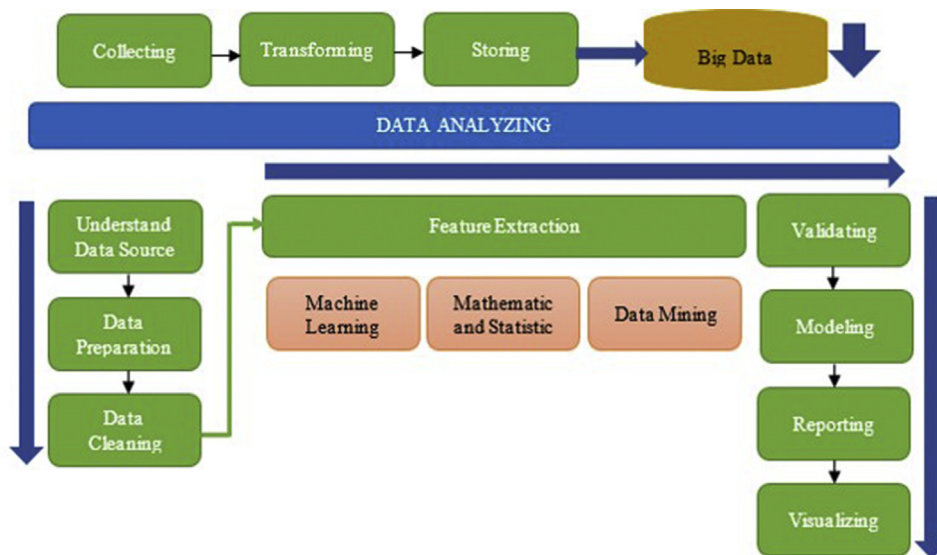
FIGURE 8.3 Big data analytics.

computation of data are the challenges in big data analytics, thus the flow in the amount of data increases the information security risk.

Making better medical decisions, enhanced patient management, monitoring systems, and efficient public health supervision are progressively viewed by the government, the general public, and the medical community as a major key to support the improvement of and reduce the cost of the healthcare arena [49,114]. The operations of advanced analytic methods on big data are known as big data analytics. With the gigantic statistical samples that big data provides, an improved result of the analytic tool is also enhanced. Various tools intended for statistical analysis and data mining are liable to be optimized for large amounts of datasets [113]. It is a general rule that the accuracy of statistics and other products of the analysis depends and increases with how large the data sample is [113,116]. The latest generation of tools for data visualization and analytics for in-database functions similarly operates and functions on big data [113,116,117].

Big data analytics simply is the procedure, techniques, and technologies used in analyzing big data to discover useful information through the use of nontraditional and advanced methods [118,119]. It is designed in such a way that it increases accuracy and scalability when compared to the conventional methods, such as the regression-based models and many other statistical models. AI, among other advanced techniques, has been acknowledged as a very significant development in the roles it plays in various application domains [120], together with other disciplines related to public health [121]. In a vacuum, big data is worthless. Its prospective usefulness can only be tapped when it is adopted for decision-making. Organizations require an effective means of revolving
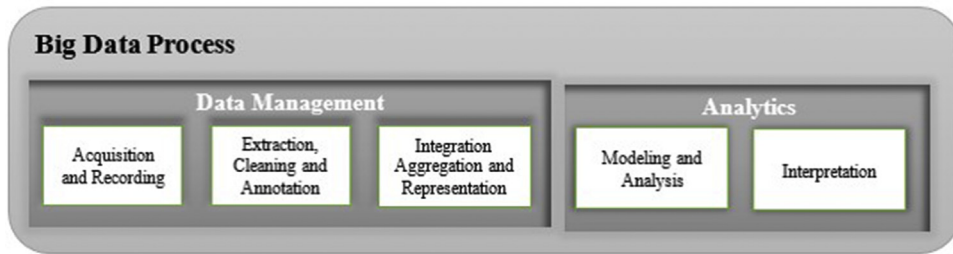
FIGURE 8.4 Processes for extracting insights from big data.

large volumes of fast-growing and varied data into useful insights to facilitate the making of evidence-based decisions.

The general techniques for extracting useful information from big data can be classified into five different stages [1,44]; Fig. 8.4 shows the five stages that depict the two major analytics and management subprocesses. Data management comprises of methods and technologies that support the means of acquiring, storing, preparing, and retrieving data for analysis. Conversely, analytics describes the technique used in analyzing and obtaining intelligence from big data. Hence big data analytics is seen as a subprocess in the overall technique of extracting insights from big data.

By combining and efficiently using big data in a digitalized manner, the healthcare system and organizations that range from single-physician workstations and multi-provider groups to big hospital networks and responsible healthcare organizations stand to gain many benefits [118,122]. The prospective advantages of big data analytics in COVID-19 consist of early disease detection, which can then be treated effortlessly and efficiently; maintaining precise individual and public health; and easy detection of healthcare frauds speedily and effectively [118].

Addressing numerous questions is made possible with big data analytics. Predictions and estimations are based on a huge amount of chronological data, which includes how long a patient will stay, numbers of patients who will prefer an elective surgery, other hospital-acquired illness/the rate of disease progression, disease causal factors, and risks of disease advancement in patients [118]. Yearly, big data analytics will facilitate over $300 billion in total savings in the US healthcare system. Two-thirds of such reductions amount to approximately 8% in the national healthcare expenses. Two biggest areas of prospective savings with $165 billion and $108 billion savings are medical operations and research and development (R&D), respectively [118,123]. Big data will possibly assist in cost reduction and ineffectiveness in three areas. Clinical operations: relative efficient research to establish added medically related and cost-efficient ways of diagnosing and treating patients.

R&D: (1) the use of predictive modeling to reduce the error rate and construct faster, compact, and more centered R&D pipeline in drugs and devices; (2) statistical tools and algorithms to progress medical trial design and patient enrollment to match treatments to patients efficiently by reducing trial failures and encouraging advanced treatments to

patients, thus reducing trial failures and bringing new treatments to the market; and (3) to examine medical trials and patient records to detect and follow-up on indications and adverse effects before the products reach the market. The speedy examination of various symptoms in diseases, and tracking of disease outbreak in real-time have really help in public healthcare systems. The BDA has speedy improved the production and development of vaccines for COVID-19 outbreak. For example, the conversion of huge amount of data to a meaningful information has been used to predict, forecast, and diagnose various infectious diseases and in the development of vaccine for the advantages of the population at large [118,123].

Hence big data analytics has a high ability to innovate the approach so that healthcare providers can use complicated technologies to provide insight from their medical and many other repositories for data to ensure making informed decisions regarding the COVID-19 epidemic and the outbreak of any associated infectious disease.

5. Conclusion

The December 2019 COVID-19 is known as a pandemic because it has cut across the globe. It has become a great threat to global health. It has caused close to 2.5 million infections and 180,578 deaths in almost 213 countries by April 22, 2020, and the daily increase in this infectious disease has become a huge challenge to the socioeconomic development of each country across the globe. The application of big data offers possibilities for forecasting and analyzing viral behavior for directing healthcare in individual nations to enhance their readiness for the COVID-19 outbreak. This can be achieved using different global databases, for instance, the Official Aviation Guide, the Tencent Location Services (Shenzhen, China), and the Wuhan Municipal Transportation Management Bureau conducted a model analysis of "now-casting" and prediction of COVID-19 development inside and outside China that could be used by public health planning and controlling authorities around the world [124]. Likewise, the WHO International Health Regulations, the State Parties Annual Reporting Guide, the Joint External Assessment Reports, and the Infectious Disease Vulnerability Index were used to evaluate the readiness and vulnerability of African countries to combat COVID-19, which will help raise awareness among the health authorities in Africa to better plan for the viral outbreak [48,125].

Big data potentially provide several promising solutions to help combat the COVID-19 epidemic. The use of big data with different analytic tools will help understand COVID-19 in terms of outbreak monitoring, virus development, disease control, and the manufacturing of vaccines. Big data coupled with intelligence-based applications can create complex prediction models using coronavirus data streams to estimate the outbreak, thus allowing health authorities to track the spread of coronavirus and plan for effective preventive measures. Big data models also support future forecasting of the COVID-19 outbreak through their ability to combine enormous quantities of data for prevention and treatment. Moreover, big data analytics from a variety of real-world

sources, including infected patients, can help implement large-scale COVID-19 investigations to develop comprehensive treatment solutions with high reliability [126,127]. This would also help healthcare providers to understand the virus development for better response to the various treatment and diagnoses. As one of the most effective solutions to combat the COVID-19 pandemic, early treatment and prediction are of importance. Hence the capability of big data in the fight against the COVID-19 outbreak can be useful and divided into four main areas of application: outbreak virus spread, monitoring forecasting, diagnosis/treatment of coronavirus, and discovering vaccines/drugs. In response to the COVID-19 virus globally, different traditional approaches were dedicated to improve future forecasting, prevention, and treatment, with other alternatives. However, these approaches are typically costly and time-consuming, have a low positive result rate, and require different materials, equipment, and resources. Moreover, most countries are suffering from a lack of testing kits because of the limitation on budget and techniques. Thus these standard methods are not suitable to meet the requirements of fast detection and tracking during the COVID-19 pandemic. Therefore the use of connected devices, along with big data, is an easy and low-cost approach for COVID-19 detection. In the COVID-19 pandemic battle, developing efficient diagnostic and treatment methods plays an important role in mitigating the impact of the COVID-19 virus.

References

[1] A. Gandomi, M. Haider, Beyond the hype: big data concepts, methods, and analytics, Int. J. Inf. Manag. 35 (2) (2015) 137−144.

[2] H.E. Brady, The challenge of big data and data science, Ann. Rev. Polit. Sci. 22 (2019) 297−323.

[3] V. Dhar, Data science and prediction, Commun. ACM 56 (12) (2013) 64−73.

[4] J. Leek, The keyword in data science is not data, it is science, Simply Stat. 12 (2013).

[5] L. Cao, Data science thinking, in: Data Science Thinking, Springer, Cham, 2018, pp. 59−90.

[6] B.T. Hazen, C.A. Boone, J.D. Ezell, L.A. Jones-Farmer, Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications, Int. J. Prod. Econ. 154 (2014) 72−80.

[7] E. Szymańska, Modern data science for analytical chemical data−a comprehensive review, Anal. Chim. Acta 1028 (2018) 1−10.

[8] I. Carmichael, J.S. Marron, Data science vs. statistics: two cultures? Jpn. J. Stat. Data Sci. 1 (1) (2018) 117−138.

[9] W. Van Der Aalst, Data science in action, in: Process Mining, Springer, Berlin, Heidelberg, 2016, pp. 3−23.

[10] D. Donoho, 50 years of data science, J. Comput. Graph. Stat. 26 (4) (2017) 745−766.

[11] Y. Wang, L. Kung, T.A. Byrd, Big data analytics: understanding its capabilities and potential benefits for healthcare organizations, Technol. Forecast. Soc. Change 126 (2018) 3−13.

[12] T.M. Choi, H.K. Chan, X. Yue, Recent development in big data analytics for business operations and risk management, IEEE Trans. Cybern. 47 (1) (2016) 81−92.

[13] B. Shi, S.S. Iyengar, General framework of mathematics, in: Mathematical Theories of Machine Learning-Theory and Applications, Springer, Cham, 2020, pp. 13−16.

[14] A. Sarkar, A. Pal, Where does data science research stand in the 21st century: observation from the standpoint of a scientometric analysis, Libr. Philos. Pract. (2019) 1−9.

[15] D.M. Blei, P. Smyth, Science and data science, Proc. Natl. Acad. Sci. USA 114 (33) (2017) 8689−8692.

[16] T. Hulsen, S.S. Jamuar, A. Moody, J.H. Karnes, V. Orsolya, S. Hedensted, et al., From big data to precision medicine, Front. Med. 6 (2019) 34.

[17] E. Baro, S. Degoul, R. Beuscart, E. Chazard, Toward a literature-driven definition of big data in healthcare, BioMed Res. Int. 2015 (2015).

[18] E. Saweros, Y.T. Song, Connecting heterogeneous electronic health record systems using Tangle, in: International Conference on Ubiquitous Information Management and Communication, Springer, Cham, 2019, pp. 858−869.

[19] C. Austin, F. Kusumoto, The application of big data in medicine: current implications and future directions, J. Intervent. Card. Electrophysiol. 47 (1) (2016) 51−59.

[20] A. Fiske, A. Buyx, B. Prainsack, Health information counselors: a new profession for the age of big data, Acad. Med. 94 (1) (2019) 37.

[21] P. Galetsi, K. Katsaliaki, S. Kumar, Values, challenges, and future directions of big data analytics in healthcare: a systematic review, Soc. Sci. Med. (2019) 112533.

[22] B. Williamson, Big Data in Education: The Digital Future of Learning, Policy, and Practice, SAGE, 2017.

[23] H.M. Krumholz, Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system, Health Aff. 33 (7) (2014) 1163−1170.

[24] P. Lacroix, Big data privacy and ethical challenges, in: Big Data, Big Challenges: A Healthcare Perspective, Springer, Cham, 2019, pp. 101−111.

[25] M. Ienca, E. Vayena, On the responsible use of digital data to tackle the COVID-19 pandemic, Nat. Med. (2020) 1−2.

[26] S. Chen, J. Yang, W. Yang, C. Wang, T. Bärnighausen, COVID-19 control in China during mass population movements at new year, Lancet 395 (10226) (2020) 764−766.

[27] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic Analysis of COVID-19 in China by Dynamical Modeling, arXiv, 2020 preprint arXiv:2002.06563.

[28] World Health Organization, Coronavirus Disease (COVID-2019) Situation Reports, 2020.

[29] CDC: Centers for Disease Control and Prevention Healthcare Professionals: Frequently Asked Questions and Answers, 2020 (Accessed 22 April 2020), www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html/.

[30] National Health Commission of the PRCCOVID-19 Epidemic Situation up to 24:00 on March 8th (2020). http://www.nhc.gov.cn/xcs/yqtb/202003/f2c83db9f 73d4be5 be0dc96af7 31813c.shtml/. (Accessed 9 March 2020).

[31] M. Marmot, R. Bell, The sustainable development goals and health equity, Epidemiology 29 (1) (2018) 5−7.

[32] C. Zhou, F. Su, T. Pei, A. Zhang, Y. Du, B. Luo, et al., COVID-19: Challenges to GIS With Big Data, Geography and Sustainability, 2020.

[33] U.G. Assembly, A Life of Dignity for All: Accelerating Progress Towards the Millennium Development Goals and Advancing the United Nations Development Agenda beyond 2015. A Report of the Secretary-General, United Nations, New York, NY, 2013.

[34] A.E. Ling, Y.S. Leo, Potential Presymptomatic Transmission of SARS-CoV-2, Zhejiang Province, China, 2020, 2020.

[35] D.S.W. Ting, L. Carin, V. Dzau, T.Y. Wong, Digital technology and COVID-19, Nat. Med. (2020) 1–3.

[36] H.N. Dai, H. Wang, G. Xu, J. Wan, M. Imran, Big data analytics for manufacturing internet of things: opportunities, challenges, and enabling technologies, Enterp. Inf. Syst. (2019) 1–25.

[37] I. Olaronke, O. Oluwaseun, Big data in healthcare: prospects, challenges, and resolutions, in: 2016 Future Technologies Conference (FTC), IEEE, 2016, pp. 1152–1157.

[38] G. Manogaran, C. Thota, D. Lopez, V. Vijayakumar, K.M. Abbas, R. Sundarsekar, Big data knowledge system in healthcare, in: Internet of Things and Big Data Technologies for Next-Generation Healthcare, Springer, Cham, 2017, pp. 133–157.

[39] K. Metaxiotis, Healthcare knowledge management, in: Encyclopedia of Knowledge Management, second ed., IGI Global, 2011, pp. 366–375.

[40] P. Halder, I. Pan, Role of big data analysis in healthcare sector: a survey, in: 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), IEEE, 2018, pp. 221–225.

[41] V. Tresp, J.M. Overhage, M. Bundschus, S. Rabizadeh, P.A. Fasching, S. Yu, Going digital: a survey on digitalization and large-scale data analytics in healthcare, Proc. IEEE 104 (11) (2016) 2180–2206.

[42] Z.S. Wong, J. Zhou, Q. Zhang, Artificial intelligence for infectious disease big data analytics, Infect. Dis. Health 24 (1) (2019) 44–48.

[43] A. Oussous, F.Z. Benjelloun, A.A. Lahcen, S. Belfkih, Big Data technologies: a survey, J. King Saud Univ. Comput. Inf. Sci. 30 (4) (2018) 431–448.

[44] S. Sagiroglu, D. Sinanc, Big data: a review, in: 2013 International Conference on Collaboration Technologies and Systems (CTS), IEEE, 2013, pp. 42–47.

[45] R.L. Villars, C.W. Olofson, M. Eastwood, Big Data: What it Is and Why You Should Care vol. 14, White Paper, IDC, 2011, pp. 1–14.

[46] K. Priyanka, N. Kulennavar, A survey on big data analytics in health care, Int. J. Comput. Sci. Inf. Technol. 5 (4) (2014) 5865–5868.

[47] C.S. Kruse, R. Goswamy, Y.J. Raval, S. Marawi, Challenges and opportunities of big data in health care: a systematic review, JMIR Med. Inform. 4 (4) (2016) e38.

[48] P.R. Ward, Improving access to, use of, and outcomes from public health programs: the importance of building and maintaining trust with patients/clients, Front. Public Health 5 (2017) 22.

[49] V.J. Munster, M. Koopmans, N. van Doremalen, D. van Riel, E. de Wit, A novel coronavirus emerging in China—key questions for impact assessment, N. Engl. J. Med. 382 (8) (2020) 692–694.

[50] C. Buckee, Improving epidemic surveillance and response: big data is dead, long live big data, Lancet Digit. Health (2020).

[51] D.B. George, W. Taylor, J. Shaman, C. Rivers, B. Paul, T. O'Toole, et al., Technology to advance infectious disease forecasting for outbreak management, Nat. Commun. 10 (1) (2019) 1–4.

[52] S. Lai, A. Farnham, N.W. Ruktanonchai, A.J. Tatem, Measuring mobility, disease connectivity, and individual risk: a review of using mobile phone data and mHealth for travel medicine, J. Trav. Med. 26 (3) (2019) taz019.

[53] C.B. Stone, A.R. Neely, M.L. Lengnick-Hall, Human resource management in the digital age: big data, HR analytics, and artificial intelligence, in: Management and Technological Challenges in the Digital Age, CRC Press, 2018, pp. 13–42.

[54] D.P. Acharjya, K. Ahmed, A survey on big data analytics: challenges, open research issues, and tools, Int. J. Adv. Comput. Sci. Appl. 7 (2) (2016) 511–518.

[55] A. Labrinidis, H.V. Jagadish, Challenges and opportunities with big data, Proc. VLDB Endow. 5 (12) (2012) 2032−2033.

[56] M.M. Najafabadi, F. Villanustre, T.M. Khoshgoftaar, N. Seliya, R. Wald, E. Muharemagic, Deep learning applications and challenges in big data analytics, J. Big Data 2 (1) (2015) 1.

[57] C.P. Chen, C.Y. Zhang, Data-intensive applications, challenges, techniques, and technologies: a survey on big data, Inf. Sci. 275 (2014) 314−347.

[58] G.D. Samaraweera, M.J. Chang, Security and privacy implications on database systems in big data era: a survey, IEEE Trans. Knowl. Data Eng. (2019).

[59] X. Wu, X. Zhu, G.Q. Wu, W. Ding, Data mining with big data, IEEE Trans. Knowl. Data Eng. 26 (1) (2013) 97−107.

[60] R. Gul, A. Ahsan, Big data and analytics: case study of good governance and government power, in: European Conference on Intangibles and Intellectual Capital, Academic Conferences International Limited, 2019, pp. 128−XI.

[61] M. Chen, S. Mao, Y. Zhang, V.C. Leung, Big Data: Related technologies, Challenges, and Future Prospects, 2014.

[62] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: overview, challenges, and the future, in: Classification in BioApps, Springer, Cham, 2018, pp. 323−350.

[63] W. Zang, P. Zhang, C. Zhou, L. Guo, Comparative study between incremental and ensemble learning on data streams: case study, J. Big Data 1 (1) (2014) 5.

[64] R.M. May, Uses and abuses of mathematics in biology, Science 303 (5659) (2004) 790−793.

[65] C. Viboud, K. Sun, R. Gaffey, M. Ajelli, L. Fumanelli, S. Merler, et al., The RAPIDD ebola forecasting challenge: synthesis and lessons learnt, Epidemics 22 (2018) 13−21.

[66] R. Kahn, A.S. Mahmud, A. Schroeder, L.H.A. Ramirez, J. Crowley, J. Chan, C.O. Buckee, Rapid forecasting of cholera risk in Mozambique: translational challenges and opportunities, Prehosp. Disaster Med. 34 (5) (2019) 557−562.

[67] G.L. Armstrong, D.R. MacCannell, J. Taylor, H.A. Carleton, E.B. Neuhaus, R.S. Bradbury, et al., Pathogen genomics in public health, N. Engl. J. Med. 381 (26) (2019) 2569−2580.

[68] R. Schiavo, M. May Leung, M. Brown, Communicating risk and promoting disease mitigation measures in epidemics and emerging disease settings, Pathog. Glob. Health 108 (2) (2014) 76−94.

[69] Y.A. de Montjoye, S. Gambs, V. Blondel, G. Canright, N. De Cordes, S. Deletaille, et al., On the privacy-conscientious use of mobile phone data, Sci. Data 5 (1) (2018) 1−6.

[70] E.L. Ray, N.G. Reich, Prediction of infectious disease epidemics via weighted density ensembles, PLoS Comput. Biol. 14 (2) (2018) e1005910.

[71] M. Cottle, W. Hoover, S. Kanwal, M. Kohn, T. Strome, N. Treister, Transforming Health Care through Big Data Strategies for Leveraging Big Data in the Health Care Industry, Institute for Health Technology Transformation, 2013. http://ihealthtran.com/big-data-in-healthcare.

[72] S. Chae, S. Kwon, D. Lee, Predicting infectious disease using deep learning and big data, Int. J. Environ. Res. Publ. Health 15 (8) (2018) 1596.

[73] S. Bansal, G. Chowell, L. Simonsen, A. Vespignani, C. Viboud, Big data for infectious disease surveillance and modeling, J. Infect. Dis. 214 (Suppl. 4) (2016) S375−S379.

[74] G. Chowell, P.W. Fenimore, M.A. Castillo-Garsow, C. Castillo-Chavez, SARS outbreaks in Ontario, Hong Kong, and Singapore: the role of diagnosis and isolation as a control mechanism, J. Theor. Biol. 224 (1) (2003) 1−8.

[75] A. Siu, Y.R. Wong, The economic impact of SARS: the case of Hong Kong, Asian Econ. Pap. 3 (1) (2004) 62−83.

[76] F. Fadel, M.D. Khalil El Karoui, B. Knebelmann, Spread of a novel influenza A (H1N1) virus via global airline transportation, Nat. Rev. Immunol. 8 (2008) 153−160.

[77] W.J. McKibbin, The Swine Flu Outbreak and its Global Economic Impact, Brookings, 2016, 2009.

[78] C. Yang, Q. Huang, Z. Li, K. Liu, F. Hu, Big data and cloud computing: innovation opportunities and challenges, Int. J. Digit. Earth 10 (1) (2017) 13−53.

[79] J.S. Hurwitz, A. Nugent, F. Halper, M. Kaufman, Big Data for Dummies, John Wiley & Sons, 2013.

[80] P. Mikalef, I.O. Pappas, J. Krogstie, M. Giannakos, Big data analytics capabilities: a systematic literature review and research agenda, Inf. Syst. E Bus. Manag. 16 (3) (2018) 547−578.

[81] S. Liu, S. Poccia, K.S. Candan, G. Chowell, M.L. Sapino, epiDMS: data management and analytics for decision-making from epidemic spread simulation ensembles, J. Infect. Dis. 214 (Suppl. 4) (2016) S427−S432.

[82] H.J. Murff, F. FitzHenry, M.E. Matheny, N. Gentry, K.L. Kotter, K. Crimin, et al., Automated identification of postoperative complications within an electronic medical record using natural language processing, JAMA 306 (8) (2011) 848−855.

[83] T.B. Murdoch, A.S. Detsky, The inevitable application of big data to health care, JAMA 309 (13) (2013) 1351−1352.

[84] B.Y. Reis, I.S. Kohane, K.D. Mandl, Longitudinal histories as predictors of future diagnoses of domestic abuse: modelling study, Br. Med. J. 339 (2009) b3677.

[85] J.L. Brown, The unasked question, J. Am. Med. Assoc. 308 (18) (2012) 1869−1870.

[86] P.B. Jensen, L.J. Jensen, S. Brunak, Mining electronic health records: towards better research applications and clinical care, Nat. Rev. Genet. 13 (6) (2012) 395−405.

[87] S.J. Mooney, D.J. Westreich, A.M. El-Sayed, Epidemiology in the era of big data, Epidemiology 26 (3) (2015) 390.

[88] S.V. Scarpino, G. Petri, On the predictability of infectious disease outbreaks, Nat. Commun. 10 (1) (2019) 1−8.

[89] C.M. Wheeler, W.C. Hunt, N.E. Joste, C.R. Key, W.G. Quint, P.E. Castle, Human papillomavirus genotype distributions: implications for vaccination and cancer screening in the United States, J. Natl. Cancer Inst. 101 (7) (2009) 475−487.

[90] J.J. Huang, COVID-19 and Applicable Law to Transnational Personal Data: Trends and Dynamics, Sydney Law School Research Paper, (20/23), 2020.

[91] B. Chen, Historical foundations of choice of law in fiduciary obligations, J. Priv. Int. Law 10 (2) (2014) 171−203.

[92] M. Douglas, Characterization of breach of confidence as a privacy tort in private international law, UNSWLJ 41 (2018) 490.

[93] D.E. Bloom, D. Cadarette, Infectious disease threats in the 21st century: strengthening the global response, Front. Immunol. 10 (2019) 549.

[94] C.I. Paules, R.W. Eisinger, H.D. Marston, A.S. Fauci, What recent history has taught us about responding to emerging infectious disease threats, Ann. Intern. Med. 167 (11) (2017) 805−811.

[95] S.S. Morse, Public health surveillance and infectious disease detection, Biosecur. Bioterrorism Biodefense Strat. Pract. Sci. 10 (1) (2012) 6−16.

[96] Z. Wu, J.M. McGoogan, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention, JAMA (2020).

[97] L. Bengtsson, J. Gaudart, X. Lu, S. Moore, E. Wetter, K. Sallah, et al., Using mobile phone data to predict the spatial spread of cholera, Sci. Rep. 5 (2015) 8923.

 [98] M. Bates, Tracking disease: digital epidemiology offers new promise in predicting outbreaks, IEEE Pulse 8 (1) (2017) 18−22.

 [99] J. Cinnamon, S.K. Jones, W.N. Adger, Evidence and future potential of mobile phone data for disease disaster management, Geoforum 75 (2016) 253−264.

[100] N. Mor, Resources for Primary Care Providers to Meet Patients Needs During the COVID-19 Epidemic, 2020.

[101] P. Mozur, R. Zhong, A. Krolik, In coronavirus fight, China gives citizens a color code, with red flags, New York Times 1 (2020).

[102] A.N. Mathur, Strategies for Solving Wicked Problems of True Uncertainty: Tackling Pandemics Like COVID-19 (Version: April 13, 2020) (No. WP 2020-04-03), Indian Institute of Management Ahmedabad, Research and Publication Department, 2020.

[103] C. Gaglione, I. Purificato, O.P. Rymkevich, COVID-19 and labour law: Italy, Ital. Lab. Law e-J. 13 (1S) (2020).

[104] G. Knieps, Internet of things, big data, and the economics of networked vehicles, Telecommun. Pol. 43 (2) (2019) 171−181.

[105] M. Rynge, K. Vahi, E. Deelman, A. Mandal, I. Baldin, O. Bhide, et al., Integrity protection for scientific workflow data: motivation and initial experiences, in: Proceedings of the Practice and Experience in Advanced Research Computing on Rising of the Machines (Learning), 2019, pp. 1−8.

[106] C. Karthikeyan, A. Benjamin, An exploratory study on business data integrity for effective business; a techno business leadership perspective, Int. J. Res. Soc. Sci. 9 (4) (2019) 167−201.

[107] M.O. Soeder, Privacy Challenges and Approaches to the Consent Dilemma, 2019. Available at: SSRN 3442612.

[108] C. Arora, Digital health fiduciaries: protecting user privacy when sharing health data, Ethics Inf. Technol. 21 (3) (2019) 181−196.

[109] D. Helbing, Societal, economic, ethical, and legal challenges of the digital revolution: from big data to deep learning, artificial intelligence, and manipulative technologies, in: Towards Digital Enlightenment, Springer, Cham, 2019, pp. 47−72.

[110] L. Dencik, A. Hintz, J. Cable, Towards data justice, Data Polit. 167 (2019).

[111] C.J. Wang, C.Y. Ng, R.H. Brook, Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing, JAMA (2020).

[112] S. Mohanty, R. Agarwal, S. Barik, A Survey on Big Data Analytics, 2013.

[113] P. Russom, Big data analytics, TDWI Best Pract. Rep. Fourth Quart. 19 (4) (2011) 1−34.

[114] H. Yang, E. Kundakcioglu, J. Li, T. Wu, J.R. Mitchell, A.K. Hara, et al., Healthcare intelligence: turning data into knowledge, IEEE Intell. Syst. 29 (3) (2014) 54−68.

[115] M. Joshi, D. Joshi, A survey of topological data analysis methods for big data in healthcare intelligence, Int. J. Appl. Eng. Res. 14 (2) (2019) 584−588.

[116] D.S. Brown, B. Drake, P.J. Fowler, J.K. Harris, K.J. Johnson, Leveraging big data analytics and informatics, Toward a Livable Life: A 21st Century Agenda for Social Work 337 (2019).

[117] T.H. Davenport, J. Dyché, Big Data in Big Companies Vol. 3, International Institute for Analytics, 2013.

[118] W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, Health Inf. Sci. Syst. 2 (1) (2014) 3.

[119] S. Schneeweiss, Learning from big health care data, N. Engl. J. Med. 370 (23) (2014) 2161−2163.

[120] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, et al., Mastering the game of go without human knowledge, Nature 550 (7676) (2017) 354−359.

[121] I.B.M. Watson, IBM Watson for Oncology, 2016.

[122] C. Burghard, Big data and analytics key to accountable care success, IDC Health Insights (2012) 1−9.

[123] J. Manyika, Big Data: The Next Frontier for Innovation, Competition, and Productivity, 2011. http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_ next_frontier_for_innovation.

[124] J.T. Wu, K. Leung, G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, Lancet 395 (10225) (2020) 689−697.

[125] M. Gilbert, G. Pullano, F. Pinotti, E. Valdano, C. Poletto, P.Y. Boëlle, et al., Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study, Lancet 395 (10227) (2020) 871−877.

[126] Q.V. Pham, D.C. Nguyen, W.J. Hwang, P.N. Pathirana, Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts, 2020.

[127] C.C. Lai, T.P. Shih, W.C. Ko, H.J. Tang, P.R. Hsueh, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges, Int. J. Antimicrob. Agents (2020) 105924.