Chapman & Hall/CRC Internet of Things

DATA SCIENCE FOR EFFECTIVE HEALTHCARE SYSTEMS

Edited by
Hari Singh
Ravindara Bhatt
Prateek Thakral
Dinesh Chander Verma

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

First edition published 2023
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

Library of Congress Cataloging-in-Publication Data
Names: Singh, Hari, 1957- editor.
Title: Data science for effective healthcare systems / edited by Hari
Singh, Ravindara Bhatt, Prateek Thakral, Dinesh Chander Verma.
Description: First edition. | Boca Raton : Chapman & Hall/CRC Press, [2023]
| Series: Internet of things: data-centric intelligent computing,
informatics, and communication | Includes bibliographical references and
index. | Identifiers: LCCN 2022003644 (print) | LCCN 2022003645 (ebook) |
ISBN 9781032105680 (hardback) | ISBN 9781032307503 (paperback) |
ISBN 9781003215981 (ebook)
Subjects: LCSH: Medicine—Data processing. | Medical informatics. |
Medicine—Information technology.
Classification: LCC R858 .D3779 2023 (print) | LCC R858 (ebook) | DDC
610.285—dc23/eng/20220401
LC record available at https://lccn.loc.gov/2022003644
LC ebook record available at https://lccn.loc.gov/2022003645

ISBN: 9781032105680 (hbk)
ISBN: 9781032307503 (pbk)
ISBN: 9781003215981 (ebk)

DOI: 10.1201/9781003215981

Typeset in Palatino
by codeMantra

Contents

Preface

This book primarily introduces the importance of data science in the field of healthcare. Data science has proven to be a gamechanger in the optimization of disease prevention, drug discovery, diagnosis, treatment, hospital operation, and post-care monitoring. Initially, the basic concepts, applications, and challenges related to data science will be covered to lay down a strong foundation for its importance in healthcare. A large volume of data is generated in healthcare in processes primarily from clinical trials, genetic information, electronic medical records (EMRs), billing, wearable data, care management databases, scientific articles, social media, and internet research. A global pandemic like COVID-19 has dramatically increased the need and the importance of data science and analytics.Data science in healthcare industry is mainly used for predictive analysis that uses artificial intelligence, machine learning, deep learning to predict possibility of an event occurring in the future based on known present variables and facts. The various available and emerging tools and techniques to manage this voluminous healthcare data results in improving patient outcomes and reduce healthcare costs. Various techniques such asregression and classification have been covered here for better insights into healthcare data analysis and its related advantages. This book also focuses on the importance of visual analytics to observe the medical history of a patient for further courses of action to cure the diseases. This book is also meant to enlighten the importance of data science in drug discovery, pandemic like COVID-19 needs fast simulation techniques to observe the reaction of drugs inthe body so that the drug can be approved by competent authorities and used to cure the disease.

In nutshell, the tools and techniques of data science have changed the way to manage, analyze, and leverage healthcare data for better and optimized prediction.The main motive of this book is to bring the research and findings together for clear directions for future work.

Mr Prateek Thakral is currently an Assistant Professor in the Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat Solan. He has earned a B.Tech in Information Technology and an M.Tech in Computer Engineering from Kurukshetra University, Kurukshetra in 2010 and 2013, respectively. Before joining JUIT, he was an Assistant Professor (on contract) at the National Institute of Technology Kurukshetra for four years. He has 8 years of teaching experience in total. He has published about 10 research papers as various reputed book chapters, journals and international conferences. He is currently serving as a member of the Computer Society of India (CSI).

Dr Dinesh Chander Verma holds a PhD(CSE), MTech(IT), MCA and MPhilin in Computer Science. He has more than 17 years of teaching experience that includes significant administrative, research, and industrial exposure. He has delivered many expert talks, chaired sessions, and reviewed a good number of research articles. His areas of interest include Machine Learning, MANET, IoT, and Data Science. He is a Life Member of the Indian Society for Information Theory and Applications (ISITA). He has published several research articles, book chapters in Scopus/UGC Care/ESCI indexed journals. He has authored two books for UG courses and published one patent too. He has conducted several conferences/seminars/FDPs on recent emerging technologies and has mentored many industrial live projects on emerging technologies.

1.2.2 Patient Sentiment and Behavior Data

Mobility sensor data includes sensor data or streaming data from home monitoring, medical monitoring, telehealth, smart devices, and sensor-based wireless. Biometric data includes fingerprints, handwriting, and iris scans.

1.2.3 Clinical Information and Notes

Eighty percent of health information is unstructured, consisting of photos, papers, transcribed notes, and clinical notes. These semi-structured to unstructured clinical records (patient discharge summaries, medical photographs, diagnostic testing reports, etc.) and papers constitute novel data sources.

1.2.4 Web- and Social Networking–Based Data

Google and other search engines, Internet consumer use, and social networking sites all contribute to web-based statistics (Twitter, LinkedIn, Facebook, blog, smartphones, and health plan websites).

1.2.5 Genomic Data

Genomic data represents a large volume of new gene-sequencing information. Genotyping, gene expression, and DNA sequencing are all examples of genomic data.

1.3 Big Data 5 V's in Healthcare

Gartner identified eight important sources of big data in health in 2016 [5]. The four V's Veracity, Volume, Velocity, and Variety have been coined by Ernst & Young and others to describe it. Similarly, McKinsey also identified five "rights" that it may provide: "right provider", "right living", "right care", "right value", and "right innovation" [6].

1.3.1 Volume

Terabytes and petabytes of data are used by healthcare systems. Personal information, 3D imaging, genomics, personal medical records, radiological images, and biometric sensor readings are all stored in these systems [7]. This complicated data structure may now be managed and analyzed by healthcare systems. Storage, manipulation, and use of such complicated data are now possible due to the advent of cloud computing. According to KPMG (Klynveld Peat Marwick Goerdeler) report, healthcare data had surpassed 150 exabytes in 2013 and is growing at a speedy rate of 1.2–2.4 exabytes each year [8].

1.3.2 Velocity

The underlying explanation for data's exponential increase is velocity [9], which states how quickly data is generated. Data is being generated at an ever-increasing rate. Due to the amount and diversity of data acquired, the speed with which unstructured and structured data is generated forces a conclusion based on its outcome.

1.3.3 Veracity

Data veracity refers to the degree of certainty that a data interpretation is consistent. Varied data sources have different levels of data reliability and dependability [10]. Unsupervised machine learning algorithms, however, are employed in healthcare to make choices that data may be useless or ambiguous [11]. The purpose of healthcare analytics is to extract useful insights from this data so that patients can be treated effectively and the best decisions can be made.

1.3.4 Variety

It refers to the data's format, whether it's unstructured or structured, medical images, video, audio, text, or sensor data. Clinical data is structured data that must be collected, stored, and analyzed using specialized equipment. Structured data accounts for only 5% to 10% of overall healthcare data. Images, audios, videos, e-mails, and healthcare data, such as prescriptions, hospital medical reports, radiographic films, and physician's observations, are examples of unstructured or semi-structured data [12].

1.3.5 Value

This V, unlike the other 4Vs, is too unique since it reflects the anticipated results of big data processing. We're continuously looking for new ways to collect and extract the true value from the vast data. Investment should be the one to store data as the quality of the governance strategy and approach determines the value of data. Another important consideration is that certain data has a different risk value at the time of collection, but that risk can change with time [13].

Five more key traits have been identified by researchers, totaling 10V's of big data. Variability, Validity, Vulnerability, Volatility, and Visualization are the five extra traits. Healthcare data is separated into structured and unstructured data, which includes Electronic Medical Records (EMR) reports, medical photographs, and so on [14]. A large amount of data helps improve the worth of healthcare by utilizing creative analyses. The huge volume of data associated to healthcare can be analyzed by distributing the process using cloud centers and big data. Not only have big data increased the size of data but it has also increased the value that can be derived from it. To put it in another way, big data has shifted the focus of Business Intelligence away from reporting and decision-making toward prediction. Understanding innovative diseases and therapies, predicting results early, taking instantaneous decisions, improving health, boosting treatment, lowering expenses, and improving healthcare quality and worth are all instances of a value in healthcare.

1.4 Big Data Analysis in Healthcare Industry

Big data analysis can change the healthcare scenario. The ability to use technological equipment by healthcare professionals and make decisions about their clinical and other data streams has been changed for better comprehend. The five processes that make up big data healthcare analytics include data acquisition, data storage, data management, data

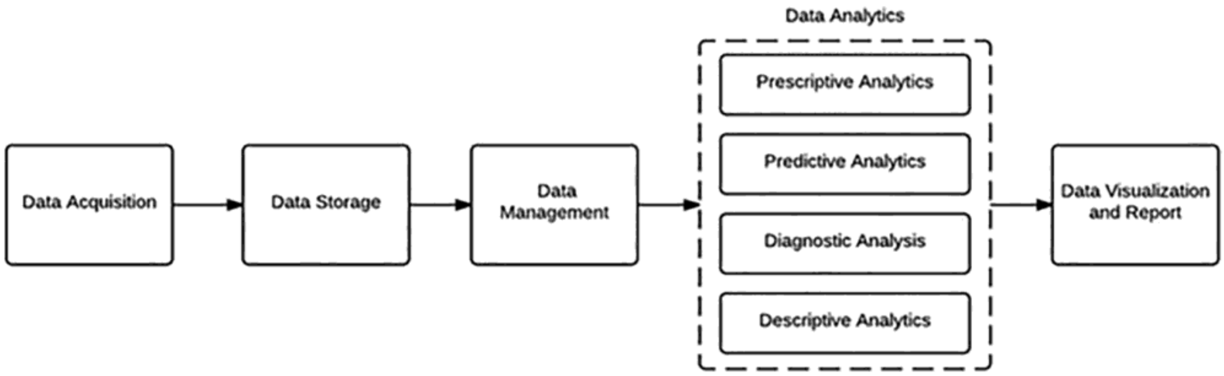

FIGURE 1.2 The process of big data analytics in healthcare [32].

analytics, and data visualization and reporting. Figure 1.2 depicts the process of big data analysis in healthcare management.

1.4.1 Data Acquisition

The type of data can be classified into unstructured, structured, or semi-structured, and it comes from both primary such as clinical decision support systems, electronic health records (EHRs), and Computerized Provider Order Entry (CPOE) and secondary resources such as laboratories, pharmacies, Health Maintenance Organization (HMOs), government sources, and insurance companies [15]. In healthcare, EHRs, social media, image processing, web-based data, and smartphones, among other things, are the most significant sources of big data.

1.4.2 Data Storage

In this epoch of technology, storage is crucial. Since the amount of data in the healthcare industry grows, we need a storage platform that is both efficient and huge. With such a big number of users, cloud computing is the most promising technology. Clouds offer the scalability and capabilities needed to obtain access to data, boost awareness, speed up the development of scalable analytics solutions, and generate income. It reduces the necessity of maintaining expensive computing gear and software.

1.4.3 Data Management

In healthcare, it encompasses structuring, cleansing, mining, and controlling the data. It also contains a mechanism for determining whether any row or omitted data exist. That information must be handled in a proper way. It aids in the risk assessment of patients and the creation of a customized discharge plan. Apache Ambari and HCatalog are the two important data management solutions. The process of obtaining files or important information from big healthcare databases is known as data retrieval. In healthcare, big data analysis usually includes information recovery and data mining [16].

1.4.4 Data Analytics

Data analytics is the procedure of converting raw data into information. The four forms of big data analytics applied in healthcare are descriptive, diagnostic, predictive, and prescriptive analytics [17].

The descriptive analysis offers specific information on each attribute such as features, number of features, and the size of the data set.

Predictive analysis aids in the "forecasting" of future events based on current data. It deciphers data insights and provides useful information to businesses in the form of recommendations. It also provides probabilistic estimates of future outcomes [18].

Prescriptive analytics makes use of predictive data outcomes to allow persons to "prescribe/determine" numerous activities to take and steer them toward a result. Before making a decision, it aims to assess the impact of future decisions and provide advice on alternative outcomes.

1.4.5 Data Visualization

Data visualization is the process of displaying analytic conclusions from healthcare data in a graphical layout to make better decisions. It can be used to interpret data patterns and associations.

1.5 Big Data Analytics Tools in Healthcare

Big data analytics is used to process the unprocessed data. The data, which is still unprocessed, is accessed, retrieved, and processed using service-oriented architecture. Another technique is data warehousing, which involves gathering data from a variety of sources and preparing it for processing, even if the data isn't available in real time. From several sources, data is processed and prepared. Although the data is unstructured or structured, the big data analytics platform can accept a variety of data formats [19]. Some of the analytics tools are listed below (Table 1.1).

1.6 Applications of Big Data in Healthcare

1.6.1 Data Analytics in COVID-19

Big data technology can save a lot of data on people who have been infected with COVID-19. It assists in gaining a complete understanding of the nature of the virus. In future, the knowledge acquired can be used to teach new preventive actions. This technology is used to save data from COVID-19 affected instances of all types (recovered, infected, and expired). This information can be used to track down cases and provide resources for better public health protection. Patient physiology, patient-reported travel, patient location, comorbidity, proximity, and existing symptoms are just a few examples of digital data modalities that can be used to provide insights at the community and population levels [24].

1.6.2 Hadoop-Based Applications

Because most of the data is unstructured, extracting relevant information about clinical operations, patient care, and research is a huge problem for the business. The Hadoop can

TABLE 1.1
Various Big data Platforms in Healthcare [20]

Platform/ Technology	Description
MapReduce	It is a programming archetype for forming Big Data applications that execute in parallel across multiple nodes. It's a method for analyzing big amounts of complex data [21].
Hadoop	Apache Hadoop is a distributed platform to deal with huge datasets through clusters of computers ranging in size from one to thousands [22].
Pig and Pig Latin	To work with a wide range of data types, Pig's programming language is used. The PigLatin programming language and its runtime version, which executes PigLatin code, are the two main modules.
Hive	Apache Hive is a query and analysis tool that runs on top of the Hadoop databases and file systems. Hive is an SQL-like Hadoop database and file system querying tool [23].
Cassandra	It is a standalone database system. It's a complex system that's meant to deal with massive volumes of data spread over several utility servers. Also, it offers continuous service and eliminates the possibility of a single point of failure.
Jaql	It is a query language for analyzing bulky data collections. It transforms "high-level queries" into "low-level queries" for parallel execution of MapReduce jobs.
Lucene	Lucene has been included in several open-source projects and is widely used for text analytics and searches. Its scope comprises the library and full-text indexing search within a Java program.
Mahout	Mahout is an open-source Apache project aimed at developing scalable machine learning models for Hadoop-based big data analytics.

assist the healthcare sector in managing this massive volume of data. In the healthcare sector, following are some of the applications of Hadoop ecosystem:

- **Genomics and Cancer Treatment:** To struggle against cancer, it is important to organize massive amounts of data efficiently. Individual genetics has an impact on cancer mutation patterns and responses, which helps to explain why some tumors are incurable. When recognizing cancer patterns, it is discovered that it is vital to provide customized therapy for certain malignancies based on patient's genetic composition. For mapping three billion DNA base pairs, researchers can use Hadoop's MapReduce technology to discover the best cancer therapy for individual patients.
- **Fraud Prevention and Detection:** In the first few years, health-based insurance companies used a variety of approaches to detect fraud and develop methods to prevent it. Companies utilize Hadoop to identify fraudsters by using data from prior health claims, voice recordings, earnings, and demographics to create apps based on a prediction model. By combining authentic medical claim bills, real-time Hadoop-based health apps, voice data recordings, weather forecasting data, and other data sources, Hadoop NoSQL database can assist in avoiding medical claim fraud at an early stage.
- **Network in Hospital:** Hadoop NoSQL database is used by various hospitals to accumulate and manage massive volumes of data from a variety of origins linked to patient care, payroll, and finances allowing them to detect risky patients along with deducting daily costs.

- **Patient's Vital Monitoring:** Using bigdata technologies, hospital employees around the world can connect their job output. Several hospitals in many parts of the world use Hadoop-based components in the Hadoop Distributed File System, such as the Impala, HBase, Hive, Spark, and Flume frameworks, to convert massive amounts of unstructured data generated by sensors that take patient vital signs, heartbeats per minute, blood pressure, blood sugar level, and respiratory rate into structured data. Without Hadoop, these healthcare professionals would be unable to analyze the unstructured data supplied by patient healthcare systems.

1.6.3 Big Data in Public Health and Behavior Research

In public behavior and health, big data concentrates on biological data collected by portable equipment, for instance vitals, electrocardiograms, wearable devices, contagion, and daily health records.

- **Electrocardiogram:** Traditional electrocardiographic tracing investigations are essential for understanding the reasons and mechanisms of arrhythmias and other related disorders; however, the data does not allow for the prediction of the initiation of cardiac damage before it occurs. As a result, new metrics must be established for use in various preclinical cardiovascular diagnoses. The introduction of Big Data which collects higher complexity information characterized by volume, variety, and velocity is making a significant contribution to the control, contrast, and management of big data sets [24].
- **Omics Data:** The term "omics" refers to large records in the molecular and organic sciences (e.g. macrobiotics, metabolomics, proteomics, genomics, etc.). The purpose of using big data is to recognize ill approaches and improve the precision of medical therapy. The advancement of genomics, proteomics, metabolomics, and other kinds of Omics knowledge in earlier eras has resulted in a massive amount of data related to molecular biology [25].

1.6.4 Source of Valuable Data

In the healthcare industry, technological advancements in the utilization of gadgets have brought with them the new possibilities for generating and collecting data. Most data repositories in healthcare, such as those in other industries and organizations, combine data from a diversity of sources, including social media, EHRs, and medical devices [26].

- **EHRs:** In a more organized sense, data in a healthcare facility is stored in EHRs to offer a multifaceted view of patients' records. The EHR is a database of electronic healthcare data that generally pertains to individuals, medical records, and administrative information. Sub-schemes such as admission, discharge, and transfer of patients' schema, recurring engagements and planning schema, the technique for inputting prescription schema, and notes on routine medical checks are all included in a traditional EHR framework [27].
- **Medical Supplies and Equipment:** Medical devices and sensors such as pulse oximeters, blood pressure monitors, glucose monitors, and other sensors generate massive amounts of data that might reveal important information about a patient's physical condition. One of the motivating factors for embracing big data

in healthcare is the explosion of the Internet of Things and its ability to provide speedy access to medical needs. Body Sensor Networks and their continued application in healthcare will offer healthcare providers the ability to scan critical parameters and, as a result, precisely predict imminent medical hazards such as epidemics and pandemics [28].

- **Social Media:** In the healthcare industry, social media is also a valuable source of information. Patient behavior data and sentiment data on patient recovery are collected. Furthermore, social media posts such as blogs, Twitter feeds, Facebook status updates, and web pages can expose and provide an indicator of a person's health, mood, and state of mind, which is important to health professionals.

1.6.5 Big Data in Medical Experiment

Molecular biology is concerned with the communication and supervision of biological activity inside cells, for example, communications between proteins, RNA, and DNA. It is an important aspect of both biological and medical research. The study of proteins and genes has a secure interaction with the features of genetics and biochemistry. Macromolecule blotting and probing, polymerase chain reaction, molecular cloning, microarrays, and other techniques are used in molecular biology. A person's body comprises organs, tissues, and cells, along with cross-sectional images of tissues and organs, which is used to display the structure of the human body for considering medical measures. Biological laboratory specimens, such as human body data sets, are collected from human bodies and stored in biorepositories. Before a new medicine, vaccine, or piece of medical equipment is used, clinical trials must be performed. A medical experiment is a form of test or study used in clinical or medical research to determine the efficacy of a new medical therapy on humans [29].

1.6.6 Medical Research Using Big Data

Research articles and organized knowledge are currently produced at a rapid rate as the medical/clinical area has developed. There are also a lot of obsolete materials under the clinical/medical section. The study creates a substantial involvement to the subject of big data in healthcare.

1.7 Challenges with Healthcare Data Management

1.7.1 Challenges Associated with Manpower

- **Human Interaction:** Despite technological advancements, there are still some areas where a human could achieve better outcomes than a machine. As a result, humans must be involved at every level of big data analytics. To offer meaningful solutions, experts from various fields (those who understand the problem and data storage) must interact with big data tools. The scarcity of skilled and knowledgeable individuals is once again an impediment to the implementation of big data analytics [30].

- **Keeping Big Data Experts:** Even though corporations may discover the greatest big data professionals with healthcare experience, hiring them is tough and expensive. Given the severe competition, retaining highly skilled data scientists and analysts is also difficult [17,31].
- **Deficiency of Talent:** When businesses choose to use big data technology, a need for qualified data analysts who can understand healthcare data arises [15]. Data scientists and analysts with experience in the healthcare field and those who can apply the right tools to the data, generate results, and analyze them to give actionable insights are in high demand. However, only a few people have the essential skills and abilities to apply analytics to healthcare [16].

1.7.2 Challenges in Data and Process

- **Data Integrity:** The authenticity and integrity of healthcare data is an issue when combining multiple forms from various origins [1,8,6]. Data from X-rays, physiological pathology reports, bedside monitors, and recordings from various tests may be used to compile patient data. Before analyzing this large amount of data, an information extraction method is required, which extracts the necessary data and delivers in an acceptable format to analyze the data. It's a technical difficulty to guarantee the accuracy and thoroughness of this process. Furthermore, sensor malfunctions and human errors may result in inadequate and incorrect data, which could result in severe health concerns and unpleasant occurrences for patients [11]. Furthermore, errors caused by poor data quality are to blame for increased expenses for healthcare companies. As a result, data cleaning and normalization, which involves removing noise and unnecessary data, is difficult in making effective utilization of data.
- **Big Data Privacy and Security:** Patient record is stored in a variety of places and can be accessed through a variety of endpoints, including medical records, pathology reports, and insurance claim information. To track patient health and enable the transmission of patient data across various parties, a variety of applications are utilized, each with varying levels of security. Out-of-date software can jeopardize the security of health data, leaving it vulnerable to cyber-attacks. Personal information, such as a person's name, occupation, and income, can cost healthcare providers a lot of money if it is accessed and misused [30]. Furthermore, disclosing private health data raises privacy concerns for patients [8,32]. Because of the unconstrained actions of data breaches, managing large data security and privacy remains a technical concern, even with the technical controls and measures in place.
- **Interoperability:** The issue of interoperability occurs when data is generated by multiple healthcare equipment and devices [1,12,5]. Because these devices use different platforms and software, the data they generate is in different formats. To make full use of this data, the devices must be able to interact and exchange data in a common format that is interoperable with other devices.
- **Data Storage and Linkage:** Traditional IT equipment in organizations were not capable to handle such vast amounts of data as the volume of healthcare data rises exponentially [8]. Furthermore, data redundancy difficulties arise as a result of data storage among multiple departments within an organization [9]. Analyzing

data that is fragmented and incomplete becomes tough. Even the most advanced algorithms fail to cope with disintegrating data. The huge volume and a variety of data from various sources make it difficult to combine and aggregate these sources into repositories [10].

- **Data Collection:** Healthcare data is multidimensional and highly segmented, and it comes from a variety of sources [1]. Due to a lack of synchronization across several data sources, there may be gaps in the information provided. Furthermore, the organization of the data available from these sources varies greatly. Healthcare data, however, contains organized patient's demographic information, physician observations, diagnostic metaphors such as MRI and CT scans, and visuals that are not in a structured format. As a result, bringing together disparate data silos from many sources and transforming them into a uniform format suitable for storage in the system is difficult [14]. Furthermore, continuous data collection is difficult due to real-time or near-real-time data production [8].

1.7.3 Overall Organizational Challenges

- **Actionable Insights at the Right Time:** With correct tools and analytics solutions, big data could manage huge information or data than intuition. Actionable intuitions are much important than simple responses to the questions posed because of the activities they stimulate. As a result, the priority is to gain timely insights that inspire action. Given the importance of timely healthcare choices, another problem to consider is the development of triable information in realtime [33].
- **The Discrepancy in Technology:** Despite the increased interest in healthcare digitization, most businesses continue to rely on outdated technologies. It's a challenge to replace a historical data storage and management system with cutting-edge technology. Due to the large volume of data held in physical records in healthcare, digitalization would necessitate a significant amount of effort, human resources, and time. Another difficult task is to bridge the technological divide [34].
- **Unknown Purpose:** Lack of clarity in the aim of big data analytics to an organization has also become a hurdle in its advancement. Organizations use big data technologies as a source of competitive advantage without clearly defining their business goals. Force-fitting new technologies will result in a lack of direction. It's once again difficult to define the business goals for using big data analytics [35].
- **Identifying Useful Information and Tools:** The next step is to find and store relevant data once firms become familiar with commercial use cases for big data technology. As a result, the appropriate tools for working with such data must be discovered. It's also difficult to figure out what tools and solutions are available.
- **Sharing:** Data sharing is another stumbling block to the successful deployment of big data analytics. Organizations must exchange data with other healthcare organizations to reap the benefits of big data technology for the community. Although more openness and data availability would allow for faster judgments, firms are wary of sharing data due to competitive pressures. This obstructs the effective use of big data analytics in healthcare.
- **Economic:** The most critical initial difficulty for healthcare businesses adopting a data-driven culture is managing the expense of data warehouses and infrastructure required to hold massive amounts of data. In reality, the required

computational resources also increase initial investment for performing big data analysis. Hence, for small and medium organizations, it would be costly, while major businesses are wary of making such an investment without knowing what they would get in return.

1.8 Conclusion

Big data analytics, which makes use of a plethora of heterogeneous, unstructured, and structured data sources, has a critical responsibility in the future of healthcare. A variety of analytics is already being used to aid the healthcare workers and patients. Genetic data processing, physiological signal processing, and medical image analysis are the major areas of interest. The exponential growth in the volume of healthcare data compels computer scientists to devise novel ways to process such a massive amount of data in a manageable amount of time. Big data analytics brings heterogeneous data together from various domains including medical informatics, medical imaging, sensor informatics, computational biomedicine, bioinformatics, and health informatics. Furthermore, the properties of big data give an excellent foundation in healthcare and medicine for developing applications to use promising software platforms.

Data scientists face hurdles in integrating and implementing a large volume of medical data acquired across multiple platforms. As a result, it is suggested that a healthcare revolution is required to bring analytics, health informatics, and bioinformatics together to encourage tailored and additional effective therapies.

References

1. Mukhtar, W. F., and Abuelyaman, E. S. 2020. *Opportunities and Challenges of Big Data in Healthcare. Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications*, 1989–2001. IGI Global. http://doi:10.4018/978-1-7998-1204-3.ch099
2. Russom, 2011. Big data analytics. TDWI best practices report, fourth quarter 19(4): 1–34.
3. Hong, Liang, Luo, Mengqi, Wang, Ruixue, Lu, Peixin, Lu, Wei, and Lu, Long. 2018. Big data in health care: Applications and challenges. *Data and Information Management*, 2(3): 175–197. https://doi.org/10.2478/dim–2018–0014
4. Mousannif, Hajar, Khalil, Ismail, and Kotsis, Gabriele. (2013). The cloud is not 'there', we are the cloud! *International Journal of Web and Grid Services*, 9: 1–17. http://doi:10.1504/IJWGS.2013.052854.
5. Sini, E. 2016. How big data is changing healthcare. Humanitas Hospital Italy.
6. The big-data revolution in US health care: accelerating value and innovation. 2013. Available from: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care (accessed August 2021).
7. Liu, W. and Park, E.K. 2014. Big data as an e-health service: 982–988. http://doi:10.1109/ICCNC.2014.6785471.
8. KGMP report, https://assets.kpmg/content/dam/kpmg/pdf/ 2016/01/healthcare-insights-dec-2015-2.pdf (accessed July 2021).
9. Raghupathi, W., and Raghupathi, V. 2014. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2: 3. https://doi.org/10.1186/2047-2501-2-3

10. Mehmood, R. and Graham, G. 2015. Big data logistics: A health- care transport capacity sharing model, *Procedia Computer Science*, 64: 1107–1114.
11. Herland, Matthew, Khoshgoftaar, Taghi, and Wald, Randall 2014. A review of data mining using big data in health informatics. *Journal of Big Data*, 1: 2. 10.1186/2196–1115–1–2.
12. MAPR, Healthcare and life science use cases. 2018. https://mapr.com/solutions/industry/healthcare-and-lifescience- use-cases/ (accessed July 2021).
13. Mallappallil, Mary et al. 2020. A review of big data and medical research. *SAGE Open Medicine*. https://doi.org/10.1177/2050312120934839.
14. Shvachko, K., Kuang, H., Radia, S., and Chansler, R. 2010. The Hadoop Distributed File System. *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*: 1–10.
15. Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., and Taha, K. 2015. Efficient machine learning for big data: A review. *Big Data Result*, 2: 87–93. https://doi.org/10.1016/j.Bdr.2015.04.001.
16. Andrzej Chluski, L. Z. 2015. The application of big data in the management of healthcare organizations: A review of selected practical solutions. *Bus. Informatics*, 1: 9–18. https://doi.org/10.15611/ie.2015.1.01.
17. Ahmed, I., Ahmad, M., Jeon, G., and Piccialli, F. 2021. A framework for pandemic prediction using big data analytics. *Big Data Research*, 25, 100190. https://doi.org/10.1016/j.bdr.2021.100190.
18. India Mag. 2018. https://analyticsindiamag.com/organizations-continue-to-face-challenges-with-big-data-lets-deep-dive/ (accessed August 2021).
19. PEX Process Excellence Network. 6 ways pharmaceutical companies are using big data to drive innovation & value. https://www.processexcellencenetwork.com/tools-technologies/whitepapers/6-ways-pharmaceutical-companies-are-using-big-data (accessed June 2021).
20. Haleem, A., Javaid, M., Khan, I. H., and Vaishya, R. 2020. Significant applications of big data in COVID-19 pandemic. *Indian Journal of Orthopedics*, 54(4): 526–528.
21. Song, Y., Wang, H., Li, J., and Gao, H. 2016. MapReduce for Big Data Analysis: Benefits, Limitations, and Extensions. In: Che, W. et al. (eds) *Social Computing. ICYCSEE. Communications in Computer and Information Science*, 623. Springer. https://doi.org/10.1007/978-981-10-2053-7_40
22. Abualkishik, Abedallah. 2019. Hadoop and big data challenges. *Journal of Theoretical and Applied Information Technology*, 97: 3488–3500.
23. Kan, Z., Cheng, X., Kim, S. H., Jin, Y., and Kan, Z. 2018. Apache hive-based big data analysis of health care data. *International Journal Pure and Application Mathematics*, 119(18): 237–259.
24. Cipresso, C., Rundo, F., Conoci, S., and Parenti, R. 2019. Big data in preclinical ECG alterations research. *Biomedical Journal*, 13(2): 9895 -9896, doi: DOI: 10.26717/BJSTR.2019.13.002384.
25. van Allen, E. M., Wagle, N., and Levy, M. A. 2013. Clinical analysis and interpretation of cancer genome data. *Journal of Clinical Oncology*, 31: 1825–1833.
26. Asare-Frempong, Justice and Jayabalan, Manoj 2017. Exploring the impact of big data in healthcare and techniques in preserving patients' privacy. *International Journal of Computer Science and Network Security*, 17: 143–149.
27. Saranga Jayawardena, D. B. A. 2013. A systematic literature review of security, privacy and confidentiality of patient information in electronic health information systems. *Sri Lanka Journal Bio-Medical Informatics*, 4(2): 25–31.
28. KupwadePatil, H. and Seshadri, R. 2014. Big data security and privacy issues in healthcare. *IEEE International Congress on Big Data*: 762–765.
29. Vesna, V. The Visible Human Project: Informatic bodies and posthuman medicine. 2000. 262–263. https://doi.org/10.1007/BF01205456
30. Labrinidis, Alexandros and Jagadish, H.V. 2012. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5: 2032–2033. https://doi.org/10.14778/2367502.2367572.
31. Dhayne, H., Haque, R., Kilany, R., and Taher, Y. 2019. In search of big medical data integration solutions - A comprehensive survey. *IEEE Access*. 7: 91265–91290. https://doi.org/10.1109/ACCESS.2019.2927491.
32. Dash, S., Shakyawar, S.K., Sharma, M. et al. 2019. Big data in healthcare: management, analysis, and future prospects. *Journal of Big Data*, 6: 54. https://doi.org/10.1186/s40537-019-0217-0.

33. Harvey, C. Big data challenges. *Datamation*. 2017. https://www.datamation.com/big-data/big-data-challenges.html (accessed in August 2021).
34. Chander, D., Singh, H., and Gupta, A. K. 2020. A study of big data processing for sentiments analysis. *Large-Scale Data Streaming, Processing, and Blockchain Security*. IGI Global. https://doi.org/10.4018/978-1-7998-3444-1.ch001.
35. Singh, Hari and Bawa, Seema. 2020. Predicting Covid-19 statistics using machine learning regression models Li-MuLi-Poly. Multimedia Systems. https://doi.org/10.1007/s00530-021-007982.