How to Cite:
Alhamoud, M. Y., Sa'ad Ziyad, K. S., & Alharbi, M. S. (2020). The role and challenges of big data in healthcare informatics and analytics: Review. *International Journal of Health Sciences*, *4*(S1), 398–409. https://doi.org/10.53730/ijhs.v4nS1.15333

The role and challenges of big data in healthcare informatics and analytics: Review

Mohammad Yahya Alhamoud
Health Information Technician, National Guard Health Affairs

Khalid Saeed Sa'ad Ziyad
Health Information Technician, National Guard Health Affairs

Murad Salman Alharbi
Health Information Technician, National Guard Health Affairs

***Abstract*---**It appears that data-related research, and in particular big data research, is ongoing and producing results faster than we can keep up with. This has challenged us to prepare an updated version of this survey that will cover these fields in the future. It appears that big data and healthcare informatics applications need more research targeted towards them. This research needs to be directed in terms of both developing the most appropriate technology that is adequate for streamlining the new healthcare practices in various regions, as well as in comparison to the current technologies and practices, and targeting research towards ways of coordinating healthcare professionals and hospital financial departments. Collaborative and interdisciplinary research among various stakeholders and professionals is also critical in overcoming the challenges related to big data and healthcare informatics. Moreover, besides the new big data technologies and practices developed alongside the fourth industrial revolution, future evolution needs to be supported by the development of appropriate regulatory frameworks. These may also embrace existing ethical considerations and constraints, particularly those related to making data available for researchers. Finally, the importance of having alternative and continued training for healthcare professionals related to data ethics and legal considerations is a must.

***Keywords*---**healthcare informatics, healthcare professionals, particular big data research.

International Journal of Health Sciences E-ISSN 2550-696X © 2020.
***Corresponding email**: Shsh-5522@hotmail.com*
Manuscript submitted: 27 Jan 2020, Manuscript revised: 09 Feb 2020, Accepted for publication: 18 March 2020

1. Introduction to Big Data in Healthcare Informatics

We live in the digital age where almost each one of us has internet. It has made the flow of information easy and quick. In addition to this, the advancements over the last two decades have revolutionized data capturing, storage, and retrieval. Terabytes of data from millions of sources are produced every day. Unfortunately, serious research on this data could not be started even after its inception because of the unavailability of required tools and computational resources, and that is why it is often referred to as 'big data.' These knowledge warehouses have the potential to transform the lives of people through personalized medicine and diagnostics, improved healthcare decision-making, and national epidemic surveillance to create policies based on real facts. The task of managing data resources with such sheer volumes, varieties, and complexities is tedious. Little meaning can be derived until it is processed, compiled, and presented in a consumable manner. Additionally, this data is not just huge but flows rapidly from various sources including electronic health records, clinical data, pathology reports, discharge summaries, lab reports, diagnostic images generated by various medical imaging equipment, transcriptomics and proteomics studies, and genomic data with the potential of 1MB data per person from hospitals, physicians, social media, health statistics, government, research, various consortia, pharmaceutical companies, and mobile apps.

1.1. Definition and Scope of Big Data

The term Big Data has no single or universally accepted definition. It is defined and understood differently from myriad perspectives and depending on the varied analytical and technological capacities and capabilities as well as user awareness. In the context considered for this paper, Big Data is conceptualized, operationalized, and defined as a large volume of complex, fast-growing, and diverse types of data being created at an ever-increasing rate and covering numerous dimensions, of high veracity and data in both structured and unstructured formats, for which sensible and insightful analysis will lead to efficient and better decision-making that enhances patient care and supports better organizational efficiency and effectiveness. In healthcare, data are depicted as significantly large, complex, integrated, de-identified, or real-time medical records, demographic, administrative, and financial data, large volumes of human genomics data, mHealth data, health social media data, and voluminous levels of operational, clinical logistics, and clinical management and operative data cascading down directly or indirectly from the IT and the IoT, among others.

Big Data in healthcare can take different forms, including some or all of the following: clinical data such as patients' medical history, medication, diagnosis, treatment, surgery, anomalies, and side effects; diagnostic lab tests; and administrative data such as patients' demography. Usually, in addition to such monumental data, most health organizations would acquire, collate, and integrate other forms of clinical, operational, logistics, financial, and patient-generated electronic medical records, all of which may come either in structured, semi-structured, or unstructured formats, robust in volume and velocity. Most of these data usually require that the information systems of these healthcare organizations be connected to wide data and network resources. A distinguishing feature of the digital-age evolution of healthcare data sets lies in the expectation of these diverse data and information resources and the data extracted from them being combined and analyzed together to broaden, deepen, or enhance patient care, disease prevention, health promotion, and explore opportunities for healthcare operational and administrative efficiency to the betterment of patients, healthcare organizations, and the wider community. However, the use of multiple data types introduces high-level complexities in data integration, data fusion, data cleaning, and data analysis.

1.2. Importance of Big Data in Healthcare

Subsection making an argument for the positive role of big data in healthcare
In a report, evidence that carefully harnessed big data enabled companies and public services to outcompete those firms still relying on traditional problem-solving approaches. This compelling evidence led to an expectation that healthcare can be transformed into a data-driven field with substantial improvements in outcomes and costs. The traditional approach of statistical sampling used for problem-solving in health informatics has little, if any, value for solving operational and policy issues in healthcare's big data-infested environment. (Avula2020)(Subrahmanya et al. 2020)

These frameworks helped create a market for personal health tracking devices such as wearables, emissivity sensors, exhalometry sensors, mobile apps, and social media monitoring. They are used to stimulate interest with catchy marketing slogans that have generated a polarized response from healthcare workers. As more useful data is obtained by using opportunistic methods, it soon becomes apparent that today's clinicians do not have the tools to interpret that data. They are poorly equipped due to a paucity of evidence-based design methods and tools in their curriculum and day-to-day work habits. Traditional approaches to evidence-based medicine, for example, use a method derived from the 19th century and amount to a series of yes or no questions that are supposed to guide the clinician on how to proceed. New tools and methods deeply integrated into the clinician's workflow are required, and it is big data's promise to enable these. There is also the possibility of using big data for population analytics to guide policy-making in areas such as public health, aged care management, waste reduction, and resource allocation.

2. Technologies and Tools for Big Data in Healthcare

The giant possibilities of big data in healthcare, thus, urge for robust data management techniques that can handle it. Technologies and tools that can help in gathering and analyzing data in healthcare for real-time analytics are the need of the time. The role of cutting-edge technology is very important to collect and analyze the volume, velocity, variety, value, and veracity of healthcare data. Various industries, and more especially information technology, are now tapping big data. Hospital systems, health insurance companies, intelligent medical devices, third-party wellness monitoring providers, and research systems for public health are some of the applications that have managed to accumulate big data. As a result, there has been an increase in the interest of the healthcare sector, which is in the process of collecting big data. Data science is one of the key tools for collecting images to gather data in healthcare in no time. IoT is one of the next technologies for data collection, gathering a large volume of data. The next technology software tool is also used to manage big data. (Deepa et al.2020)

There have been calls for more accurate and less time-consuming ways of diagnosing health problems. Toward this end, recommendations have been made to use big data to guide the development of a new science of health. Various industries adopted the big data approach because of the amount of data and the publication of tools that made the big data era possible. The accumulation of factual evidence about the complex relationships among an organism's genetic makeup, the environment, and disease can also be detailed in big data. Advances in genotyping and other technologies have already begun to yield an impressive amount of data with great potential to catalyze a new scientific understanding of the molecular basis of health and illness. It's helping to create accurate models. There are increasing numbers of use cases for graph databases and columnar databases in data science. This is to make the most possible study of data during the particular time frame. The process shifts from finding the correct tools to choosing the best machine-learning AI technology. It is breaking the traditional concept against using 20% of data in real-time. The storage cost for real-time analysis is growing lower, and as a result, technology enterprises are looking to the adoption of big data tools. There is an increased time-to-time which also helps make go-to-market data quicker. There is a reduction in data movement except for applying cloud computing services, particularly in the business and business intelligence models. It has relevance to firms in e-commerce since their data is readily accessible. The importance of becoming fast and data-driven cannot be overlooked in the face of competition and client demand. Organizations, therefore, end up choosing custom analytics models that consider their industry's big data trends and analytics prominence. However, making the right selection is extremely important for industry organizations because implementing new custom big data tools without regard to requirements that ideally fit the industry can add enormous value to the businesses. Keeping the specific use cases of healthcare in mind, the upcoming section covers the existing technologies that can be used to derive knowledge from such an enormous amount of data and present it in accessible forms. Each sector in healthcare typically has a range of technologies and faculties that process big data. Healthcare systems can use them to guarantee the ability of a proper supply of healthcare-related information and share ensured e-health data so that essential data insights are generated. These sectors would, in turn, further facilitate implementations in analytics capabilities and knowledge. In the following section, details on some of the most important related technologies are presented.

2.1. Data Collection and Storage

Healthcare data can be collected from various sources, including electronic health records, hospital information systems, picture archiving and communication systems, prescription drug orders, clinical notes, journals, publications, teleconferences, and exchange of emails with colleagues, browsers, search engines, and social media, to mention only a few. A substantial amount of healthcare data comes from nonclinical data sources, such as patient-generated data from smart devices, i.e., mobile phones; vital signs; smartphones; wearable devices; smart clothes; smartwatches; activity trackers; biosensors and other wearable sensors; and next-generation wearable sensors and devices, cyber data, and technologies.

Some of these data sources have already been established as a part of large-scale health analytics systems. Genomic databases, in particular, contain a wealth of data valuable for health and disease prediction and management for precision medicine and healthcare. Another potentially important source of health data is in situ and nonintrusive data scouting from online web interfaces of centralized knowledge resources. The wealth of information calls for scalable storage of biomedical data on community cloud platforms, which have capabilities to store big data and are scalable to requirements with acceptable performance.

Data migration among hospitals and from hospital facilities to common cloud platforms is inevitable since the size of the collected health data has already exceeded petabytes in most hospitals. To use such data to design newer treatment modalities, it will be necessary to have a common city archive for record keeping and retrieval. Therein lies the criticality of cloud-based cluster storage with a smaller scalable cluster file system. This minimizes the access time of the large data repositories. In short, the challenges today in data storage are in the design and deployment of specialized file systems besides database servers to support not only fast data analysis but also scalable storage of biological, biomedical, and healthcare data. The quality of data in healthcare is very important. There are distinct challenges in data quality, data integrity, and security of healthcare data because of the massive data generated every second. Data cleaning and preprocessing are two of the most fundamental aspects of data processing which ensure that the data used for processing and mining is error-free with acceptable quality. Adherence to standards and strict regulations is required for collecting and using patient information and data for analytics. Collecting healthcare data without proper consent can not only create serious legal and ethical issues, but countries can also ban the import of products that do not conform to the privacy compliances that they follow. Upcoming technologies and products can solely rely on analyzing a wealth of the patient's healthcare data for treatments and assisting doctors in enhancing patient healthcare. These products, however, need to implement the highest system designs with special attention to ethics and privacy laws and rules to fulfill the trust and compliance requirements by the stakeholders. The healthcare data can be further extended to analytics using non-intrusive in situ data scouts from hospitals and patient online websites associated with centralized knowledge resources facilitating additional supporting proof of healthcare analysis for a given product in different countries. (Wang et al. 2020)(Praveen et al. 2020)(Bhandari & Kenett, 2020)

2.2. Data Processing and Analysis

Data processing includes the techniques and methodologies utilized for data analysis. Sometimes, in the literature, analysis is suggested as a sub-section of data processing. Descriptive Analytics (DA), Predictive Analytics (PA), and Prescriptive Analytics (PrA) are separate. DA focuses on historical data and answers the questions 'What has happened?' and 'How many?' or 'How often?' DA looks at past and current data to get an estimation of the future. PA uses advanced algorithms to predict future outcomes based on historical data. It helps healthcare providers make informed decisions about the future. PrA extends and enhances PA by identifying optimal decision-making strategies. Advanced algorithms in PA and PrA include various machine learning methods. Machine learning is based on the ability of computers to learn without being explicitly programmed. There are three categories of machine learning: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning predicts an output given an input. It is widely applied in prediction, e.g., medical imaging, diagnostics, biological, and e-health applications. Unsupervised learning finds hidden patterns and intrinsic structures in input data. It is often applied in segmentation, clustering, and anomaly detection. Reinforcement learning learns how to behave when in unsupervised situations. It can effectively manage medical treatments and recommendation systems, e.g., for patients, and is faster than the policy optimization methods.

There are several challenges and constraints in the analytics field to analyze big data from healthcare, social, and biomedicine domains. The main challenges are algorithmic biases, which can potentially evolve in the use of machine learning and big data, and the computational and financial costs. Another challenge is the necessary skilled personnel, such as data analysts or data scientists, that are required to use the big data analytics tools, the complexity in patterns that need to be found, and the necessity of real-time analysis. The minimal access to cloud data processing facilities in the clinical environment adds to this list of problems that analysts have. Moreover, the use of high-throughput technology generates big data in healthcare to process in real-time with the aim of improving the quickness and resolution of services to target more personalized healthcare. To build a comprehensive study for practice, less emphasis is placed on practical research on big data visual analytics in healthcare, so despite its potential advantages, little is known systematically about big data visual analytics in healthcare. Consequently, this limitation motivates the production of this review. Data scientists convert complex big data into information, designing, gathering, interpreting, and reading big data management to provide meaningful results. The rise of personalized health services is further increasing the demand for big data analytics.

3. Applications of Big Data in Healthcare Informatics

In recent years, the potential applications of big data in healthcare informatics and analytics have been explored to improve patient outcomes. This has given birth to the concept of big data for Health, focusing on the value-creation aspect of healthcare communications and patient care. Through big data analysis, healthcare providers are attempting to make more evidence-based and informed decisions. Big data aids in extracting information from information pools and data silos via analytics to enable evidence-based decisions. Big data also allows for the development of state-of-the-art clinical decision support systems based on specialized and basic clinical knowledge and data networks. Predictive analytics, for example, collectively trigger an alarm when a patient’s condition begins to deteriorate, stopping or decreasing the variation of care and improving the patient experience, are employed as a solution to evaluate patient and system-level results. Predictive analytics can be used to forecast clinical reminders of chronic diseases in order to interrupt the progression of illness and inform healthy patients’ needs based on clinical factors and major social determinants of health. Big data and analytics are also used for infectious disease surveillance and outbreak monitoring to improve public health and the healthcare system. Public health is now more concerned with two challenges. These entail a shift from communicable diseases to long-term multi-morbidities that do not respect national boundaries and a greater opportunity and demand by service users for treatment. In both cases, policy must start at the level of politics. Healthcare users must be able to exercise a greater voice and control.3

Clinical Decision Support Systems (CDSS) is an expanded model of the DSS domain that aids clinicians' and researchers' decision-making. The impact of CDSS on patient care has been reported; the system automates many of the mundane tasks and allows them to keep up-to-date with scientific findings. Clinical Decision Support (CDS) systems analyze data points in order to help evidence-based decision-making, focusing primarily on patient-related information, clinical, environmental, and other data previously systematically and rigorously collected over a period of time. With adequate Big Data, along with advanced data analysis from the system, clinicians can easily diagnose the disease status and provide personalized support treatment to a patient. Moreover, the system also helps clinicians with various other remedies, including screening, prognosis, injury prevention, health improvement, and more.

Big Data provides enriched datasets and entrenched system intelligence that imposes accuracy on clinical decision-making. A system that is able to learn from data analysis technologies from Artificial Intelligence and Machine Learning gives doctors and other clinicians the capacity to stretch their capabilities and develop a personalized treatment plan based on patient personal and clinical data, matching symptoms to all new clinical and scientific evidence. The system mentors in information we can share to provide data analytics to assist us in health. When the DSS was applied for patient health monitoring, the system had a significant impact on the early prediction of patient deterioration that followed with downstream outcomes.

Clinical Decision Support has the potential to be a transformative tool for healthcare practitioners, moving the patient from being reactive in "fix me now," to concentrating on "I do not want to become that heart failure patient," to a personal health plan. The potential of CDSS can be limitless as technology and data advance. In an environment of fast-paced technological development, it is imperative that the existing ethical and regulatory framework guiding the development and appropriate use of CDSS shields, supports, and guides the emerging methodologies and their application. Albeit the ethical conundrum of utilizing such tools, steps should be taken from "can we" to "should we." Such mechanisms should underpin the appropriate use, with unambiguous guidelines for oversight and the clinical governance required to assure actionability. As recently expressed, "the reliance on digital tools to deliver more efficient health and care is essential in the 21st century. But digital tools should never replace human oversight and judgment. We need to be mindful of the limitations of technology."

3.2. Disease Surveillance and Outbreak Prediction

The impact of big data on public health also includes the domain of disease surveillance and outbreak prediction. Through predictive analytics, it is possible for public health authorities to know, based on measured disease activities, what to expect concerning new cases in the next time period. A number of techniques in healthcare analytics are used to monitor disease trends and to understand patterns in order to contribute to resource allocation and timely interventions. Case examples of epidemics have underscored the importance of upfront public health information, collective monitoring, and decision-making.

Next to these individual datasets, it is possible to combine multiple datasets to create more precise and informative insights. Examples are the combination of demographic data with environmental data to predict the incidence of diseases and the combination of demographic and clinical data to predict the severity or progression of an illness. Unfortunately, the quality, reliability, and timeliness of the data mostly deteriorate with an increasing spatiotemporal resolution. The use of different, complementary data sources is beneficial because they are less likely to suffer the same measurement or sampling defects and biases. Concerning the reduction of the burden of disease, biological systems, humans, and pathogens are three important variables that are monitored in real time to reduce the transmission and burden of disease. Concerning humans, the main public health monitoring system for daily surveillance is that of medical consultations and hospital admissions. In hospitals, sentinel surveillance systems are being used to monitor emerging diseases and other important health changes as well. The positive predictive value of such systems is close to 1, but the negative predictive value is low, and an unusual peak in patients or patient complaints is most often found one or two days after the beginning of an outbreak, by which time control efforts in the early stages of an epidemic may be less effective. Temporal trends such as diseases that follow a seasonal pattern are often detected substantially later, if at all, because the usual volume of cases in such diseases is high enough in number not to cause any alarm. Integration of the previous thinking would improve the specificity of outbreak prediction to better deal with false positives.

Despite the potential benefits in terms of both individual care and public health, the question of access to data and, particularly, who can see what information about whom and under which circumstances is complex, legally and politically sensitive, and open to social abuse. Additionally, data concerns in public health will be incorporated through transparent governance bodies, legislatures, and public scrutiny and involvement. Therefore, it is particularly important to be able to disclose, monitor, and enforce detailed rules concerning data handling and analytical validation, taking into account ethical, cultural, and societal norms, particularly in public health. Like other scientific fields, public health research generally includes a very complex system of patient consent and is subject to confidentiality-based rules. Conflicts can arise between using medical data for public health surveillance and the importance of protecting sensitive patient information.

4. Challenges and Ethical Considerations

A number of critical challenges accompany the utility of big data in healthcare analytics. One primary concern is data security, where the issues faced are similar to those in other sectors seeking to use big data analytics, and involve protecting against data breaches, which may compromise proprietary and confidential information, as well as unauthorized access to individuals' private data. Healthcare data are subject to especially stringent regulations around privacy, and ad hoc approaches to data security can be catastrophic to those affected, as such information as date of birth, address, and protected health information can be significant in instances of identity theft. Cyberattacks on organized healthcare institutions have been on the rise, with major data breaches disclosing protected health information from over 80 million individuals since 2009. Given the stakes involved, the healthcare industry has the responsibility to ensure that big data applications comply with existing and future privacy protections and regulations.

Beyond issues and challenges associated with data sharing, there are additional considerations regarding the use of big data in healthcare that speak to the moral responsibility of healthcare professionals and the broader fabric of ethical society. These ethical considerations include thoughtfully examining questions about what information is collected and shared; respecting the autonomy and privacy of research participants; and evaluating when and under what conditions it is ethically permissible to simply observe the effects of disease and in what circumstances it is an ethical requirement to intervene. Enhanced data volume means more opportunities to misuse data and, consequently, increased possibilities of societal discrimination. Indeed, biases such as sexism and racism can be amplified and reinforced by data. Accordingly, protecting vulnerable communities is a priority for ethical big data practices and would need to be enshrined in the informed consent process. Moreover, ethical data use is achieved through best practices that are shared and developed across stakeholders and that also seek to manage the acquisition and maintenance of accurate data as much as possible. Big data can be used to improve healthcare, the lives of its users, and the effectiveness of its surrogate outcomes. Proper stewardship and practice will need to ensure that data security is robust, that privacy is protected, and that ethical considerations are consonant with the emerging use of big data in healthcare.

4.1. Data Security and Privacy

One of the major debacles in sharing big data in healthcare is data security. Data security is an essential aspect of an individual as well as any society. Healthcare data is vulnerable to cyberattacks, data breaches, unauthorized data access, etc. As healthcare systems continue their move to cloud solutions developed by third parties, their data storage and networking services may still be vulnerable. Last year saw ransomware penetrate several healthcare facilities. Regular, consistent IT advancements conducted to mitigate and prevent future attacks are welcomed. However, to undermine appropriation efficiency, security researchers highlight potential new privacy and security risk vectors via data sharing and reuse. The simple solution would be to encrypt any healthcare data shared, such as with homomorphic encryption. While enticing, it does not prevent privacy attacks and is very lossy compared to clear text analytics.

Various strategies can be employed to tackle the challenges, risks, and issues associated with data sharing in healthcare. Some of them include data access control, protection by design, anonymization, and transparent data licensing and contracts that specify how a dataset should be accessed, managed, stored, and possible uses thereof. Organizations are encouraged to report incidents and seek guidance to address concerns arising from data in their care. Healthcare providers are encouraged to inform patients of any plans to share their medical data within or across borders. Not all data sharing is detrimental to privacy, and it is possible to engage in a secure, transparent, and ethical data sharing service. Data sharing need not compromise patient privacy; a culture of privacy awareness, trust, and openness should be nurtured. Healthcare providers must conform to the privacy requirements of the Protection of Personal Information Act and the legal jurisdiction of South Africa.

4.2. Ethical Use of Patient Data

Healthcare data of individuals have long been recognized as sensitive, and its collection requires the trust and cooperation of the persons whose information is being recorded. Ethical handling of patient information, whether aggregated or not, requires it to be collected with respect for patient autonomy and consent. Data cannot be collected without the patient being informed and providing explicit consent, unless there is a legitimate and urgent need. Once collected, patient data should be de-identified and anonymized in order to protect individual privacy. There is a tension between using patient personal record data for research or other purposes and protecting individuals as countenanced in ethical norms, which is likely to be perceived as illegitimate by research participants or members of governing health organizations. Additionally, aggregation increases the opportunity for re-identification to occur. If it is demonstrated that de-identified data are re-identifiable, individuals have an ethical obligation to guard against re-identification. The secondary use of already collected data for new purposes poses special ethical questions. Because the data have already been collected, the need to obtain explicit and informed consent for their alternate use is diminished but not eliminated.

If big data research initiatives are to be trusted, they must operate with informed concern for ethics relating to the data used. Current trends in informatics research risk encouraging philosophers of technology to be conceived primarily as advocates for big data. They are not. Ethical aspects of handling patient data and safeguarding privacy and confidentiality are highly important if trust in healthcare organizations is to be sustained. A critical concern for anyone who supports the use of big data in healthcare research is that a major breach involving big data will compromise healthcare organizations and their ability to leverage big data research. At stake is societal acceptance of a variety of healthcare organizations. There is one way to deal with ethical concerns over using big data in healthcare research: build upon the ethical consideration of data and minimize the risk of legitimate privacy concerns through implementing defensible ethical guidelines and quality control for handling big data. Best big data practices should adhere to high ethical standards as a matter of institutional integrity. Where individuals or governing bodies in healthcare organizations have concerns regarding the values held in a big data initiative, they should be taken seriously. Upholding ethical concerns is a legitimate reflection on an organizational mission. In the same breath, it is legitimate for those entities desiring to enhance the quality of reanalysis of patient records to push for appropriate big data research policy in order to deepen our ethical commitment to being trustworthy in our activities.

5. Conclusion and Future Directions

The potential role of big data in transforming the delivery of healthcare and improving patient outcomes through better and more accurate analytics is very significant. Big data can be applied to diverse areas in both clinical settings and healthcare research. Technologies that are used for storing and processing raw big data from healthcare environments need to be well understood. This includes technologies that relate to databases, data storage, data analysis, data extraction, data transformation, data loading, as well as data visualization. Ethical considerations linked to the provision of patient data for big data and AI applications in healthcare analytics have been discussed. Several key challenges or concerns related to big data use in healthcare informatics include "The data will not be used correctly," "Inefficiency trap," "Electronic healthcare record centers of excellence can finite reward," "Personal health records," and "Big data capabilities are overrated."

It appears that data-related research, and in particular big data research, is ongoing and producing results faster than we can keep up with. This has challenged us to prepare an updated version of this survey that will cover these fields in the future. It appears that big data and healthcare informatics applications need more research targeted towards them. This research needs to be directed in terms of both developing the most appropriate technology that is adequate for streamlining with the new healthcare practices in various regions, as well as in comparison to the current technologies and practices, and targeting research towards ways of coordinating healthcare professionals and hospital financial departments. Collaborative and interdisciplinary research among various stakeholders and professionals is also critical in overcoming the challenges related to big data and healthcare informatics. Moreover, besides the new big data technologies and practices developed alongside the fourth industrial revolution, it is important for future evolution to be supported by the development of appropriate regulatory frameworks. These may also embrace existing ethical considerations and constraints, particularly those related to making data available for researchers. Finally, the importance of having alternative and continued training for healthcare professionals related to data ethics and legal considerations is a must.

References

Avula, R. (2020). Data-driven decision-making in healthcare through advanced data mining techniques: A survey on applications and limitations. International Journal of Applied Machine Learning and Computational Intelligence, 12(4), 64-85. neuralslate.com

Bhandari, D. R. & Kenett, R. (2020). Paradigm shift of statistical big data in healthcare: management, analysis and future prospects. Pravaha. nepjol.info

Deepa, N., Pham, Q. V., Nguyen, D. C., Bhattacharya, S., Prabadevi, B., Gadekallu, T. R., ... & Pathirana, P. N. (2020). A survey on blockchain for big data: Approaches, opportunities, and future directions. Future Generation Computer Systems, 131, 209-226. [PDF]

Praveen, S. P., Murali Krishna, T. B., Anuradha, C. H., Mandalapu, S. R., Sarala, P., & Sindhura, S. (2020). A robust framework for handling health care information based on machine learning and big data engineering techniques. International Journal of Healthcare Management, 1-18. [HTML]

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. Z., Paul, R., Smriti, K., ... & Somani, B. K. (2020). The role of data science in healthcare advancements: applications, benefits, and future prospects. Irish Journal of Medical Science (1971-), 191(4), 1473-1483. springer.com

Wang, M., Li, S., Zheng, T., Li, N., Shi, Q., Zhuo, X., ... & Huang, Y. (2020). Big data health care platform with multisource heterogeneous data integration and massive high-dimensional data governance for large hospitals: design, development, and application. JMIR Medical Informatics, 10(4), e36481. jmir.org