

Health sector and application of big data: A case study of India

**Roma Mitra DEBNATH, Indian Institute of Public Administration (IIPA),
India**

Apaar DHINGRA, Research Scholar, India

Received: 21.05.2021, Revised: 20.11.2021, Accepted: 03.12.2021

doi: <http://10.29015/cerem.923>

Aim: The paper aims to study and present the case study of the health sector of India. The paper also aims to identify the opportunities for the application of Big data in the health sector. The major stakeholders of the system viz. doctors, hospitals, clinics, insurance companies, pharmaceutical companies, research, and development organizations, industries manufacturing medical instruments, laboratories, medical data analysts, and many more are utilizing big data and predictive analytics in their critical decision making. The predicted revenue was expected to reach 280 billion by 2020 as per the statistics given by the Indian Brand Equity Foundation.

Research methods: A critical review has been conducted using electronic sources between 2015 and 2020, limited to English language articles and reports published from 2015 onwards. The reviews will be classified to identify the opportunities for future application of Big Data.

Conclusions: The paper presents a trend in the use of Big Data Analysis in the health sector. The paper also explores and identifies the areas of future application of big data to increase the efficacy of the system.

Originality/value of the article: This is an original piece of article in the context of India in terms of documenting the big data applications in the health sector and identifying the opportunities for the future application of the same.

Implications of the research: This research holds a significant contribution towards the implications of the application of Big Data in the health sector. The newly identified areas of the health sector, which can be improved by using the big data analytics, are important for the policy makers of the organizations, including the Government.

Correspondence address: Dr. Roma Mitra DEBNATH, Associate Professor, Indian Institute of Public Administration (IIPA), New Delhi, India, Email: roma.mitra@gmail.com. Apaar DHINGRA, Research Scholar, India, Email: dhingraapaar@gmail.com.

Limitations of the research: The research has been conducted based on the secondary data, which area available in the public domain. However, due to COVID 19, there could have been more innovations in the health sector in terms of using the big data, which may not have been published or are available in the public domain. Also, collection of the primary data in terms of an interview with the administrators/management can be explored in the future study.

Keywords: health care, big data, India

JEL: I00, I15

1. Introduction

There has been a paradigm shift in the health care system. The way healthcare is delivered has been undergoing a major transformation for some time now. The change has been from volume-based care to value-based care.

We are all familiar with the traditional service model, where the patients pay by the number of visits, procedures, or tests. A change has been there from the hospital-centered and professional-focused approach towards a distributed patient-centered care model, where many care elements will be delivered virtually and by “informal” carers, meaning carers without formal professional training.

The arrivals of mobile phones and portable computers in the late 1980ies were the start of an information communication technology revolution that has fundamentally changed the health care sector too, like other sectors. However, the recent digital disruption has catalyzed healthcare’s adoption of big data.

It is safe to assume that we all are familiar with the concept of big data. Big data is a part of every industry, and it’s going to continue to grow rapidly as digitization and connectivity increase. Big Data is characterized by the 5V’s. The five Vs are (i) Volume, (ii) Velocity, (iii) Variety, (iv) Veracity and (v) Value.

The Volume speaks about the quantity of the data, or how voluminous is the collection of data. The sheer volume of data generated these days by real-time applications is enormously large and it runs to petabytes and exabytes of data. Big Data technology enables us to store this amount of data on distributed systems.

The Velocity is the rate at which data is generated. For example, the data generated through the phone calls at a call center in a minute or the amount of credit

card transactions that happen in a second. As the data is generated at a high speed, it requires a high-speed analysis too.

The variety defines the multiple sources of the data and multiple formats of the data. The data is available in different formats like structured, semi-structured, and unstructured data. Unlike, the statistical data, big data can be unstructured or semi-structured. It is generally believed that 90% of the data generated is unstructured in nature.

Veracity defines the trustworthiness of the data. Even though the data is generated, to what extent the data is genuine is one of the characteristics of the big data.

The final characteristic of the big data are defined by Value, which refers to techniques of deriving value/usefulness from data. There is an intrinsic value that the data may possess and must discover for analysis. It is the most significant aspect of the big data, as the decision-making process depends on the usefulness of the data.

In healthcare, big data is being generated at all corners of the ecosystem, and consumers, providers, and payers are all contributing to this system. A number of data sources viz. administrative data such as healthcare claims data, clinical data, self-generated data such as that produced by fitness and wellness mobile applications and wearable devices, and patient-reported data such as surveys are driving this growth.

Apart from the health-related data, the health sector has been affected by the external data too. For example, the data related to consumer research or data on education, social media data which captures the sentiment of the people on various aspects can be used by the health sector appropriately to improve health care.

The primary objective of the paper is to explore the application of big data in the health sector, studying them to explore the strengths and challenges of the same. The paper has been organized into five sections. Section I provides an introduction, Section II provides an extensive literature review of the application of big data, Section III presents the findings and the concluding section IV concludes the research article.

2. Literature review

Mählmann et al. (2017) have discussed the seemingly important role that big data will play in the public health policy-making process. In this context, health data cooperatives (HDC) are a key component and core element for public health policy-making and for exploiting the potential of all the existing and rapidly emerging data sources. HDC is an ecosystem where stakeholders' data, knowledge, materials, and services are integrated into a cloud-based framework. At first, its heterogeneous components might include data sources such as national electronic patient records and non-electronic systems, research data, repositories of biological material, genomic and proteomic data, etc. are processed, analyzed, and mapped comprehensively in order to create a network of knowledge and actionable insights. Treating research data as a public good, creating HDC to empower citizens through citizen-owned health data, and allowing data access for research and the development of new diagnostics, therapies, and public health policies will yield the transformative impact of digital health. The HDC model for data governance is an arrangement, based on moral codes, that encourages citizens to participate in the improvement of their health. This then enables public health institutions and policymakers to monitor policy changes and evaluate their impact and risk on a population level.

Van Staa et al. (2016) have mentioned the efficient use of large scale health data has the potential to benefit patient care, public health, and research. The handling of such data, however, raises concerns about patient privacy, even when the risks of disclosure are extremely small. Failures in the implementation of data-sharing projects have eroded public trust. In the wake of NHS England's decision to close down its care data programme, the study also examines better ways to do it.

Belle et al. (2015) have discussed the pivotal role of the rapidly expanding field of big data analytics in the evolution of healthcare practices and research. Big Data has provided tools to accumulate, manage, analyze, and assimilate large volumes of disparate, structured, and unstructured data produced by current healthcare systems. Big data analytics has been recently applied towards aiding the process of care delivery and disease exploration. However, the adoption rate and research

development in this space is still hindered by some fundamental problems inherent within the big data paradigm. The study highlights the issues and tries to answer them.

Kaur et al. (2018) have conducted an empirical study to analyze the role of big data in the healthcare industry. Over the years, significant work has been done using big data in the healthcare sector. It is intricate to envision the way machine learning and big data can influence the healthcare industries. Most of the authors who implemented the use of machine learning and big data analytics in disease diagnosis do not give significant weightage to the privacy and security of the data. Here, a novel design of smart and secure healthcare information system using machine learning and advanced security mechanism has been proposed to handle big data of the medical industry.

Dash et al. (2019) have discussed that the healthcare industry data from all the sources including biomedical research requires proper management and analysis in order to derive meaningful information. Otherwise, seeking a solution by analyzing big data quickly becomes comparable to finding a needle in the haystack. The study points that in order to provide relevant solutions for improving public health, healthcare providers are required to be fully equipped with appropriate infrastructure to systematically generate and analyze big data. Efficient management, analysis, and interpretation of big data can change the game by opening new avenues for modern healthcare. With a strong integration of biomedical and healthcare data, modern healthcare organizations can revolutionize medical therapies and personalized medicine.

Mehta et al. (2019) have mentioned the application of analytics, machine learning and artificial intelligence over big data enables the identification of patterns and correlations and hence provides actionable insights for improving the delivery of healthcare. The study gives a comprehensive view of the current state of research and its application.

Madanian et al. (2019) have discussed the use of mobile devices in health (mobile health/mHealth) coupled with related technologies promises to transform global health delivery by creating new delivery models that can be integrated with existing health services. The study identifies challenges related to the current status

of India's healthcare system – with a specific focus on mHealth and big-data analytics technologies. To address these challenges, they have proposed a framework for integrating the generated mHealth big-data and applying the results in India's healthcare. Further, a model is proposed that utilizes generated data from mHealth devices for big-data analysis that could result in providing insights into the health status of the Indian population. The insights could be important for public health planning by the government towards reaching Universal Health Coverage.

Kumar and Manjula (2014) have discussed the big gap in the knowledge about innovations in public and private health financing and delivery Health care as one of the greatest concerns in India. While those living in cities and big towns have access to high-end health services, the millions of people living in rural India, particularly in the remote parts of the country face problems of inadequate facilities and poor access to healthcare. The inefficiencies and inequities in the public health care access in India have pushed forward the need for creative thinking and innovative solutions to strengthen the same. Further, they also addresses the critical computing and analytical ability of Big Data in processing huge volumes of transactional data in real-time situations to turn the dream of Svasth Bharath (Healthy India) into reality.

Pramanik et al. (2019) have discussed the considerable benefits and opportunities of Big Data. Big data has attracted the momentous attention of all the stakeholders in the healthcare industry. Handling big data always remains a big challenge. The chapter identifies all the possible opportunities and challenges in realizing the benefits of big data in healthcare, along with a brief survey of the tools and platforms, architectures, and commercial infrastructures for healthcare big data.

Duggal et al. (2016) have discussed an overview of Big Data, its applicability of in the Healthcare system, opportunities that it holds to improve Indian healthcare. In India, increase in income levels, an aging population, rising health awareness, and changing attitudes towards preventive healthcare are expected to boost healthcare services demand in the future. The massive growth of the volume, velocity and variety of digital health data creates both manageability issues and opportunities for greater patient insights. Finding a way to successfully manage Big Insights to eventually “predict, personalize, and prevent” could lead to early detection of

potential problems, which means healthier people, as well as fewer complications and admissions.

Khanra et al. (2020) have performed a systematic literature review (SLR) to synthesize prior research on the applicability of big data analytics (BDA) in healthcare. The findings suggest that applications of BDA in healthcare can be observed from five perspectives, namely, health awareness among the general public, interactions among stakeholders in the healthcare ecosystem, hospital management practices, treatment of specific medical conditions, and technology in healthcare service delivery.

Raja et al. (2020) have provided a systematic review study on healthcare big data based on the systematic literature review (SLR) protocol. In particular, the study highlights some valuable research aspects on healthcare big data according to the defined inclusion-exclusion criteria. The study determines the extent of healthcare big data analytics together with its applications and challenges in healthcare adoption. Besides, the big data produced by these healthcare systems, its characteristics, and various issues in dealing with it, as well as how big data analytics contributes to achieve a meaningful insight on these data set.

Pastorino et al. (2019) have discussed the potential of Big Data in healthcare relies on the ability to detect patterns and to turn high volumes of data into actionable knowledge for precision medicine and decision-makers. In several contexts, the use of Big Data in healthcare is already offering solutions for the improvement of patient care and the generation of value in healthcare organizations. This approach requires, however, that all the relevant stakeholders collaborate and adapt the design and performance of their systems. The research is an overview of best practice initiatives in Europe related to Big Data analytics in public health and oncology sectors, aimed to generate new knowledge, improve clinical care and streamline public health surveillance.

Benjelloun et al. (2015) have discussed the overview of big data opportunities applications, and tools in the modern world across several business sectors. They were of the view that big data brings valuable insights, particularly for the healthcare sector. Several applications of big data have been tested to improve private and public medical services and to better support patients and medical practitioners. This

can help bring a revolution in the health domain by supporting the optimization of operational services, better understanding of disease evaluation supporting medical decision-makers a better prevention strategy, and customization of medical services. However, the challenges like those of integration of various disparate sources medical entities, and medical records cannot be overlooked. Big data with its features like distributed storage, massively parallel processing, fault-tolerant, and scalable system, can be a reliable answer to these challenges. New database is like a neoSQL, new SQL help in searching and indexing the complex datasets in real-time and hence can be an answer to complex problems of traditional searching methods. But the security issues of big data, like how to securely manage large and structured, and heterogeneous types of data sets, integration of security mechanisms into distributed platforms, and ensuring a good performance at the same time is going to be a challenging task. Data owners have the responsibilities to set clear security clauses and policies to be respected by outsourcers, they also have to analyze security risks of combining different evolving big data technologies.

Cavanillas et al. (2016) have discussed several developments in the healthcare sector, such as escalating healthcare costs, increased need for healthcare coverage, and shifts in provider reimbursement trends, triggering the demand for big data technologies to improve the overall efficiency and quality of care delivery. Big data technologies and health data analytics provide the means to address the efficiency and quality challenges in the health domain. But the highest impact of big data applications in the healthcare domain is achievable when it becomes possible to acquire data from various data sources such that different aspects can be combined to gain new insights. The availability and integration of all related health data sources, such as clinical data, claims, cost, and administrative data, pharmaceutical and R&D data, patient behavior, and sentiment data is highly relevant. The availability of the technologies will not be sufficient for fostering widespread adoption of big data in the healthcare domain. The critical stumbling block is the lack of business cases and business models. As big data fosters a new dimension of value proposition in healthcare delivery, with insights on the effectiveness of treatments to significantly improve the quality of care, new reimbursement models that reward quality instead of quantity of treatments are needed.

Galetsia et al. (2020) present a systematic overview of the literature to determine the way Big Data Analytics has managed to improve the healthcare domain. They concluded that the most popular analytical techniques that scientists use to make meaningful interpretations of data are: modeling, machine learning, data mining, visualization, and statistical analysis. In particular, machine learning is the most applied technique across almost all created values and data types that offer immense potential in the healthcare predictive analytics arena to improve outcomes in many domains of research. BDA analysis indicates that most users use clinical or medical structured or unstructured data for their studies to build new approaches for the diagnosis of personalized healthcare and to invent entirely new business models to reduce time, cost of search, or processing while maintaining quality. It is clearly shown that there is demand for research in health analytics to focus on improving the technological aspects. Further, there is a definite need in healthcare for systems that support or improve the decision-making ability of clinical experts. The progress that has been made via programs like Hadoop and MapReduce has increased performance by reducing time and pre-computing computationally intensive jobs. The main difficulty with big data in healthcare is that most data are often unstructured, which means that there are obstacles to computationally processing the largest part of them. Scientists are in a continuous effort to advance infrastructure in order to achieve the greatest possible analysis and to further develop computational methods in order to extend systems' capabilities. It is expected that more investment will be given to IT infrastructure and to BDA experts in the healthcare sector, and make these accessible to relevant professionals.

Alonso et al. (2017) present a review of the existing researches in the literature regarding big data sources and techniques in the health sector they try to identify the most used prediction of chronic diseases. Recent developments in information technology like social networks, mobility, cloud computing, big data, and the internet of things imply advanced usage and greater challenges across all sectors. This will require a considerable effort of adaptation by the organizations specializing in the healthcare sector. Also, it is necessary to make a series of changes, in the currently existing information systems of the healthcare system. The objectives of these changes should be majorly aimed to manage and analyze large

volumes of data from very diverse sources recorded in very heterogeneous formats. Big Data also offers the opportunity to allow an effective and accurate medicine by stratification of the patient. This is indeed a key task towards personalized healthcare. Better use of medical resources through personalization can lead to well-managed health services that can overcome the challenges of a rapidly growing and aging population. Therefore, advances in Big Data processing for bioinformatics, chronic disease detection, genomics, and biomedicine will have a major impact on future clinical research. Unlike traditional solutions, Big Data can manage large volumes of data of all kinds, especially unstructured, at a rate much higher and much lower consumption of resources. Another of the great advantages of this technology is the possibility of using advanced analytical techniques, such as predictive analytics, intending to operate on massive data and construct predictive quality models to support decision making.

Salas-Vega et al. (2015) have identified initiatives aimed at promoting the use of big data in European Union (EU) health care, highlight expected challenges, and use these to evaluate EU big data policy developments to the extent that they are able to advance health sector priorities. Several ongoing big data initiatives in health across the EU, at both national and European levels like data centers, confidentiality, and data security, e-health and m-health, and genomics and bioinformatics aim at strengthening of health system effectiveness, accessibility, resilience, quality, and performance, as well as the promotion of health research. However, EU policy makers are yet to tailor data policy to accommodate conceptual challenges to health sector development like, quality and performance improvement—that fall within European legal competencies and responsibilities in health. As EU policy makers begin to call for greater integration of health data sets it is important to bear in mind that as in any security domain, the weakest link can break the chain. Several key challenges that present an obstacle to big data use in health care are confidentiality and data security, access to information, data reliability, interoperability, and management and governance. Nevertheless, European Commission (EC) discussions on big data policy are still in their infancy, as confirmed by communications with key EU policy stakeholders. Additional progress in the merger between big data policy and sectoral objectives may therefore be expected soon as the EC embarks on

this new field of policy. At this time, however, it remains unclear how big data developments will advance health sector objectives, casting doubt on optimistic predictions of the return on big data investments in the EU.

Wyber et al. (2015) have explored current and potential applications of big data to public health and healthcare delivery in low- and middle-income countries. The paper also talks about the benefits, risks, and opportunities for big data in health and recommends the use of big data in the delivery of healthcare services in low- and middle-income countries. The big data approach facilitates the development of learning systems and enables precise management of data to improve the health of entire populations. Sheer size increases both the potential risks and potential benefits of the approach. Steps like Collaborative governance, careful analysis, and technical partnerships are needed to minimize the risks. The complexities of the process should not be underestimated. In low- and middle-income countries, the shepherding of the transition from paper records to petabytes of digital storage provides another opportunity for global health institutions to offer useful governance.

Kumar and Singh (2015) have discussed huge amounts of structured, unstructured, and semi-structured data that have been generated by various institutions around the world and, the need of the health industry sector to manage the big data being produced by various sources. The impact of big data in healthcare, and various tools available in the ecosystem for handling it have also been discussed. The big data industry could very well revolutionize the healthcare sector by providing the most suitable patient diagnosis. This will help the system provide the patient with the right living right care right provider right value and right innovation. Big data-based applications can provide data-based solutions to several prevailing issues like treatment of cancer and genomics, monitoring of patient vitals, healthcare intelligence, prevention and detection of frauds, hospital networks, etc. The paper also proposes the use of a conceptual architecture for solving healthcare problems in big data using Hadoop-based technologies, involving the utilization of big data, generated by different levels of medical stakeholders and the development of methods for analyzing this data to obtain answers to medical questions. Combining of Big data analytics along with healthcare provided can lead to treatments that are patient effective and patient-specific by providing the ability to

prescribe appropriate medications for each individual rather than those that work for most people.

Panda et al. (2017) have proposed a Digitised HealthCare Model for India. The paper focuses on the development of a mobile/web application, through which patients send their symptomatic queries to the doctors through a server. The idea is to develop a mobile application equipped with first aid instructions, according to the nature and severity of the symptoms, the patients will either be directed to respective departments or given emergency help for further treatment. Within the time a huge amount of data is collected from users and doctors, this big data will be used to train machines to automate the tasks to some extent. The information gained from analyzing massive amounts of aggregated health data can provide useful insight to improve quality and efficiency for providers and insurers alike. This makes the patients reach out for healthcare solutions easily and cheaply and makes healthcare an easy reach for the unprivileged also. Thus, this unified model can serve as a data collection, delivery as well as and analytic tool in the healthcare domain. It will play an important role in preventive, promotive, and curative health. The major advantage of such a system is to detect and predict diseases accurately, easily, and faster with the help of machine learning.

Weiwei et al. (2021) have collected evidence from the existing literature and organized it through the process to benefit practitioners to identify the right techniques for their specific purpose of research. The study presented a review of the existing techniques of big data insights and scientific programming in the industry of healthcare. The literature in this field has increased drastically during 2017-2019. With the IT revolution, the technology of healthcare also is getting advanced at a high rate. The volume of healthcare data is always rising and, such data can make it tougher to recognize a useful form of data-generating meaningful information. The data are always playing a significant role in organizations and industry for their daily activities to function smoothly. This rise in big data volume can be a challenging task concerning analyzing the data for normal industry and IoT.

Akinnagbe et al. (2017) try to find out the impact of big data applications on healthcare services in Africa especially during epidemics and through the public health system. A systematic literature review has been conducted, to present cases of

big data applications in healthcare in the African region, and further, to explore potential ethical challenges of such applications. Against all challenges and issues, the case studies point out that big data analytics can transform the healthcare system there. Although limited uses and benefits have been reported in non-communicable and chronic diseases, with an increase in the use of mobile technologies and social media, alongside emerging investments into big data technology, rapid, extensive implementation and use of big data analytics in healthcare across the continent is highly probable. To make the gains of big data to be sustainable in the region, there is a need for stakeholders in the health industry to put in place a business intelligence roadmap so that data from diverse data sources can be enabled, mined, and transformed into useful actionable knowledge for the benefit of the community.

Olivera et al. (2018) have discussed how big data approaches would impact the future of Inflammatory Bowel Disease (IBD). Big data have been successfully used in many different areas, including finance and politics, and more recently have been increasingly implemented in health care. Big data analytics are innovative approaches to help disentangle the complexity of IBD. Potential applications of big data in the field of IBD might include precise phenomapping, the development of predictive models, precision medicine, epidemiological models, and drug discovery.

The paper concludes that researchers will face several potential limitations and challenges when using big data approaches in IBD, including ethical and legal restrictions, heterogeneous data sources, poor quality data, and the need for validation. The idea of big data analysis in health care is still in its infancy, and IBD research will benefit from its many promises in the coming years. Chakrabarty and Das (2020) have discussed about progressions and rapid transformation of the IT sector which has endorsed the legitimacy of abundant data, multiple dynamic variables & critical complexities in the modern world. The paper tries to put light on n relevance of 'Big Data Analytics (BDA)', concerning Healthcare sector, which is one of the vibrant socio-economic variables which have correlations with other aspects of life. Health is the basic input for a holistic developmental process but it also is the outcome of various developmental factors. BDAs are being used across various sectors of the economy, but have farfetched implications for health care. The developed nations have been yielding the most feasible solutions using various

forms of analysis of big data. In the health sector, the application of BDAs has been attempted and experimented within the developed nations which have resulted in prolific and sustainable solutions to the most typical cumbersome problems. The chapter demonstrates how BDAs can make progressive reforms in a developing country like India, where the Healthcare sector hasn't been very great and continuously faces emerging challenges

3. Findings

The availability of health-related Big Data can have a positive impact on medical and healthcare functions. The data-sharing approach can improve outcomes for patients and evidence-based healthcare decision-making as reported during the workshop on 'Digitalization and Big Data: implications for the health sector'. The use of Big Data in healthcare, in fact, can contribute at different levels

- (i) increasing earlier diagnosis and the effectiveness and quality of treatments by the discovery of early signals
- (ii) widening possibilities for prevention of diseases by identification of risk factors for disease
- (iii) improvement of pharma co-vigilance and patient safety through the ability to make more informed medical decisions based on directly delivered information to the patients.
- (iv) prediction of outcomes

All these aspects should eventually lead to a reduction in inefficiency and improvement in cost containment for the healthcare system. Examples of Big Data analytics for new knowledge generation, improved clinical care, and streamlined public health surveillance are already available.

Some important opportunities are identified as

1. e-Health File: The creation of an e-Health care file for each patient, where all health care providers and patients themselves were able to submit information with the consent of the patient). Both subjective data, symptom diaries, lab data, image diagnostics, pathology reports, etc., could be filed. This could be further

extended to e-prescriptions. Paper-based prescriptions are archaic and lead to several miseries each year due to prescription errors. But if every doctor is provided with an electronic prescription system, it would improve safety by making prescriptions easier to read and providing instant checks on drug interactions, dosages, and a patient's medication history.

2. **Electronic Medical Records:** Medical Experts agree that electronic medical records (EMRs) are a must for better health care in India. But, at present only a few hospitals are maintaining EMR's, mainly because of cost, privacy issues, and the lack of one compatible, easy-to-use infrastructure.
3. **Stop Unnecessary Treatments:** Doctors should avoid trial and error type of medication. The problem must be examined thoroughly by performing the required diagnostic tests during the preliminary days of disease. The right treatment should be suggested at the first visit only which avoids the disease to become more critical. Most of the issues are arising with misdiagnosis and wrong treatment during the early stages.
4. **Tele-Medicine:** Doctors can often diagnose or prescribe without seeing the patient. The patient has to physically appear before the nearby health center, where the nurses or health workers will diagnose at first level, note the symptoms, and inform the high-level specialist doctors about the case. After examining the reports, the specialist doctor suggests the treatment through a health worker which reduces costs and creates satisfaction by virtual communication of patients and doctors to discuss medication changes and test results through an online system.

However, like other technological advances, the success of these ambitious steps would apparently ease the present burdens on healthcare especially in terms of costs. It is believed that the implementation of big data analytics by healthcare organizations might lead to a saving of over 25% in annual costs in the coming years. Better diagnosis and disease predictions by big data analytics can enable cost reduction by decreasing the hospital readmission rate. Big data analytics can also help in optimizing staffing, forecasting operating room demands, streamlining patient care, and improving the pharmaceutical supply chain. All of these factors

will lead to an ultimate reduction in the healthcare costs by the organizations. Further, the usage of services like mHealth and big-data analytics will make healthcare more affordable, available, and accessible for wider communities; so the Indian healthcare system could benefit from modern technologies. Integrating mHealth into the Indian healthcare system, mHealth can create new opportunities for healthcare, especially in resource-poor environments where healthcare systems have limitations in terms of infrastructure, expertise, and human resources. mHealth could play a key role in Indian healthcare, especially for rural communities whose access to healthcare could be limited. For these communities, mHealth could facilitate access to basic health services and the analysis of it generate data may be useful for shaping the healthcare demands. As healthcare is moving towards evidence-based medicine globally, having a continuous systematic review of clinical data is necessary for effective and efficient decision-making. This requires high accuracy in operations and decision-making while time is a vital factor. Therefore, increased use of health data may be an important way to improve efficiency and effectiveness in the sector.

4. Major challenges identified

The process of using big data in the field of health is not free of challenges. The first major challenge is that of storage of data. Storing a large volume of data is one of the primary challenges, but many organizations are comfortable with data storage on their premises. It has several advantages like control over security, access, and up-time. However, an on-site server network can be expensive to scale and difficult to maintain. It appears that with decreasing costs and increasing reliability, cloud-based storage using IT infrastructure is a better option which most of the healthcare organizations have opted for. Organizations must choose cloud partners that understand the importance of healthcare-specific compliance and security issues. Once stored properly, the data needs to be cleaned or scrubbed to ensure accuracy, correctness, consistency, relevancy, and purity after acquisition. This cleaning process can be manual or automated using logic rules to ensure high levels of

accuracy and integrity. More sophisticated and precise tools use machine-learning techniques to reduce time and expenses and to stop foul data from derailing big data projects. This means that the data needs to be converted into a unified format. Patients produce a huge volume of data that is not easy to capture with traditional HER format, as it is knotty and not easily manageable. It is too difficult to handle big data especially when it comes without a perfect data organization for the healthcare providers. A need to codify all the clinically relevant information surfaced for the purpose of claims, billing purposes, and clinical analytics. Therefore, medical coding systems like Current Procedural Terminology (CPT) and International Classification of Diseases (ICD) code sets were developed to represent the core clinical concepts. However, these code sets have their own limitations. Even after this, the accuracy of the data is a challenge. Some studies have observed that the reporting of patient data into EMRs or EHRs is not entirely accurate yet, probably because of poor EHR utility, complex workflows, and a broken understanding of why big data is all-important to capture well. All these factors can contribute to the quality issues for big data all along its lifecycle. The security of data is another challenge. There have been many security breaches, hackings, phishing attacks, and ransom ware episodes that data security is a priority for healthcare organizations. After noticing an array of vulnerabilities, a list of technical safeguards was developed for the protected health information (PHI).

Further, to have a successful data governance plan, it would be mandatory to have complete, accurate, and up-to-date metadata regarding all the stored data. The metadata would be composed of information like time of creation, purpose and person responsible for the data, previous usage (by who, why, how, and when) for researchers and data analysts. This would allow analysts to replicate previous queries and help later scientific studies and accurate benchmarking. This increases the usefulness of data. Metadata would make it easier for organizations to query their data and get some answers. However, in absence of proper interoperability between datasets the query tools may not access an entire repository of data. Also, different components of a dataset should be well interconnected or linked, and easily accessible otherwise a complete portrait of an individual patient's health may not be generated. Medical coding systems like ICD-10, SNOMED-CT, or LOINC must be

implemented to reduce free-form concepts into a shared ontology. If the accuracy, completeness, and standardization of the data are not in question, then Structured Query Language (SQL) can be used to query large datasets and relational databases. After all the aforementioned challenges are taken care of, we need to keep in mind that the patients may or may not receive their care at multiple locations. In the former case, sharing data with other healthcare organizations would be essential. During such sharing, if the data is not interoperable then data movement between disparate organizations could be severely curtailed. This could be due to technical and organizational barriers. This may leave clinicians without key information for making decisions regarding follow-ups and treatment strategies for patients. The healthcare providers will need to overcome develop a big data exchange ecosystem that provides trustworthy, timely, and meaningful information by connecting all members of the care continuum. Time, commitment, funding, and communication would be required before these challenges are overcome.

To develop a healthcare system based on big data that can exchange big data and provide us with trustworthy, timely, and meaningful information, we need to overcome every challenge mentioned above. Overcoming these challenges would require investment in terms of time, funding, and commitment.

5. Conclusion

Data is not only our most valuable tool, but it's a requirement for succeeding in the experience and transformation economies. By the year 2020, healthcare data will be doubling every 72 days, says Stephen Gold of IBM (Tailor 2016).

The world is changing and has become more digital in nature. Globally, there are more phones than people in the world. People use virtual assistants, and search the Web for any symptom of ill health. With each digital action, the digital footprint is created and saved in the cloud.

Like any other sector, the health sector has been receptive to the benefits of big data but, the adoption of this innovation is relatively slow as compared to other sectors like entertainment, transport, sports and education.

HEALTH SECTOR AND APPLICATION OF BIG DATA: A CASE STUDY OF INDIA

The paper has attempted to identify a few opportunities and the challenges in terms of application of the big data in the health sector.

The integration of big data in the healthcare system provides immense opportunities within many areas. For example, it can improve the wellness of the people, as the accessibility to health care becomes easy. It can also increase the engagement of the patient in terms of enhancing the communication between the caregivers and the patients. This would also enable to enhance the education of the patients. With more and more information from the patients, it would help to predict/forecast the diseases and the other health-related risks that are relevant in modern times. This would also enable the policymakers to a better disease management. The information would lead to insights into health planning with accuracy, financial planning and which would finally provide a better world to live in.

However, the experience says that the consequences of big data are not just limited to the health, but also have ethical implications. Therefore, the sector has to be cautious about the legal implications, which requires policy intervention.

References

- Akinagbe A., Peiris K.D.A., Akinloye O. (2018), Prospects of big data analytics in Africa healthcare system, "Global Journal of Health Science", vol. 10 no. 6, pp. 114-122.
- Alonso S.G., de la Torre Diez I., Rodrigues J.J., Hamrioui S., Lopez-Coronado M. (2017), A systematic review of techniques and sources of big data in the healthcare sector, "Journal of Medical Systems", vol. 41 no. 11, pp. 1-9.
- Belle A., Thiagarajan R., Sorousmehr S.M., Navidi F., Beard D.A., Najarian K. (2015), Big data analytics in healthcare, "BioMed Research International", <https://www.hindawi.com/journals/bmri/2015/370194/> [27.01.2022].
- Benjelloun F.Z., Lahcen A.A., Belfkih S. (2015), March. An overview of big data opportunities, applications and tools, in: 2015 Intelligent Systems and Computer Vision (ISCV). Conference proceedings, pp. 1-6.
- Cavanillas J.M., Curry E., Wahlster W. (2016), New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe, Springer Open, <https://link.springer.com/content/pdf/10.1007%2F978-3-319-21569-3.pdf> [27.01.2022].
- Chakrabarty A., Das U.S. (2020), Big data analytics in excelling health care: achievement and challenges in India, in: Big data analytics and intelligence: a perspective for health care, Emerald Publishing Limited, pp. 55-74.

Dash S., Shakyawar S.K., Sharma M., Kaushik S. (2019), Big data in healthcare: management, analysis and future prospects, "Journal of Big Data", vol. 6 no. 1, pp. 1-25.

Galetsis P., Katsaliaki K., Kumar S. (2020), Big data analytics in health sector: theoretical framework, techniques and prospects, "International Journal of Information Management", vol. 50, pp.206-216.

He W., Nazir S., Hussain Z. (2021), Big data insights and comprehensions in industrial healthcare: an overview, *Mobile Information Systems*, [https://www.hindawi.com/journals/misy/2021/6628739/\[27.01.2022\]](https://www.hindawi.com/journals/misy/2021/6628739/[27.01.2022].).

Healthcare Industry in India, <https://www.ibef.org/industry/healthcare-india.aspx> [31.01.2021].

Kaur P., Sharma M., Mittal M. (2018), Big data and machine learning based secure healthcare framework, "Procedia Computer Science", vol. 132, pp. 1049-1059.

Khanra S., Dhir A., Islam A.N., Mäntymäki M. (2020), Big data analytics in healthcare: a systematic literature review, "Enterprise Information Systems", vol. 14 no. 7, pp. 878-912.

Kumar S., Singh M. (2018), Big data analytics for healthcare industry: impact, applications, and tools, "Big Data Mining and Analytics", vol. 2 no. 1, pp. 48-57.

Mählmann L., Reumann M., Mehta N., Pandit A., Shukla S. (2019), Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study, "Journal of Biomedical Informatics", vol. 100, art. 103311.

Madanian S., Parry D.T., Airehrour D., Cherrington M. (2019), mHealth and big-data integration: promises for healthcare system in India, "BMJ Health & Care Informatics", vol. 26 no. 1, <https://informatics.bmj.com/content/bmjhci/26/1/e100071.full.pdf> [28.01.2022].

Muni Kumar N., Manjula R. (2014), Role of big data analytics in rural health care: a step towards svasth bharath, "International Journal of Computer Science and Information Technologies", vol. 5 no. 6, pp. 7172-7178.

Olivera P., Danese S., Jay N., Natoli G., and Peyrin-Biroulet, L. (2019), Big data in IBD: a look into the future, "Nature Reviews Gastroenterology & Hepatology", vol. 16 no. 5, pp. 312-321.

Panda M., Ali S.M., Panda S.K. (2017), March. Big data in health care: a mobile based solution, in: 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), IEEE, pp. 149-152.

Pramanik P.K.D., Pal S., Mukhopadhyay M. (2022), Healthcare big data: a comprehensive overview, in: Research anthology on big data analytics, architectures, and applications, Information Resources Management Association (ed.), IGI Global, Hershey, pp. 119-147.

Pastorino R., De Vito C., Migliara G., Glocker K., Binenbaum I., Ricciardi W., Boccia S. (2019), Benefits and challenges of big data in healthcare: an overview of the European initiatives, "European Journal of Public Health", vol. 29 supplement_3, pp. 23-27.

Reena D., Balvinder S., Kumar K.S. (2016), Opportunities and challenges of using big data analytics in Indian healthcare system, "Indian Journal of Public Health Research & Development", vol. 7 no. 4.

HEALTH SECTOR AND APPLICATION OF BIG DATA: A CASE STUDY OF INDIA

Raja R., Mukherjee I., Sarkar B.K. (2020), A systematic review of healthcare big data, "Scientific Programming", <https://www.hindawi.com/journals/sp/2020/5471849/> [28.01.2022].

Salas-Vega S., Haimann A., Mossialos E. (2015), Big data and health care: challenges and opportunities for coordinated policy development in the EU, "Health Systems & Reform", vol. 1 no. 4, pp. 285-300.

Taylor K. (2015), The patient revolution: how big data and analytics are transforming the health care experience, John Wiley & Sons, Hoboken.

Van Staa T.P., Goldacre B., Buchan I., Smeeth L. (2016), Big health data: the need to earn public trust, "BMJ", vol. 354.

Wyber R., Vaillancourt S., Perry W., Mannava P., Folaranmi T., Celi L.A. (2015), Big data in global health: improving health in low- and middle-income countries, "Bulletin of the World Health Organization", vol. 93, pp. 203-208.