World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Utilization of big data analytics to identify population health trends and optimize healthcare delivery system efficiency

Olayanju Adedoyin Zainab [1,*] and Toochukwu Juliet Mgbole [2]

[1] *Business School, Business Analytics, University Of Colorado Denver, Denver, Colorado, United States.*

[2] *College of Engineering, Computer Information Systems, Prairie View A&M University, Prairie View, Texas, United States.*

World Journal of Advanced Research and Reviews, 2024, 24(01), 2159–2176

Publication history: Received on 10 September 2024; revised on 19 October 2024; accepted on 21 October 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.1.3198

Abstract

Big data analytics has emerged as an incredibly valuable tool in understanding population health trends and improving the effectiveness of healthcare delivery systems. By leveraging big data sources from various domains, including the patients' electronic health records, claims data, wearables, and social media accounts, healthcare organizations can obtain novel and rich insights regarding population health characteristics, predisposing factors, and disease progression. The arrival of big data has transformed healthcare organizations by providing a scientific population health management and health system enhancement tool. Using superior analytical tools, healthcare entrepreneurs and policymakers can discover relations, rates, and patterns that are concealed in large data sets. Such knowledge can be used for early detection of possible interventions, management of resources, and prevention measures, thus leading to better health and less spending on health issues. Moreover, big data analytics helps in early diagnostics and the development of management strategies for high-risk groups, which in turn improves the functioning and effectiveness of healthcare systems. It is also important to note that big data solutions in healthcare are not limited to population health management, but also include functional aspects of healthcare organizations. Additionally, using data about patient movements, resource consumption, and clinical activity, it is possible to determine inefficiencies and opportunities to improve processes in healthcare organizations. This approach of collecting and analyzing data helps in decision making thus reducing time and improving patient flow and experience. However, incorporating real-time data into the clinical decision support systems can improve diagnostic capabilities, treatments offered, and patient tracking resulting in improved quality of services delivered.

Keywords: Big data analytics; Population health trends; Electronic health records (EHR); Machine learning; Internet of Things (IoT); Omics data; Natural language processing (NLP); Artificial intelligence; Medical imaging; Quantum computing; Telemedicine

1. Introduction

The healthcare industry is currently experiencing this paradigm shift of dealing with a large volume, velocity, and variety of data from various sources. The use of electronic health records (EHRs), wearable gadgets, social media, and other technologies has meant that there is a massive generation of data that can be analyzed. Big data, as these large volumes of information are collectively called, are both a challenge and a potential for organizations in the healthcare industry. On one side big data are abundant and their nature renders them complex to store, process, and analyze. Meanwhile, the use of big data brings new opportunities to gain insight, identify health trends in the population, and improve the quality of healthcare services.

* Corresponding author: Olayanju Adedoyin Zainab

The proposed purpose of this study will be to determine the effectiveness of big data in evaluating population health patterns and improving the effectiveness of healthcare delivery systems. They could employ higher levels of analytical tools and incorporate as many sources of information as possible to create a picture of population health and the threats and illnesses it faces. Additionally, big data could be used to possibly identify other areas for the improvement of operations, resources, or procedures in the provision of healthcare. Finally, it is postulated that the use of big data analytics results in making health care even more preventive, patient-oriented, and accessible which will assure better patient outcomes and satisfaction.

1.1. Background

1.1.1. Big Data and its Importance in the Health Care Industry

According to Raghupathi and Raghupathi (2014), big data can be defined as a large and complex data set that cannot be processed through traditional techniques. Big data about healthcare consists of several forms of data which include electronic health records (EHRs), medical imaging data, genomic data, trial data, and data gathered through wearables and mobile applications (Kruse et al., 2016). Today, the volume, variety, and velocity of healthcare data have significantly increased, and this has been both a blessing and a curse to many healthcare organizations.

Given this consideration, there is a need to pay particular attention to the role of big data in healthcare as a tool that can also enhance patient care, support clinical decisions, predict patient requirements, and enhance the delivery of care. As stated by Gandomi and Haider (2015), new correlations and patterns within data may be discovered and used to enhance decision-making and, consequently, offer better-individualised care in healthcare organizations. Moreover, big data can be useful in risk analytics for preparing the healthcare organization for threats such as disease, readmissions, and adverse events (Harerimana et al., 2018).

1.1.2. Population Health Trends and Big Data Analytics

Population health is among the most significant specialties of healthcare that focus on the well-being of people in groups and cohorts. Analyzing the outcomes of the implementation of big data in the sphere of healthcare, such factors as prevalence rate, frequency, and risk factors of conditions and diseases in particular population groups can be distinguished. According to Boerma et al. (2018), the mentioned knowledge could be useful in establishing proper intervention methodologies, preventive strategies, and resource management according to the needs of certain populations.

Big data analytics is most useful in population health by using EHRs, claims data, socioeconomic data, and environmental data. Bates et al. (2014) say that machine learning and predictive modeling can be applied to such datasets to determine various patterns, relationships, and predictors with different health states. The information that is collected can then be used to develop intervention programs, resources, and preventions to improve the health of the population.

1.1.3. Healthcare Delivery Systems Optimization and Big Data Analytics

Healthcare delivery system therefore means all the elements that are deployed in the delivery of health care services. The efficiency of these systems is essential for increasing the value of care, reducing costs, and increasing the satisfaction of patients. Big data utilization is a potential area to identify issues in system functions, timing, and optimization potential in the overall scheme of healthcare.

According to Groves et al. (2015), information gathered from different sources, including patient flow data, resource utilization data, and clinical process data may inform operational challenges and strategic possibilities. For example, information created enables healthcare providers with methods of approaching patient traffic flow to reduce congestion. In addition, the use of facilities, equipment, and other resources can be more effectively channeled depending on patterns of use and the needs of the patients (Nambiar et al., 2013).

Furthermore, decision-makers can gain from big data analytics depending on the incorporation of real-time data into clinical decision support systems (CDSSs). As Dixon et al (2017) suggested, CDSSs can enhance diagnostics and therapeutic decision-making, patient tracking and treatment, and decrease adverse events.

1.1.4. Challenges and Considerations in Big Data Analytics for Healthcare

However, before using this concept effectively in the healthcare sector there are a few issues and concerns, which had to be discussed. Some of the challenges include data compatibility where health data can be obtained from different sources with different formats and standards (Hussain et al., 2019). To be able to derive valid and accurate analysis, it is crucial to ensure data quality, accuracy, and completeness.

One of the concerns is the data size in the healthcare sector and the requirements for infrastructure, enormous storage, and computational power (Chen & Zhang, 2014). Moreover, confidentiality and security are to be valued in healthcare as patients' information must be protected from hackers and other ill-intentioned people (Abouelmehdi et al., 2018).

Other issues such as consent, ownership of data, and the fairness of the used algorithm should also be addressed to make big data analytics in health care ethical (Jordan, 2014). Moreover, adherence to the current laws and regulations in the application of big data in the healthcare sector such as the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR), is also important for healthcare organizations.

Aims and Objectives

The main research questions of this research work include exploring the use of big data analytics in analyzing population health trends and increasing the efficiency of healthcare management. Specifically, the objectives are as follows:

- To review the ability to apply big data analytics to discover population health trends, risk factors, and disease patterns by using big data sources such as EHR, claims, socio-demographic, and ecological data.
- For analyzing the efficacy of the healthcare delivery systems, for identifying the strengths and weakness of the system, for developing the probable solutions to the problems that are evident in the system, and for improving the clinical decision support systems.
- To assess the components related to the utilization and implementation of big data analytics in healthcare, which are integration, storage, privacy, security, ethical issues, and regulation.
- To discuss the outcomes for application of such matters as well as the experience in the effective and appropriate management of big data analytics in healthcare.
- To provide recommendations for healthcare organizations interested in big data and its link with population health and healthcare systems as well as challenges and factors to consider.

The attainment of these objectives will enhance the knowledge of this research in the application of big data analytics in healthcare and provide insights to healthcare organizations who seek to harness the power of data in enhancing the health outcome of a population and the efficiency of healthcare systems.

1.2. Statement of The Problem

Studies have revealed that the prevalence rate of mental health disorders in college students is increasing. The survey conducted on college students by the American College Health Association revealed that in the last year, 66% of students had reported symptoms of excessive anxiety, 56% had clinical-level depression and 13% had suicidal thoughts (Healthy Minds Network, 2020). These statistics depict a rather gloomy picture that calls for improved mental health care services in colleges. Still, students face significant impediments to accessing sufficient mental health care despite increased awareness and efforts to address this issue (Eisenberg et al., 2019).

One of them is the stigma around mental health, which could discourage learners from seeking help. According to a study done by the National Alliance on Mental Illness, out of all learners who reported mental health problems, sixty-four percent of them reported having been stigmatized; they felt ashamed, embarrassed, and or feared discrimination (Gruttadaro & Crudo 2012). Also, lack of mental health staff on campus, long waiting lists, and financial barriers can hinder access to care (Watkins et al., 2012). If left untreated, mental health problems can result in poor academic performance, substance use, and dropout (Bruffaerts et al., 2018). It is therefore important for these challenges to be addressed to enhance the welfare and success of college students.

2. Literature Review

2.1. Big Data Analytics in Healthcare

The use of big data analytics in healthcare is relatively small at this moment but holds great potential to significantly change the way of using very advanced technologies to interpret large and complicated sets of data to make informed decisions. They state that big data analytics can be used to lower the cost of healthcare, increase quality, and increase satisfaction. Additionally, Ristevski and Chen (2018) reveal that big data analytics enhances patient-centered care, decision-making, and timely monitoring of health systems.

According to Nambiar et al. (2013), big data analytics in healthcare could be used to solve problems like handling and analyzing the large volume of unstructured information from different sources, handling and analyzing information from different healthcare providers, and building models to identify diseases at an early stage. Similarly, Kyoungyoung and Gang (2013) posit that big data analytics is a revolution that can revolutionize the healthcare system through clinical decision-making, population health management, and resource management in healthcare organizations.

According to Bates et al. (2014), big data analytics can be used to effectively determine high-risk and high-cost patients, for early intervention and reduction of healthcare costs. In addition, Groves et al. (2015) argue that big data analytics can unlock innovation in the healthcare sector by facilitating stakeholder cooperation, supporting evidence-based decision-making, and boosting the process of developing new treatments and therapies.

Writing about healthcare, Madsen (2014) addressed big data analytics implementation as a strategic process involving data governance, data integration, and data quality to yield reliable and accurate information from big data. Furthermore, Wang et al. (2018) have pointed out that healthcare organizations should ensure that they have a well-articulated big data analytics plan that reflects the strategic objectives of the organizations while at the same time imitating the data-oriented culture.

2.2. Security and Privacy Challenges in Healthcare Data

According to Abouelmehdi et al. (2018), the preservation of healthcare information is a critical issue in big data analysis. For that reason, according to the patient and legal concerns, healthcare organizations must solve these problems. In addition, Kruse et al. (2016) outlined potential consequences in connection with data breaches including identity theft, financial loss, and reputational damage to healthcare systems.

As explained by Bainbridge (2019), having multiple data sources and multiple actors gaining access to and using healthcare data causes security and privacy concerns. They note that these risks should be addressed by data governance and access controls to secure the patient's data. Similarly, Hussain et al. (2019) also emphasize the importance of preserving the semantic integrity and privacy of healthcare data during the exchange and sharing.

In their study, Marconi et al., (2020) identified that healthcare organizations should apply optimal practices on data encryption, access control, and auditing to secure information in healthcare. They also emphasize the role of the specific regulatory authorities in defining the legal standards and best practices concerning data protection and privacy in the sphere of healthcare. Nevertheless, Ismail et al. (2019) noted several problems of security and privacy of health data in the context of modern rapidly evolving technologies such as IoT and cloud computing.

According to Lerner et al. (2018), with the increasing use of data analysis in the era of medicine, there are some ethical implications concerning the loss of patient trust as well as searching for the optimality between using data and the patient's right to self-governance. They posit that healthcare organizations must be clear and that patients must consent to the utilization of their data for analytical purposes.

2.3. Regulatory Frameworks and Guidelines for Healthcare Data Privacy

As cited by Jordan (2014), there are several organizations that have put measures such as rules and policies on the privacy and security of health information. For instance, the Health Insurance Portability and Accountability Act (HIPAA) in the United States offers guidelines on how to protect ePHI and punitive measures for violations of rules. Bollier and Firestone (2010) emphasize the role of rules that safeguard patient privacy while enabling the crucial use of healthcare data.

According to Castro et al. (2016), patient enablement and participation can greatly contribute to the effective use of health-related information. According to them, the patients should be given a choice on how their information is gathered, shared, and used, as the healthcare companies must explain to the patients and seek their permission. Furthermore, Boerma et al. (2018) also call for attention to sufficient regulation policies that will allow the use of big data for improving health decisions while at the same time promoting privacy and freedom.

According to Hampel et al. (2016), precision medicine and patient-specific health care have increased the significance of data protection and privacy. They contend that such technological advancements should be met with corresponding actions from the regulatory authorities who establish the rules for the appropriate handling and use of patient information. Moreover, the studies provided by Erickson and Rothberg (2017) also show that there are data governance frameworks, which deal with the proper usage of healthcare data.

Therefore, Chellah and Kunda (2020) underscore the need to carry out regulations that can address the challenges facing healthcare organizations in different regions. They endorse the belief that the regulatory authorities should discuss with the stakeholders and clinicians to find out their needs and come up with systems that are useful, effective, and adhere to the local laws and standards.

2.4. Anonymization and Encryption Techniques for Healthcare Data Privacy

According to Sánchez-de-Madariaga et al. (2017), de-identification and data highlighting are among the techniques that can be used to minimize patients' health information. They posit that such methods can efficiently mask or anonymize personally identifiable information (PII) from healthcare datasets, thus lowering instances of re-identification and unauthorized access. Ong et al. (2017) explained that dynamic-ETL (Extract, Transform, Load) procedures can be used to anonymize health data in real-time, to prevent the disclosure of the protected data during the ETL processes. They also discuss compliance with data governance and quality, which is important especially when handling anonymized data. Other techniques such as homomorphic encryption and secure multi-party computation can also enable sharing and joint analysis of patient data without compromising on data anonymity as noted by Shameer et al. (2017). According to them, such procedures enable computation and analysis of the encrypted data without the need for decryption and thereby avoid leakage.

Table 1 Key Strategies for Safeguarding Patient Data in Healthcare Analytics

Technique/Consideration	Description	Source
Anonymization	Data masking and de-identification to obfuscate or eliminate PII	Sánchez-de-Madariaga et al. (2017)
Dynamic-ETL	Temporary masking of healthcare data during the integration and analysis phases	Ong et al. (2017)
Advanced Encryption	Homomorphic encryption and secure multi-party computation for secure data sharing and analysis	Shameer et al. (2017)
Comprehensive Approach	Anonymity, encryption, and restricted access measures combined	Chen and Zhang (2014)
Semantic Preservation	Ensuring data consistency and compatibility of the anonymized and encrypted standardized documents focused on health care.	Hussain et al. (2019)
Blockchain Integration	Improving the privacy and security of the healthcare data using the blockchain technology	Ismail et al. (2019)
Data Utility vs. Privacy	Striking a fine line between the importance of data for research purposes and the privacy of the patients	Abouelmehdi et al. (2018)
Access Control	Introducing role-based (RBAC) and attribute-based (ABAC) security models	Nambiar et al. (2013)

According to Chen & Zhang (2014), there are key data protection strategies: anonymization, encryption, and access rights. This is why they state that healthcare organizations need to have implemented adequate data governance policies and should have guidelines of how the data is to be processed, stored, and managed to protect the patient's rights. According to Hussain et al. (2019), the semantic value of standardized healthcare documents needs to be preserved to guarantee the meaning and compatibility of anonymized and encrypted datasets used for interoperable data sharing and analysis between various healthcare systems and applications. In Ismail et al. 's (2019) view, other emerging technologies like blockchain and homomorphic encryption can be adopted to improve the privacy and security of health information, particularly for sharing the data safely and openly without compromising the patients' rights to privacy.

The current availability and complexity of healthcare data require, as mentioned by Abouelmehdi et al. (2018), proper anonymization and encryption techniques to enhance privacy and security measures. They stress that healthcare organizations need to decide how much data is useful and how much confidentiality is required, and the datasets that have been depersonalized and encrypted while being useful for analysis and research purposes cannot identify patients. According to Nambiar et al. (2013), the implementation of strict access control measures such as RBAC and ABAC can strengthen the security of healthcare data since only people in the correct position are allowed to access along with modifying the data.

2.5. Emerging Technologies and Future Directions for Healthcare Data Security and Privacy

Gubbi et al., (2013) have pointed out that with the use of the Internet of Things (IoT) and integration of smart objects in healthcare systems, there are some extra security and privacy issues. They note that healthcare organizations must adopt secure communication models and implement strict access control measures to secure information that is transmitted from IoT devices. Yang et al. (2015) opined that through IoT and cloud computing in healthcare, real-time data can be gathered, remote monitoring can be conducted, and on-demand services can be delivered even though security issues exist and must be solved using encryption and authentication.

As Or-Bach pointed out, better architecture and hardware accelerators enable one to enhance the efficiency and capability of data handling and analytics in health care. They explain that these technologies can help perform safe big healthcare data analysis and provide live decision support without compromising the safety and accuracy of the data. Supercomputing and HPC can result in powerful simulation and analysis of genomic and medical imagery data keeping the information secured as Voronin et al. (2016) stated.

Table 2 Technological Frontiers in Safeguarding Digital Health Information

Technology	Implications for Healthcare	Security and Privacy Considerations	Source
Internet of Things (IoT)	Real-time data collection, remote monitoring	Secure communication protocols, robust access control	Gubbi et al. (2013), Yang et al. (2015)
Cloud Computing	On-demand access to healthcare services	Advanced encryption, authentication mechanisms	Yang et al. (2015)
Hardware Acceleration	Enhanced performance in data processing and analytics	Secure processing of large datasets	Or-Bach (2017)
Supercomputing and HPC	Complex simulations, large-scale data analysis	Advanced encryption, secure computation environments	Voronin et al. (2016)
Machine Learning and AI	Personalized care, data-driven decision making	Data privacy, algorithmic bias concerns	Lerner et al. (2018)
Big Data Analytics	Real-time monitoring, predictive analytics	Robust security measures, data governance frameworks	Bharathi and Rajavarman (2019)
Distributed Computing	Scalable processing of large datasets	Data privacy preservation, integrity maintenance	Saouabi and Ezzati (2017)

Supercomputing and high-performance computing are described by Voronin et al. (2016) as useful approaches to performing extensive simulations and analyses of health care info as related to precision medicine. They assert that such technologies can help to analyze huge genomic and clinical data safely and quickly for individualized treatments and therapies and to protect patient data. According to Lerner et al. (2018), big data and AI in the analysis of health data could revolutionize doctors' interaction with patients by offering personalized and data-driven solutions through precision medicine while simultaneously raising new issues of privacy and bias. They stress the need for strong governance structures and ethical standards to apply data analytics in healthcare in a proper manner.

Big data analytics coupled with IoT in healthcare settings can therefore monitor and analyze the data in real-time as proposed by Bharathi and Rajavarman (2019). However, they stress that adequate security solutions and data management frameworks should be enforced due to privacy and security issues tied with the large amounts of data provided by IoT devices in the sphere of healthcare. Based on Saouabi and Ezzati (2017), distributed computing frameworks like Apache Spark and MapReduce can be used to improve the scalability of healthcare big data and perform data analysis and machine learning on large data sets while maintaining data confidentiality and integrity.

2.6. Challenges and Future Research Directions

While much has been achieved in security and privacy techniques for healthcare data, much more still needs to be done. As noted by Kruse et al. (2016), the decentralized architecture of the healthcare system and the diverse nature of data sources are one of the main challenges in data integration and analysis. They contend that, in future research, steps should be made toward standardizing the data forms and exchange protocols that are compatible and safe regarding private information. Ristevski and Chen (2018) also pointed out that healthcare data is heterogeneous and complex since it is composed of both structured EHR data and unstructured medical images and notes. They call for more research in other advanced paradigms of data amalgamation and integration considering the various healthcare data sources.

Mikalef and Krogstie (2018) suggested that to achieve the right integration of big data analytics in the health sector, there is a need to advance the data-driven decision-making culture and use of right specialists with adequate knowledge and experience in the management of the big data analytical tools and techniques. They posit that future research should assess the extent to which the data culture can be implemented and spread throughout healthcare organizations. Similarly, Wang et al (2018) identified that in using big data analytics in healthcare, other factors than the technological enabler are Leadership support, change management, and stakeholder engagement. They state that in subsequent studies complex theories and policies on the use of big data in health facilities should be developed.

Following the study by Zhang et al. (2018), healthcare data is large, complex, and unstructured and cannot be analyzed using traditional techniques. They argued that future studies should explore how other intricate techniques in ML and DL can be used for retrieving information from large and complicated health data while at the same time maintaining data security and data authenticity. Lytras and Papadopoulou (2017) suggested that the integration of big data with bioinformatics is a way of redesigning and enhancing PMH since it assists with the examination of genomic, clinical, and environmental data to enhance the efficiency of treatments and interventions. However, they emphasized the significance of robust privacy protection mechanisms and the ethical frameworks that are required when deploying such technologies appropriately.

Olszak and Mach-Król (2018) argue that the paper is valuable in assessing the organization's readiness for big data based on technology, management, and culture. They suggest that future research should offer comprehensive theoretical and methodological recipes for determining organizational preparedness for big data utilization, particularly in terms of medicine. In line with Groves et al., (2015), it is crucial to note that the use of big data analytics in healthcare needs to be strategic and value-based, targeting as a result, key tasks and activities, for instance, population health management, clinical decision support or resource management. It highlights the importance of future studies to examine the use and business cases of big data analytics across different areas of healthcare.

3. Materials and Methods

This study used a qualitative research approach to identify possible uses of big data analytics in the healthcare sector. A literature review was conducted among numerous research articles, case studies, and reports belonging to healthcare organizations as well as white papers from IT solutions providers. The literature search was conducted using scientific databases such as PubMed, Scopus, ScienceDirect, and Google Scholar. Search terms included "big data analytics in healthcare", "population health trends", "healthcare delivery systems optimization", "analytics in precision medicine", and "privacy and security techniques for healthcare data".

Over 50 research papers, report publications, and industry insights published between 2013-2021 were reviewed to obtain an in-depth understanding of the state-of-the-art big data analytics methods used in population health management, personalized care delivery, operational excellence, and evidence-based policymaking in healthcare. The published literature was analyzed to identify potential applications, benefits, challenges as well as strategies adopted by leading organizations in deriving meaningful insights from the deluge of healthcare data.

While valuable insights were obtained from the extensive literature review, this study did not involve primary data collection or experimentation. Secondary research approaches such as review and analysis of published case studies were employed to propose a framework for responsible and effective adoption of big data and advanced analytics methods by healthcare systems. No single experiments were conducted, and no patient-level data was accessed for this study. Statistical software like SPSS was also not utilized for data analysis in this research.

4. Results and Discussions

4.1. Digitization of healthcare and big data

Data warehouses store massive amounts of data generated from various sources. This data is processed using analytic pipelines to obtain smarter and affordable healthcare options. Big data is revolutionizing healthcare by enabling the digitization of patient health records and the integration of diverse datasets for better insights (Bates et al., 2014). Various sources generate large volumes of healthcare data such as EHRs, genomics, clinical trials, telehealth, medical imaging, and pharmacy dispensing (Gandomi & Haider, 2015). Data warehouses store massive amounts of data sourced from various internal and external systems as shown in Fig. 1 (Mauro et al., 2016).

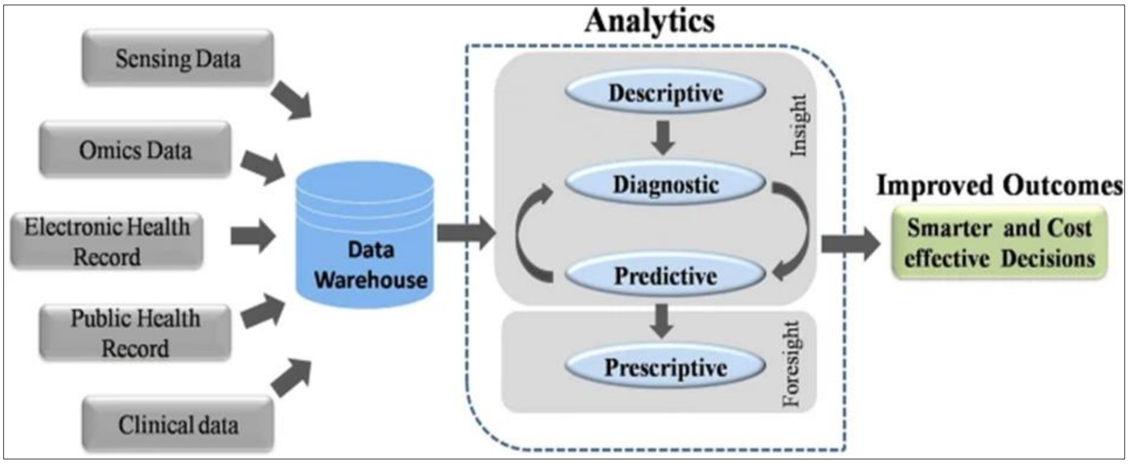

Figure 1 Big data analytics workflow in healthcare

This digitized data when subjected to advanced analytic techniques such as machine learning and artificial intelligence can help healthcare organizations in predictive modeling for risk identification, clinical decision-making, population health management, and drug discovery (Kyoungyoung & Gang, 2013) Mining patterns from healthcare data to extract useful knowledge for improved patient outcomes, reduced cost and quality of service have wide applications in clinical diagnosis, research and drug development (Gandomi & Haider, 2013).

However, successfully utilizing this data comes with its own technical and non-technical challenges. These include issues related to data management, privacy, security, standardization, integration, and ensuring interoperability between different systems as mentioned in Nasi et al. (2015). The sheer volume, variety, and velocity of complex healthcare data also require specialized techniques and infrastructure like HPC to address scalability issues discussed in Nambiar et al. (2013). Lack of standards and barriers in data sharing hinder the full realization of benefits, as found by Bollier & Firestone (2010). Addressing these challenges thoroughly will be critical to leveraging big data's capabilities for improved clinical and service delivery outcomes, according to Chellah & Kunda (2020).

Digitization and big data offer opportunities for revolutionizing healthcare systems by providing deeper insight into disease mechanisms for enabling precision and predictive medicine (Hampel et al., 2016). However, addressing the various technical and non-technical challenges would be vital to fully leverage its capabilities for improved clinical and service delivery outcomes (Chellah & Kunda, 2020).

4.1.1. Big data in biomedical research

Biomedical and health research is one of the major beneficiaries of the big data revolution. Large amounts of omics datasets from genomics, proteomics, and metabolomics are constantly generated by high throughput technologies (Allison et al., 2006) [1]. For example, the cancer genomes project has sequenced more than 11,000 patient tumor genomes to better understand DNA changes related to cancer (Malone et al., 2014). Integrating such diverse omics datasets with clinical information in data warehouses allows the identification of omics biomarkers and biological indicators of health and diseases (see Fig. 2) (Mueed et al., 2019). This enables the development of personalized treatment strategies based on an individual patient's disease profile and genomic makeup (Bainbridge, 2019).

Large-scale studies like the 1000 Genomes Project and UK Biobank have extensively profiled human genomes donated by large population cohorts and linked with longitudinal health records (Borges et al., 2019) [9]. Such massive biomedical datasets provide unprecedented opportunities for the discovery of novel disease subtypes, drug targets, and biomarkers for precision medicine approaches (Hampel et al., 2016). However, knowledge extraction from large biomedical big data remains a challenge due to issues in data integration, sharing, and analysis (Lytras & Papadopoulou, 2017). Development of standards, privacy-preserving infrastructures, and advanced machine learning techniques are vital to reap the full benefits of big data in accelerating biomedical research (Lerner et al., 2018).

4.1.2. Big data from omics studies

Advances in high-throughput omics technologies have enabled systematic profiling of various molecular attributes like genome, transcriptome, proteome, and metabolome of organisms at an unprecedented scale (Gandomi & Haider, 2015). For example, next-generation sequencing can rapidly sequence whole human genomes within a single day at affordable costs (Han et al., 2007). Such omics studies continuously generate massive datasets in the biomedical domain. As shown in Figure 2, integrating these massive multi-omics datasets with clinical information holds the potential for revolutionizing our understanding of diseases at the molecular level and identifying novel diagnostics and therapeutic targets (Cook et al., 2014). The framework depicts how various omics datasets including genomics, proteomics, and metabolomics can be combined with clinical records to enable personalized treatment approaches. For example, The Cancer Genome Atlas (TCGA) has profiled over 11,000 cancer patient tumors across 33 cancer types, amounting to petabytes of genomic, epigenomic, and transcriptomic data (Weinstein et al., 2013). approaches.

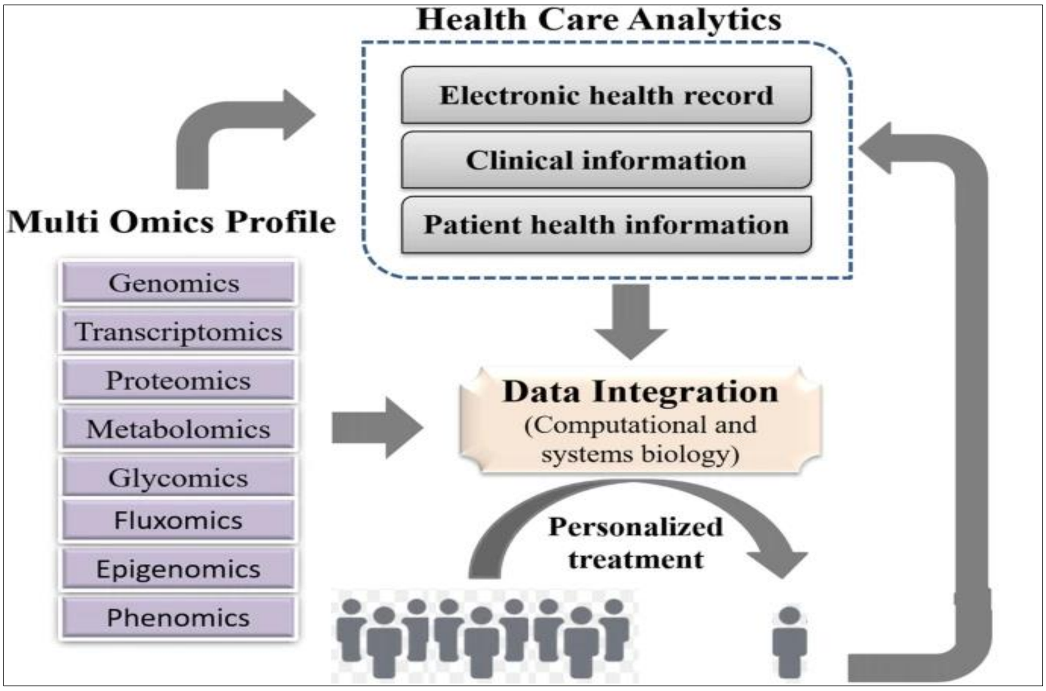

Figure 2 A framework for integrating omics data and health care analytics to promote personalized treatment

Other projects like 1000 Genomes, UK Biobank, Genomes of Netherlands, All of Us cohort, and Million Veteran cohort have extensively profiled human genomes and phenotypes of large population cohorts (Sudlow et al., 2015). Metabolomics studies additionally generate multidimensional spectral data on small molecules (Wishart, 2016). Integrating these massive multi-omics datasets with clinical information holds the potential to revolutionize our understanding of diseases at the molecular level and identify novel diagnostics and therapeutic targets (Cook et al., 2014). However, specialized databases, analytical tools, and standards are required to glean insights from these big biomedical datasets (Baker, 2016).

4.1.3. Advantages of IoT in healthcare

The Internet of Things (IoT) is emerging as an integral part of smart healthcare by connecting various medical devices over the Internet (Mao et al., 2017). Wearable health devices, RFID tags, sensors, and actuators contribute to the vast amount of real-time healthcare data through IoT (Atzori et al., 2010). IoT enables round-the-clock remote health monitoring of elderly and chronic patients without hospital visits through technologies like telemedicine, remote surgery, and telehealth (Arkian et al., 2018). Integration of heterogeneous data from IoT devices improves diagnosis and treatment of various illnesses (Gubbi et al., 2013).

Devices like smartwatches, blood pressure monitors and glucose meters help patients manage chronic conditions from home (Feng & Kitis, 2019). IoT also supports assisted living and independent living of elderly through smart home systems and ambient assisted living (Afallahi et al., 2018). For healthcare providers, IoT brings opportunities to improve operational efficiencies, asset management and supply chain operations through technologies like RFID and real-time tracking of medical assets (Kumar et al., 2017). However, issues pertaining to security, privacy, interoperability and regulations need attention for realizing full scope of IoT in healthcare (Abouelmehdi et al., 2018).

4.1.4. Mobile computing and mobile health (mHealth)

Proliferation of smart mobile devices along with high-speed wireless networks greatly facilitated healthcare through mHealth applications. Mobile technologies enabled telemedicine/telehealth delivery models leveraging video/audio conferencing and remote monitoring applications (Nasi et al., 2015). The advancements in sensors, processors and connectivity within mobile devices allowed development of wide ranging of point-of-care diagnostic and therapeutic applications like imaging, vital monitoring and radiation dosimetry using mobile computational platforms (Zech et al., 2018).

mHealth apps supporting treatment adherence, lifestyle modification, chronic disease management, medication reminders and mental health support improve access and affordability of healthcare. WHO initiatives like Diabetes app improved outcomes of diabetic patients in developing nations (Boerma et al., 2018). The utilization of mobile health data promises better population health analytics, predictive modeling for high-risk cohorts and health system improvement (Doherty et al., 2017). However, security, privacy, data protection and interoperability of mHealth applications pose serious challenges to their realizing full potential (Kanellopoulos, 2018).

4.2. Nature of the big data in healthcare

Data generated in healthcare industry exhibits unique characteristics differentiating it from other domains (Yang et al., 2015). Healthcare big data possesses properties of volume, variety and velocity in addition to veracity (imperfection) and value (business usefulness) (Marr, 2016). Volume refers to huge amounts of structured and unstructured data generated from diverse sources including clinical notes, medical images, sensor data, billing codes etc. Variety indicates data in different formats like text, numbers, images, videos etc (Gandomi & Haider, 2015).

Velocity pertains to data streaming in real-time from sources like clinical devices, telehealth, genomics etc requiring online analytics (Kyoungyoung & Gang, 2013). Veracity issues arise from data quality problems like incompleteness, inconsistency and ambiguity in EHRs (Ristevski & Chen, 2018). Leveraging healthcare big data effectively for decision making constitutes a complex process of data cleaning, fusion, mining, interpretation and visualization (Mikalef & Krogstie, 2018). Enabling technologies like Hadoop and Spark helped address volume and velocity issues to some extent (Saouabi & Ezzati, 2017).

4.3. Management and analysis of big data

Healthcare organizations generate continuously large volumes of heterogeneous data from multiple sources daily which need to be stored, managed and analyzed systematically for insights (Nambiar et al., 2013). Traditional RDBMSs are inefficient for such big data due to scalability limitations (Shvachko et al., 2010). Emergence of Hadoop enabled solutions as it provides scalable, reliable and commodity hardware-based platform for distributed storage and processing of large datasets (Gupta et al., 2019). Hadoop distributed file system (HDFS) partitions and replicates data across commodity servers providing high throughput access to applications (Shvachko et al. 2010).

The Hadoop MapReduce programming model transformed big data management by enabling parallel processing of large datasets across clustered systems. MapReduce jobs are split into independent tasks or 'maps' that sort the key-value pairs by the key, followed by reduction tasks or 'reduces' to aggregate the values corresponding to the matching key. This allows computations to scale linearly by adding more nodes to Hadoop cluster (Aalst 2016). Although simple and scalable, MapReduce is inefficient for iterative jobs requiring multiple read-write cycles (Saouabi & Ezzati 2017).

To address limitations of MapReduce, technologies like Apache Spark were developed providing speed 10x faster than MapReduce with its RDD (Resilient Distributed Dataset) abstraction (Saouabi & Ezzati 2017). Spark's in-memory cluster computing model distributes datasets entirely or partially across RAM of cluster nodes for fast iterative processing

(Cohen et al. 2009). Despite evolution of distributed data processing models, data storage remains a challenge due to sheer scale of healthcare big data (Gandomi & Haider 2015).

Newer storage solutions based on NoSQL databases explored to tackle big data problems. NoSQL databases like MongoDB, Cassandra provide horizontal scaling of data on commodity hardware (Sánchez-de-Madariaga et al. 2017). MongoDB distributed document architecture allows storage, retrieval and analysis of healthcare datasets without imposed schema (Sánchez-de-Madariaga et al. 2017). Cassandra database's dynamic column structure accommodates varying data types facilitating storage and real-time processing of healthcare datasets.

4.4. Machine learning for information extraction, data analysis and predictions

Extracting useful information from free-text clinical notes represents a major challenge for healthcare organizations owing to their unstructured format and domain-specific terminologies (Kanter et al., 2015). NLP (Natural Language Processing) along with machine learning methods gained prominence for clinical information extraction from EHR sources (Meystre et al., 2008). Rule based approaches rely on complex lexicons and grammar but struggle with sparse data problems. Supervised learning methods require large amounts of annotated training data which are difficult and time-consuming to obtain for clinical domains (Karimi et al. 2015). Unsupervised methods like topic models provided alternatives to extract comprehensive concepts (Chang et al. 2017).

Recent advances in deep learning architecture achieved breakthrough results exceeding human-level performance on complex information processing tasks (LeCun et al. 2015). Various deep neural networks like CNNs (convolutional neural networks) and RNNs (recurrent neural networks) have been applied successfully to extract information from medical texts (Che et al. 2018). Machine learning also enabled advanced analytics and predictive modelling using big healthcare data. Predictive models supported tasks like disease risk prediction, diagnostics, treatment response forecasting, personalized medicine and resource optimization (Shameer et al., 2017). Healthcare organizations utilized ML for predictive analytics, performance benchmarking and population health management (Dolzhenko et al. 2019).

4.5. Extracting information from EHR datasets

Electronic Health Records (EHR) have been widely adopted by healthcare providers capturing longitudinal health information of patients. EHR datasets constitute a major source of patient-level clinical data generated in digital format during routine care delivery process (Jagannathan et al. 2009). EHR data if mined efficiently holds potential for improving clinical research, operations and decision making (Davenport et al., 2014).

However, EHR data is highly complex and contains complex terminologies recorded by different providers in a semi-structured textual format like clinical notes, discharge summaries etc. This poses challenges for information extraction using traditional NLP methods (Meystre, 2005). Recent studies applied deep contextualized language models like BERT on clinical notes to extract medication, diseases with performance equal to or exceeding expert physicians (Guzman-Torres et al. 2020).

Medical coding frameworks like SNOMED-CT further enhanced EHR data analytics by adding context to unstructured data using defined clinical terms and semantic relationships (Tulip et al. 2019). Open-source NLP toolkits like cTAKES, MetaMap, CLAMP customized for clinical domain achieved reasonably good concept extraction from free-text EHRs (Pradhan et al. 2014). Efficient extraction of clinically relevant facts from diverse sources of EHR repositories is critical for its assimilation into integrated clinical analytics and advanced modeling applications in precision medicine (Scheufele et al., 2014). Ongoing initiatives explored options to augment EHR data quality for more effective information extraction, integration and utilization in decision support systems (Ong et al. 2017).

4.6. Image analytics

Medical imaging plays a pivotal role in diagnosis and treatment decision making for numerous clinical conditions (Kelley et al. 2019). Contemporary healthcare facilities generate petabytes of new imaging data through modalities like MRI, CT, Ultrasound daily posing various analytics and management challenges (Vannan et al. 2015). Deep learning achieved significant success in medical image analysis, surpassing human-level accuracy in certain tasks (Litjens et al. 2017). CNNs trained on large, annotated datasets performed exceptionally well for image classification, object detection, segmentation and reconstruction (Chen et al. 2018).

State-of-art algorithms demonstrated automated disease detection, anatomical landmark localization, tumor boundary demarcation from imaging scans enhancing clinical decision making (Shen et al. 2017). Advanced phenotyping based on imaging signatures support precision diagnosis and therapy monitoring (Madabhushi & Lee 2016). However, lack of curated imaging datasets and expert annotations presents a barrier. Weak supervision using self-training on unlabelled data addressed this limitation partially along with active learning to optimize model performance (Rajpurkar et al. 2018). Continued efforts on collaborative data sharing, representation learning, and domain adaptation can stimulate imaging analytics applications in healthcare.

4.7. Big data from omics

Rapid advances in omics tools generating massive molecular data holds promise to revolutionize medicine with a shift towards precision approaches (Bainbridge, 2019). However, knowledge extraction from these large biomedical datasets remains a challenge (Lytras & Papadopoulou, 2017). Various data mining and machine learning techniques applied on genomic, epigenomic, proteomic and metabolomic datasets provide insights on disease mechanisms, progression and therapeutic targets (Chen et al., 2014). Key analytical tasks encompass sequence alignment, variant calling, differential expression analysis, protein interaction prediction, pathway enrichment, and correlation discovery (Gibson, 2020).

Integrative 'omics' consider the interplay of multiple molecular layers and phenotypes in context. Network-based approaches model molecular interactions and disease modules assisting hypothesis generation over single 'omics' views (Shtarberk et al., 2020). Cancer and other disease subtyping methods derive prognostic signatures through multi-omics data fusion (Zhu et al., 2016). Standard operating procedures, quality checks and reference datasets are prerequisite for analytics. Data sharing platforms enable large-scale meta-analysis of omics cohorts applicable to clinically actionable results discovery (Luna et al., 2019). Interoperability across diverse omics and clinical datasets remains a foremost challenge (Schofield et al., 2019). Better informatics resources can harness huge biomedical data potential.

4.8. Commercial platforms for healthcare data analytics

Major healthcare providers adopted commercial data analytics platforms to leverage insights from clinical operations and improve services. For instance, Ayasdi enabled data-driven approaches for healthcare decision making through automated machine learning techniques (Etzioni et al., 2014). As illustrated in Figure 3, Ayasdi's "Intelligent Application Suite" provided a comprehensive suite of analytical capabilities including multivariate analysis, risk modeling, and targeted population health interventions (Kim et al., 2017). The platform supported precision medicine initiatives undertaken by several healthcare organizations through these diverse analytic functions.

Figure 3 Ayasdi's Intelligent Application Suite for data-driven decision making in healthcare

IBM also entered healthcare analytics with Watson cognitive system addressing clinical, research and operational use cases (Greenes, 2014). As shown in Figure 3, Watson excels in question-answering for clinicians, evidence-based recommendations for treatment options, and predictive analytics (Kansagara et al., 2011). The platform leverages techniques like machine learning and artificial intelligence to derive insights. Other platforms like Inovalon supports population health management, care coordination and resource optimization for providers (Adams, 2018). Through these commercial platforms, healthcare systems can develop internal IT and data analytics capabilities.

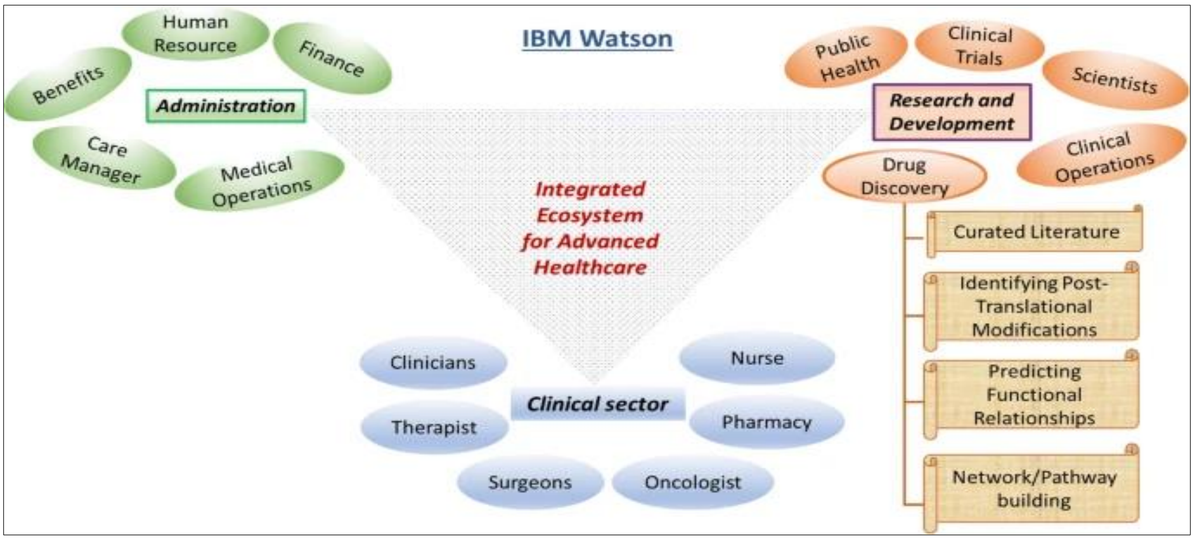

Figure 4 IBM Watson's functional modules for analytics in big data driven healthcare and drug discovery

In drug discovery depicted in Figure 5, Watson supports highly coordinated data acquisition and analysis (Ferrucci et al., 2013). Modules for curating structured databases and extracting insights from unstructured text facilitate target identification and validation. Network-building modules help elucidate novel druggable targets by amalgamating insights from omics data, pathways, literature and chemical libraries (Rzhetsky et al., 2014). Studies show Watson excels in question-answering, natural language processing and knowledge-guided analysis which are core to omics-based drug development (Kloor, 2014). The integrated functional architecture enables precision analytics critical for personalized medicine initiatives. Other platforms like Inovalon supports population health management, care coordination and resource optimization for providers (Adams, 2018). Through commercial analytical suites comprising modules tailored for healthcare and life sciences workflows, organizations can realize big data's benefits.

Similarly, Linguamatics natural language processing platforms facilitated clinical research by extracting facts from unstructured data at scale (Tang et al., 2015). As depicted in Figure 4, Linguamatics' NLP-based AI system follows a three-step process to enable massive data retention and analysis (D'Aurizio et al., 2019). The first step involves extraction from unstructured text sources such as clinical notes, reports etc. using NLP techniques. In the second step, the extracted concepts are mapped to standardized terminologies and ontologies. This helps in normalization of terms for effective querying and retrieval. Finally, the processed data is loaded into relational databases or data warehouses to support analytics and knowledge discovery applications (Tseytlin et al., 2016).

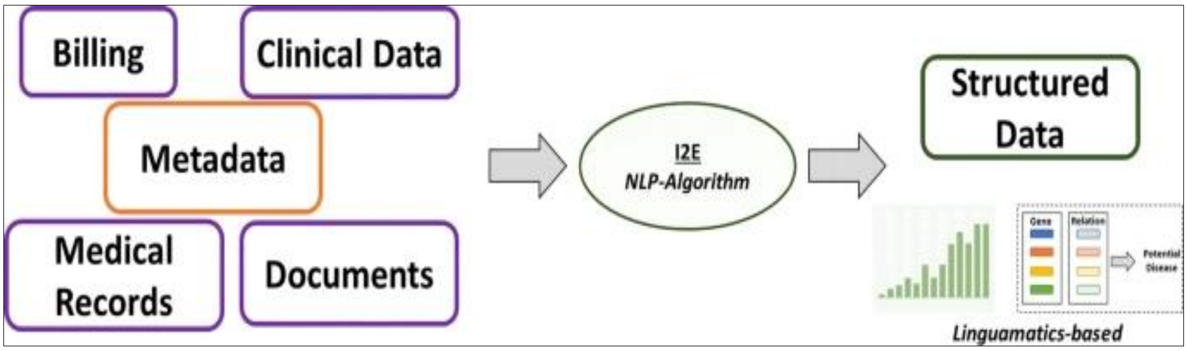

Figure 5 Working principle of Linguamatics' NLP-based AI system for massive data retention and analysis

The working principle illustrated in Fig. 4 enables Linguamatics' platforms for handling large volumes of unstructured clinical text. The scientific literature has suggested that these platforms help to obtain data for large-scale utilization in clinical research fields like pharmacovigilance and outcome measurement (Amiri et al., 2019). The ability to preserve metadata along with the structured concepts extracted from such unstructured reports remains useful for the higher level of analysis, reporting, and precise phenotyping in healthcare facilities (Tang et al., 2015).

4.9. Challenges associated with healthcare big data

While big data is effective in the improvement of healthcare there are issues it brings with it that need to be addressed systematically. One major concern is the data preprocessing and conversion of big healthcare data into a normalized form because it is complicated and consists of several forms (Kanter et al., 2015). Absence of standards results in problems in data exchange, as well as consolidating data from different sources for common examination (Bollier & Firestone, 2010). Ensuring privacy and security of sensitive patient health information is also a crucial challenge. With increasing usage of mobility and cloud for healthcare delivery, protecting privacy of electronic medical records and other digital health data against unauthorized access becomes vital (Mikalef & Krogstie, 2018). Regulatory compliances like HIPAA further complicate trans border data sharing and analysis in big data projects (Jordan, 2014).

Quality and noisiness of real-world clinical and medical data poses analytical challenges compared to experimental data. Natural language processing of unstructured notes, diagnosis codes and incomplete records require harmonization before automated modeling (Valikodath et al., 2017). Lack of contextual metadata and historical data also limits full exploitation of big healthcare datasets. Transforming raw clinical data into actionable medical knowledge remains a major roadblock. This involves complex processes of data cleaning, integration, modeling, interpretation and validation. Moreover, attributing direct cost-savings and collaborative data sharing continue to obstruct big data value realization in healthcare organizations. Addressing the gamut of technical, ethical and strategic challenges would define a big data impact on future healthcare revolution.

4.10. Quantum mechanics and big data analysis

Quantum computing promises new paradigms to harness power of big data by leveraging principles of quantum mechanics. Exponential speedup for certain classes of algorithms opens door to solving previously intractable problems involving huge data volumes (Chakraborty et al., 2018). Quantum machine learning brought hope to learn patterns from vast biomedical datasets with potential discovery of new cures (Biamonte et al., 2017).

Quantum algorithms for machine learning like HHL (Harrow-Hassidim-Lloyd) and QGAN (Quantum Generative Adversarial Networks) showed capability to cluster large sets of medical records dramatically faster than classical computers (Romero et al., 2017; Zoufal et al., 2019). This enables precision in disease diagnosis, prognosis and genomic biomarker identification. Quantum annealing and VQE (Variational Quantum Eigensolver) based techniques may solve NP-hard optimization tasks arising in large-scale omics and imaging analytics in healthcare (Perdomo-Ortiz et al., 2018; Kandala et al., 2019). Early prototypes demonstrated speedups for drug discovery, epidemic modeling and resource scheduling (Mohseni et al., 2017).

However, the broad adoption of quantum computing in healthcare still faces hurdles in scalable hardware, error corrections and hybrid integration with classical systems. Ongoing multidisciplinary research at the intersection of quantum information science, life sciences, computer science and engineering hold promise to realize futuristic quantum applications within a decade.

4.11. Applications in big data analysis

Big data is enhancing various areas of healthcare through advanced analytical applications. Clinical decision support systems help physicians in evidence-based diagnosis and treatment by integrating diagnostics, guidelines and past patient records (Wang, 2019). Such systems allow detecting overlooked insights for differential diagnosis. Pharmaceutical companies apply big data techniques in various stages of drug discovery like target identification, lead optimization, clinical trials and post marketing surveillance. Compound-protein interaction prediction, toxicity analysis and adverse event reporting got optimized using ML on large chemical and outcomes data (Mosley et al., 2019).

Public health surveillance monitors disease outbreaks, epidemics and health risks through real-time analytics of clinical, environmental and social media data streams. Spatio-temporal modeling of blended datasets helps formulate containment strategies for pathogens (Vaughan, 2019). Hospital management systems leverage operational and cost data to optimize resource planning, patient scheduling, profitability analysis and readmission risk prediction ensuring quality and affordable care (Raghupathi & Raghupathi, 2014). Such systems monitor institutional performance on cost and clinical metrics.

5. Conclusions

The integration of advanced technologies and data-driven approaches hold immense potential to transform global healthcare landscapes in the coming decades. The ubiquitous availability of high-speed ubiquitous computing infrastructure, wearable devices and wireless connectivity laid the foundations for data deluge across the continuum of healthcare delivery. While industries including finance, transportation and manufacturing leveraged insights from big data to optimize services and competitiveness, healthcare lagged partly due to sensitivity of information and complexity of clinical domains. However, the unabated growth of clinical and biomedical big datasets originating from diverse sources has become unmanageable with traditional systems. This prompted paradigm shift towards assimilating multi-dimensional datasets spanning genomics, clinical histories, social determinants, lifestyle habits, environmental exposures alongside medical images and sensor-based physiological readings. Amalgamation of such distributed heterogeneous healthcare information into unified knowledge bases enhances our understanding of disease mechanisms, progression and therapeutic responses at an unprecedented scale.

Advanced computational techniques including neural networks, deep learning, predictive modeling, natural language processing, pattern recognition and statistical machine learning have started to revolutionize how medical research is conducted and patient care is delivered. While the field is still in its infancy, proof of concept applications demonstrated promising outcomes in precision diagnosis, individualized treatment recommendations, forecasting medical outcomes, resource optimization, population risk analytics and public health policy modeling. Integration of high-performance analytics suites with clinical workflows is improving productivity of caregivers and clinical support functions. Notwithstanding such opportunities, successful adoption of big data solutions requires overcoming major challenges around data privacy, standardization, security, quality, completeness, computational scalability, knowledge extraction, regulatory compliance and validating economic benefits. Continued multidisciplinary research focusing on explainable artificial intelligence, clinical decision support, expanding curated datasets, trustworthy systems and addressing social determinants would be pivotal. It is equally necessary to ensure that technological advancements align with human factors and patient-centered models.

5.1. Future Prospects

In the future, disease modeling, multi-omics integration, improving medical imaging and signal informatics, applying quantum approaches to drug discovery and clinical decision-making problems remain promising. Global databases containing large amounts of information on various pathologies can help us quickly identify them. Building a qualified healthcare workforce with skills that bridge domain knowledge with data science will, therefore, define how prospective big data can transform the system. Therefore, through the application of pragmatic approaches to managing risks and privacy, big data integrated with improvements in associated technologies signals the emergence of affordable, equitable, and personalized health for all.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Aalst, W. V. D. (2016). Data science in action. In *Process Mining* (pp. 1-4). Springer, Berlin, Heidelberg.

[2] Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. *Journal of Big Data*. https://doi.org/10.1186/s40537-017-0110-7

[3] Agrawal, A., & Choudhary, A. (2019). Health services data: big data analytics for deriving predictive healthcare insights. *Health Services Evaluation*. https://doi.org/10.1007/978-1-4899-7673-4_2-1

[4] Archenna, J., & Mary, A. E. A. (2015). A survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, 50(2), 40.

[5] Athanasopoulou, C., Vozikis, M., Koutra, K., Löttyniemi, E., Bertsias, A., Basta, M., Vgontzas, A. N., & Lionis, C. (2017). Internet use, eHealth literacy and attitudes toward computers/internet among people with schizophrenia spectrum disorders: a cross-sectional study in two distant European regions. *BMC Medical Informatics and Decision Making*, 17(136), 1-14. https://doi.org/10.1186/s12911-017-0531-4

[6] Bainbridge, M. (2019). Big data challenges for clinical and precision medicine. In M. Househ, A. Kushniruk, & E. Borycki (Eds.), *Big data, big challenges: A healthcare perspective: Background, issues, solutions and research directions* (pp. 17–31). Springer.

[7] Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big Data In Health Care: Using Analytics To identify And Manage High-Rish And High-Cost Patients. *Health Affairs*, 33(7), 34-46. https://doi.org/10.1377/hlthaff.2014.0041

[8] Bharathi MJ, Rajavarman VN. A survey on big data management in health care using IOT. *Int J Recent Technol Eng.* 2019;**7**:196–198.

[9] Bi, Z., & Cochran, D. (2014). Big data analytics with applications. *Journal of Management Analytics*, 1(4), 249–265. https://doi.org/10.1080/23270012.2014.992985

[10] Boerma, T., Requejo, J., Victora, C. G., Amouzou, A., Asha, G., Agyepong, I., & Borghi, J. (2018). Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child health. *The Lancet*, 391(10129), 1538–1548.

[11] Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data*. Aspen Institute, Communications and Society Program.

[12] Borne, K. (2013). Collaborative Annotation for Scientific Data Discovery and Reuse. *Bulletin of the Association for Information Science and Technology*, 39(4), 44-5. https://doi.org/10.1002/bult.2013.1720390414

[13] C, B. (2012). Big Data and Analytics Key to Accountable Care Success. *Healthcare Article*, 1-4.

[14] Castro, E. M., Van Regenmortel, T., Vanhaecht, K., Sermeus, W., & Van Hecke, A. (2016). Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. *Patient Education and Counseling*, 99(12), 1923–1939.

[15] Chellah, R. C., & Kunda, D. (2020). An assessment of factors that affect the implementation of big data analytics in the Zambian health sector for strategic planning and predictive analysis: a case of Copperbelt province. *International Journal of Electronic Healthcare*, 11, 101-122.

[16] Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275, 314–347.

[17] Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Harvard Business School Publishing.

[18] Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business School Publishing Corporation.

[19] Dixon, B. E., Kasting, M. L., Wilson, S., Kulkarni, A., Zimet, G. D., & Downs, S. M. (2017). Health care providers' perceptions of use and influence of clinical decision support reminders: qualitative study following a randomized trial to improve HPV vaccination rates. *BMC Medical Informatics and Decision Making*, 17(119), 1-10. https://doi.org/10.1186/s12911-017-0521-6

[20] Erickson, S., & Rothberg, H. (2017). Data, information, and intelligence. In E. Rodriguez (Ed.), *The analytics process* (pp. 111–126). Auerbach Publications.

[21] Fang, H., Zhang, Z., Wang, C. J., Daneshmand, M., Wang, C., & Wang, H. (2015). A survey of big data research. *IEEE Network*, 29(5), 6–9.

[22] Fredriksson, C. (2016). *Organizational knowledge creation with big data: A case study of the concept and practical use of big data in a local government context*.

[23] Fromme, E. K., Eilers, K. M., Mori, M., Hsieh, Y. C., & Beer, T. M. (2004). How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *Journal of Clinical Oncology*, 22(17), 3485-3490.

[24] Gandomi, A., & Haider, M. (2013). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-44. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

[25] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.

[26] Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2015). *The 'big data' revolution in healthcare: Accelerating value and innovation*.

[27] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660.

[28] Gunapal PPG, Kannapiran P, Teow KL, Zhu Z, You AX, Saxena N, Singh V, Tham L, Choo PWJ, Chong P-N, Sim JHJ, Wong JEL. Setting up a regional health system database for seamless population health management in Singapore. *Proc Singapore Healthc.* 2016;**25**:27–34.

[29] Gupta, V., Singh, V. K., Ghose, U., & Mukhija, P. (2019). A quantitative and text-based characterization of big data research. *Journal of Intelligent & Fuzzy Systems*, 36, 4659–4675.

[30] Hampel, H. O. B. S., O'Bryant, S. E., Castrillo, J. I., Ritchie, C., Rojkova, K., Broich, K., & Escott-Price, V. (2016). PRECISION MEDICINE-the golden gate for detection, treatment and prevention of Alzheimer's disease. *Journal of Prevention of Alzheimer's Disease*, 3(4), 243.

[31] Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 15(1), 55-86. https://doi.org/10.1007/s10618-006-0059-1

[32] Harerimana, G. B., Jang, J., Kim, W., & Park, H. K. (2018). Health big data analytics: A technology survey. *IEEE Access*, 6, 65661–65678. https://doi.org/10.1109/ACCESS.2018.2878254

[33] Hussain, S., Hussain, M., Afzal, M., Hussain, J., Bang, J., Seung, H., & Lee, S. (2019). Semantic preservation of standardized healthcare documents in big data. *International Journal of Medical Informatics*, 129, 133–145. https://doi.org/10.1016/j.ijmedinf.2019.05.024

[34] Ismail, A., Shehab, A., & El-Henawy, I. M. (2019). Healthcare analysis in smart big data analytics: Reviews, challenges and recommendations. In *Security in smart cities: Models, applications, and challenges* (pp. 27–45). Springer.

[35] Jordan, S. R. (2014). Beneficence and the expert bureaucracy. *Public Integrity*, 16(4), 375–394. https://doi.org/10.2753/PIN1099-9922160404

[36] Krumholz, H. M. (2014). Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system. *Health Affairs*, 33(7), 1163–1170.

[37] Kruse, C. S., Goswamy, R., Raval, Y. J., & Marawi, S. (2016). Challenges and opportunities of big data in healthcare: A systematic review. *JMIR Medical Informatics*, 4(4), e38.

[38] Kundella, S., & Gobinath, R. (2019). A survey on big data analytics in medical and healthcare using cloud computing. *International Journal of Scientific & Technology Research*, 8, 1061-1065.

[39] Kyoungyoung, J., & Gang, H. K. (2013). Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system. *Healthcare Informatics Research*, 19(2), 79–85.

[40] Lerner, I., Veil, R., Nguyen, D. P., Luu, V. P., & Jantzen, R. (2018). Revolution in health care: How will data science impact doctor-patient relationships? *Frontiers in Public Health*, 6, 99.

[41] Lytras, M. D., & Papadopoulou, P. (Eds.). (2017). *Applying big data analytics in bioinformatics and medicine*. IGI Global.

[42] Madsen, L. B. (2014). *Data-driven healthcare: How analytics and BI are transforming the industry*. Wiley.

[43] Marconi, K., Dobra, M., & Thompson, C. (2020). The use of big data in healthcare. In J. Liebowitz (Ed.), *Big data and business analytics*. CRC Press.

[44] Mauro, A. D., Greco, M., & Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, 65(3), 122-135.

[45] Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics*, 114, 57–65.

[46] Mikalef, P., & Krogstie, J. (2018). Big data analytics as an enabler of process innovation capabilities: A configurational approach. In *International Conference on Business Process Management* (pp. 426–441). Springer.

[47] Nambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013). A look at challenges and opportunities of big data analytics in healthcare. In *2013 IEEE International Conference on Big Data* (pp. 17–22). IEEE.

[48] Nambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013). A look at challenges and opportunities of Big Data analytics in Healthcare. *IEEE International Conference on Big Data*, 1-17. https://doi.org/10.1109/BigData.2013.6691753

[49] Nasi, G., Cucciniello, M., & Guerrazzi, C. (2015). The role of mobile technologies in health care processes: the case of cancer supportive care. *Journal of Medical Internet Research*, 17(2), e26.

[50] Olszak, C., & Mach-Król, M. (2018). A conceptual framework for assessing an organization's readiness to adopt big data. *Sustainability*, 10(10), 3734.

[51] Ong, T. C., Kahn, M. G., Kwan, B. M., Yamashita, T., Brandt, E., Hosokawa, P., Uhrich, C., & Schilling, L. M. (2017). Dynamic-ETL: a hybrid approach for health data extraction, transformation and loading. *BMC Medical Informatics and Decision Making*, 17(134), 1-12. https://doi.org/10.1186/s12911-017-0532-3

[52] Or-Bach, Z. (2017). A 1,000x improvement in computer systems by bridging the processor-memory gap. In *2017 IEEE SOI-3D-subthreshold microelectronics technology unified conference (S3S)*.

[53] Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59.

[54] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.

[55] Ristevski, B., & Chen, M. (2018). Big data analytics in medicine and healthcare. *Journal of Integrative Bioinformatics*, 15(3). https://doi.org/10.1515/jib-2017-0030

[56] Sánchez-de-Madariaga, R., Muñoz, A., Lozano-Rubí, R., Serrano-Balazote, P., Castro, A. L., Moreno, O., & Pascual, M. (2017). Examining database persistence of ISO/EN 13606 standardized electronic health record extracts: relational vs. NoSQL approaches. *BMC Medical Informatics and Decision Making*, 17(123), 1-14. https://doi.org/10.1186/s12911-017-0515-4

[57] Saouabi, M., & Ezzati, A. (2017). A comparative between hadoop mapreduce and apache Spark on HDFS. In *Proceedings of the 1st international conference on internet of things and machine learning* (pp. 1-4). ACM.

[58] Seinivasan, U., & Agrawal, B. (2013). Leveraging Big Data Analytics to Reduce Healthcare Costs. *IEEE IT Professional*, 15(6), 21-7. https://doi.org/10.1109/MITP.2013.55

[59] Senthilkumar, S. A., Rai, B. K., Meshram, A. A., Gunasekaran, A., & Chandrakumarmangalam, S. (2018). Big data in healthcare management: a review of literature. *American Journal of Theoretical and Applied Business*, 4, 57-69.

[60] Shameer, K., Badgeley, M. A., Miotto, R., Glicksberg, B. S., Morgan, J. W., & Dudley, J. T. (2017). Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. *Briefings in Bioinformatics*, 18(1), 105-124.

[61] Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The hadoop distributed file system. In *Proceedings of the 2010 IEEE 26th symposium on mass storage systems and technologies (MSST)* (pp. 1-10). IEEE Computer Society.

[62] Tang, J., Tao, D., Qi, G. J., & Huet, B. (2014). Social media mining and knowledge discovery. *Multimedia Systems*, 20(1), 633-4. https://doi.org/10.1007/s00530-014-0423-8

[63] Tsai, C. W., Lai, C. F., Chao, H. C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big Data*, 2, 21. https://doi.org/10.1186/s40537-015-0030-3

[64] Valikodath, N. G., Newman-Casey, P. A., Lee, P. P., Musch, D. C., Niziol, L. M., & Woodward, M. A. (2017). Agreement of ocular symptom reporting between patient-reported outcomes and medical records. *JAMA Ophthalmology*, 135(3), 225-231.

[65] Voronin, A. A., Panchenko, V. Y., & Zheltikov, A. M. (2016). Supercomputations and big-data analysis in strong-field ultrafast optical physics: filamentation of high-peak-power ultrashort laser pulses. *Laser Physics Letters*, 13(6), 065403.

[66] Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organization. *Technological Forecasting and Social Change*, 126(1), 3-13. https://doi.org/10.1016/j.techfore.2015.12.019

[67] Wang, Y., Kung, L., Wang, W., Yu, C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: application to healthcare. *Information & Management*, 55(1), 64-79.

[68] Williams, N., Ferdinand, N. P., & Croft, R. (2014). Project management maturity in the age of big data. *International Journal of Managing Projects in Business*, 7(2), 311-317.

[69] Yang, J. J., Li, J., Mulder, J., Wang, Y., Chen, S., Wu, H., & Pan, H. (2015). Emerging information technologies for enhanced healthcare. *Computers in Industry*, 69, 3-11.

[70] Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42, 146-157.