Big Data in Healthcare: Acquisition, Management, and Visualization Using System Dynamics

Student:Ashiat Ashake Adeogun
College of Basic and Applied Sciences, Healthcare Informatics
Middle Tennessee State University
Murfreesboro, TN, United States
aaa2ev@mtmail.mtsu.edu

Supervisor:Misagh Faezipour
Department of Engineering Technology
Middle Tennessee State University
Murfreesboro, TN, United States
misagh.faezipour@mtsu.edu

***Abstract*—Integrating big data, electronic health records (EHR), and health information systems (HIS) has yielded transformative benefits in healthcare. However, it has also introduced critical risk factors. These risks encompass data privacy and security concerns, necessitating robust encryption techniques and adherence to regulations like the Health Insurance Portability and Accountability Act (HIPAA). Additionally, data inconsistencies stemming from human behavior and decision-making can compromise data quality, urging strict data quality management and governance. Data fragmentation, primarily due to disparate data storage systems and limited interoperability, challenges comprehensive data integration. The accuracy of data in extensive datasets poses another challenge, with adaptations of quality management techniques required. Re-identification risk, wherein supposedly anonymous data can be traced back to individuals, underscores privacy breaches. Lastly, the evolving landscape of healthcare data governance presents its unique challenges. Using system dynamics in big data analytics can significantly improve healthcare management, decision-making, and patient outcomes. System dynamics is a modeling technique that helps understand complex systems, such as healthcare systems, by representing them as interconnected feedback loops and variables. When applied to big data in healthcare, it can provide valuable insights and support evidence-based decision-making. Addressing these risk factors is imperative to ensure big data's success and ethical use in healthcare through system dynamics.**

***Index Terms*—Big data, Healthcare, Electronic health records (EHR), Health information systems (HIS), and System Dynamics (SD).**

I. Introduction

A. Definition of Big Data

Big data is the term used to describe the enormous amounts of data that are more than what conventional software or internet-based platforms can efficiently manage, store, process, and analyze. It includes large datasets that are beyond the capabilities of traditional techniques and equipment. Big data refers to the size and complexity of massive data collections, which call for specialist management and insight extraction techniques [17].

Big data in healthcare is characterized by volume, variety, velocity, veracity, variability, and value. These features highlight the complexity and potential of healthcare data for improved outcomes and innovation [41]. Douglas Laney's definition of big data is widely recognized and widely accepted.

1

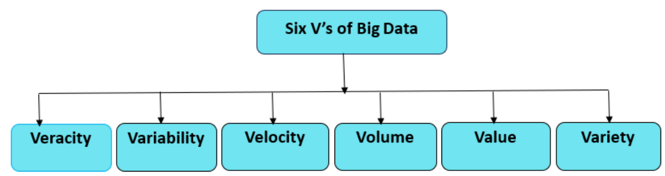

Fig. 1. Six V's of Big Data

He identified three key dimensions that characterize big data: volume, velocity, and variety, often called the three V's [33]. Velocity refers to the speed of data collection, enabling real-time analysis. Variety encompasses the diverse data types that can be collected, including structured and unstructured formats. While additional V's have been proposed, the most accepted fourth V is "veracity," emphasizing the significance of data quality and reliability [18]. Variability in healthcare data refers to its fluctuations and diversity during handling. It enhances data's value by revealing valuable, unforeseen insights and making it more attractive for analysis [47]. Value in big data refers to extracting useful information through analytics. Big data analytics enables organizations to uncover patterns and trends, informing decision-making and driving innovation across various sectors, including healthcare [41].

B. Healthcare as a big-data repository

The healthcare system has been transformed by adopting EHRs and HISs. EHRs digitize patient medical histories, enabling seamless access and sharing of information among healthcare professionals. HIS encompasses data management, analysis, and integration from diverse sources, empowering big data analytics to extract valuable insights. By integrating healthcare data with genomics, environmental factors, and social determinants of health, personalized medicine is achieved, tailoring treatments to individual needs. However, ensuring data privacy, security, and ethical use is vital. Upholding patient confidentiality and complying with regulations like the Health Insurance Portability and Accountability Act (HIPAA) are essential considerations. Balancing data utilization with privacy rights remains a challenge. The transition to EHRs and HISs has revolutionized healthcare, providing a comprehensive data repository that enhances decision-making, coordination among providers, research initiatives, and public health interventions through big data analysis. This data-driven approach can potentially improve healthcare outcomes and advance medicine [19].

Digitizing clinical exams and medical records is now standard in healthcare systems, improving efficiency, accessibility, and communication. EHR enables accurate diagnoses, personalized treatments, and patient engagement. Data security and privacy are vital for success in this digitized healthcare landscape [17].

II. Methodology - System Dynamics

System dynamics (SD) is an analytical modeling approach deeply rooted in the overarching framework of general systems theory, initially proposed by von Bertalanffy in 1968 [8]. However, the foundational roots of system dynamics can be traced back to the 1950s at MIT, where Jay Forrester conducted pioneering research in industrial dynamics [23].

The fundamental principle of system dynamics is that a system's behavior is determined by its underlying structure. In simpler terms, the way various components within a system are interconnected and exert influence on each other directly shapes the overall behavior of the entire system. This often results in emergent behavior that can be counterintuitive, necessitating a comprehensive examination of the individual components to understand the reasons behind these unexpected outcomes better [10]. SD comprises two discernible components: one qualitative and one quantitative. The qualitative facet entails the creation of causal loops or influence diagrams, which visually illustrate the interrelationships among the elements within the system. The primary objective is to enrich comprehension of a given problem or situation by examining the system's structure and the connections among pertinent variables [10].

A. Feedback Loops in SD

These feedback mechanisms encompass the various interactions that can occur within a system. These interactions involve:

1) Reinforcing Loop in SD: A reinforcing or positive feedback loop (R) intensifies changes, leading to exponential growth or decline as one variable's shift triggers processes amplifying the difference in the same direction [10]. For example, as more patient data is collected and analyzed, it can lead to better diagnoses, treatment strategies, and patient outcomes. These positive outcomes encourage healthcare providers to invest more in data acquisition, management, and visualization technologies, leading to more data-driven insights and improvements. This self-reinforcing cycle can contribute to continuous enhancements in healthcare practices and outcomes through big data. In the context of the previous example, the reinforcing loop diagram features arrows that signify the direction of the loop, with the positive sign (+) indicating growth in the forward movement. Similarly, the reinforcing loop consistently adheres to a clockwise format as shown in figure 2.

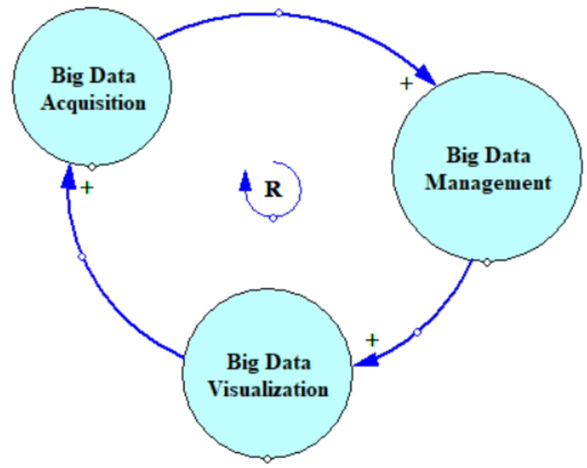

Fig. 2. The Reinforcement Loop Visualization for Big Data in SD

2) Balancing Loop in SD: A balancing loop, also known as a negative feedback loop (B), responds to changes by activating processes that oppose or neutralize those changes. It stabilizes systems, maintaining them near a setpoint or equilibrium [10]. For example, inconsistencies in big data healthcare applications present a substantial hurdle, risking inaccurate diagnoses, treatment errors, and unreliable research outcomes. Addressing these challenges involves data cleaning, quality assurance, and robust data governance. Once these foundational issues are effectively managed, healthcare institutions can focus on enhancing data acquisition, management, and visualization processes, ultimately leading to better-informed decisions and improved patient outcomes. This balancing loop illustrates how efforts to address data inconsistencies through data cleaning, quality assurance, and data governance can decrease the negative impact of inconsistencies, ultimately resulting in more reliable data for decision-making in healthcare. The loop follows an anticlockwise direction, symbolizing the desired reduction in data inconsistencies as shown in Figure 3.

III. Exploring the Interplay of Risk Factors in Big Data within the Healthcare Context

System dynamics relies on an intricate causal model incorporating diverse variables and feedback loops, including balancing loops that maintain equilibrium and reinforcing loops that amplify effects. These loops interact within healthcare systems, shaping the dynamics of patient care, resource allocation, and policy implementation. In the subsequent sections, we will explore these loops to elucidate how System Dynamics is a

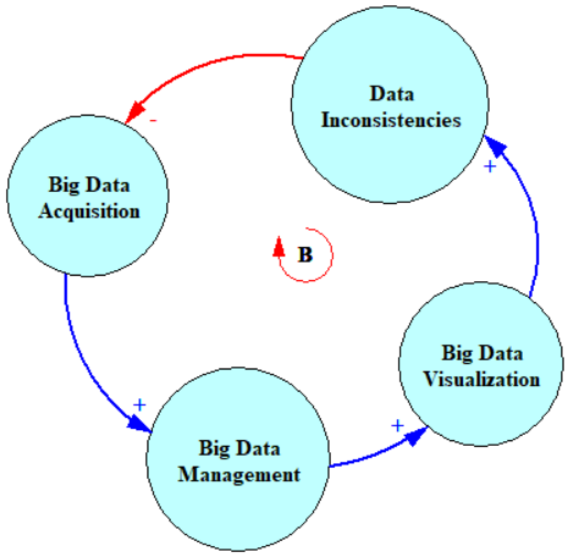

Fig. 3. The Balancing Loop Visualization for Big Data in SD

potent healthcare acquisition, management, and visualization tool, as shown in Figure 4.

1) Big Data Acquisition, Management and Visualization (R1): Data acquisition is the conversion of real-world physical phenomena into electrical signals, which are then measured and converted into a digital format for computer processing, analysis, and storage [38]. Healthcare data can originate from primary sources such as Computerized Physician Order Entry (CPOE) systems, clinical decision support systems, and electronic health records, as well as secondary sources like laboratories, insurance companies, government agencies, pharmacies, and Health Maintenance Organizations (HMO) [3] the most significant sources of big data in EHRs. Data management in healthcare encompasses activities such as organizing, cleaning, retrieving, mining, and governing data. It involves ensuring the data's accuracy and completeness by validating for any irrelevant or missing values and removing such data as necessary [5]. Data visualization translates complex healthcare data into visual or graphical formats, aiding understanding and informed decision-making. It uncovers patterns and correlations, enabling the effective communication of insights. Presenting analytic results visually enhances comprehension and supports data-driven actions in the healthcare domain [41]. As shown in Figure 4, in healthcare's big data realm, a reinforcing loop (+) means that better data acquisition and management lead to enhanced visualization and positive outcomes. These drive ongoing investment and innovation, creating an improving cycle for healthcare that benefits patients and processes.

2) Data Documentation (R2): Data documentation in acquisition, management, and visualization forms a reinforcing loop (+). Comprehensive documentation enhances data quality and trust during acquisition, leading to increased data acquisition. Effective data management, guided by documentation, maintains data quality and compliance. High-quality data improves visualization accuracy, fostering better data understanding among stakeholders. Improved performance encourages further engagement with data, resulting in more robust insights. This positive feedback loop emphasizes the importance of consistent and thorough data documentation throughout the lifecycle to enhance data-driven processes in various domains, including healthcare, research, and business, as shown in R2 in Figure 4.

3) Big Data Acquisition, Management, Visualization, and Data Privacy and Security (R3): Data collection, transmission, and processing are the three components that make up the big data gathering process. Big data collection is the process of gathering and retrieving limitless amounts of raw data from many sources, whether unstructured, semi-structured, or structured. [35]. Data movement from data sources into storage management systems for data processing and analysis is called big data transmission. This process guarantees improved and adequate data for analysis and storage. It is necessary to preprocess and enrich the acquired data by removing redundant, noisy, incomplete, and unnecessary data to reduce storage requirements and enhance analytical accuracy. Obtaining low-density data must also be combined with other data to provide value [14]. The main objective of clinical data visualization is to support clinical decision-making. This can be accomplished by offering displays that clarify the temporal relationship between data points and enable direct comparison of relevant data [45]. The efficiency and accuracy of data review can be improved, and a deeper comprehension of the organ system or disease process to which the data relates can be facilitated as secondary purposes of data visualization [46].

Healthcare organizations handle vast amounts of data crucial for efficient care delivery, but securing this data has long been a formidable challenge. The healthcare sector is particularly vulnerable to data breaches, which can expose sensitive information due to attackers employing data mining techniques. As security measures evolve, so do the tactics of sophisticated attackers, including advanced persistent threats [1]. These attacks aim to extract valuable data surreptitiously. Consequently, patient privacy in big data analytics is an escalating concern, demanding that organizations simultaneously address data security and privacy. Data security controls data access across its lifecycle, while data privacy aligns access with policies and laws governing who can access personal, financial, medical, and confidential data. A recent incident highlighted by Forbes underscores the urgency of patient privacy protection [27]. Data privacy and security are critical enablers for various aspects of big data in healthcare. They foster trust, compliance, and responsible data handling practices, ultimately enhancing data acquisition, management, and visualization. By prioritizing data privacy and security, healthcare organizations can maximize the value of big data

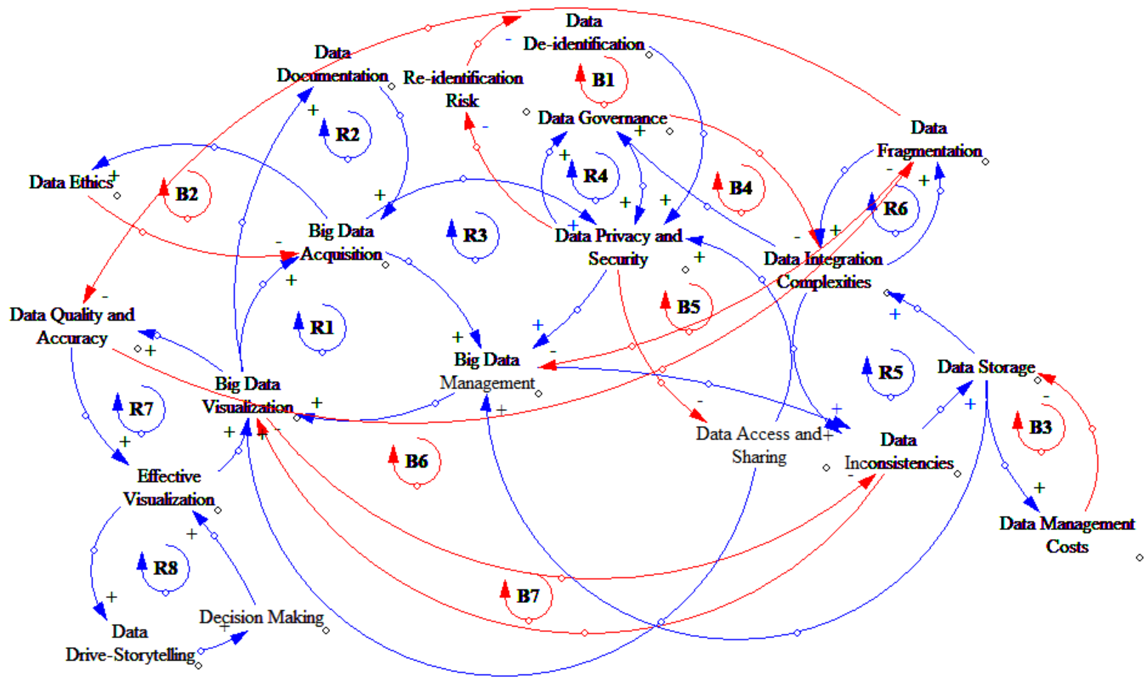

Fig. 4. The Visualization of Causal Model for Big Data in Healthcare

analytics while safeguarding patient information and complying with legal and ethical standards, as shown as a reinforcing loop in R3 in Figure 4.

4) Data Privacy and Security and Data Governance (R4): Information privacy refers to the right to control the collection and use of personal data, ensuring it remains confidential to those authorized to access it. A significant concern in user privacy is preventing the exposure of personal information while transmitting it online [40]. On the other hand, security involves safeguarding information and its assets using technology, processes, and training to prevent unauthorized access, disclosure, disruption, modification, inspection, recording, or destruction [28]. In 2018, Bruno M.C. Silva and researchers introduced novel data encryption techniques for healthcare systems, focusing on mobile health systems. The method aimed to enhance data security by combating challenges such as phishing and unauthorized access in the healthcare sector. Their paper emphasized using cryptography as the ideal solution to safeguard patient information in the rapidly evolving digital healthcare landscape [43]. In 2019, a research article by Scholars Mbonihankuye and colleagues highlighted data security in healthcare, focusing on Hospital Information Systems and HIPAA compliance [9]. They proposed six proactive defensible actions, Deny, Disrupt, Degrade, Deceive, and Contain, to safeguard healthcare data, ensuring its confidentiality, integrity, and availability. These measures bolster security, prevent unauthorized access, and create a safer environment while adhering to industry regulations, protecting patient data from potential risks and breaches [36].

Data governance is a cornerstone for effectively organizing and managing enterprise data assets. It enables the integration of clinical and business policies into data management, facilitating automated collection and reporting in alignment with these policies. In turn, data governance ensures the circulation of quality information throughout the organization [22]. As shown in R4 in Figure 4, the reinforcing loop suggests that data privacy, security, and governance are pivotal in healthcare big data. They protect sensitive patient data, maintain trust, and ensure compliance with regulations like HIPAA. These measures enable responsible and secure data utilization, enhance healthcare outcomes, and preserve data integrity.

5) Data Integration Complexities, Data Storage and Data Inconsistencies (R5): Data integration offers consolidated data access by gathering information from various origins and delivering users a cohesive data perspective [25]. Ziegler and Dittrich [50] explain the motivation for integrating multiple data sources is twofold: firstly, to streamline information access by creating a unified view of existing information systems, and secondly, to establish a broader foundation by merging data from diverse, complementary information systems. The complexity of big data is well acknowledged, comprising a wide range of data types, including structured, unstructured, and semi-structured data. Moreover, within domains such as clinical and genomic data, intricate inherent semantic relationships and intricate networks of connections link various data entities [48]. For example, data Integration in Mental Health Care, Brain, and mental health research in Big Brain Data has become a burgeoning field for data analysts and neuroscientists. Electroencephalography (EEG) is the primary method for studying brain function, utilizing advanced sensing technology and signal processing algorithms. EEG has proven invaluable in supporting healthcare needs, such as identifying ketamine responses in treatment-resistant depression through wearable forehead EEG devices [12], Investigating the complexity of resting-state EEG patterns preceding migraine episodes [12], and indexing the dynamics of cortical activity in the brain to detect signs of driving fatigue and drowsiness [11].

In big data, storage is paramount, especially in healthcare, where data keeps expanding. A robust, large-scale storage platform is crucial. Cloud technology is the most promising solution, offering scalability, efficiency, and accessibility. Cloud environments provide the flexibility to handle large datasets, improve data accessibility, expedite scalable analytics, and unlock substantial value. Cloud computing, a potent technology, enables vast data storage and facilitates complex computing, all while eliminating the need for costly hardware, software, and dedicated physical space [41]. Data generation, acquisition, aggregation, transformation, and representation inevitably introduce inconsistencies in big data. These disparities often stem from various aspects of human behavior and decision-making. In large datasets, these inconsistencies can manifest across different levels of knowledge content, from raw data and information to knowledge, meta-knowledge, and expertise. Such disparities can potentially detrimentally impact the quality of outcomes in extensive data analysis [2].

Inconsistencies may also emerge in the reasoning, heuristics, or problem-solving approaches employed for various analytical tasks, leading to challenges in comprehensive data analysis [13]. It consists of four essential types of inconsistencies in big data. They include temporal inconsistencies, spatial inconsistencies, inconsistencies found in unstructured natural language text, and inconsistencies stemming from violations of functional dependencies [49]. As seen in Figure 4 in R5, increasing data integration complexities in healthcare drive the demand for more storage capacity. Complex integration processes generate larger data volumes. This increased storage capacity, in turn, reinforces the ability to manage integration complexities. However, accumulating diverse data from various sources can lead to inconsistency, potentially affecting data quality and decision-making. Careful data governance and quality assurance are crucial to maintaining the reliability of healthcare data while managing integration challenges.

6) Data Integration Complexities, Data Fragmentation (R6): In healthcare, the challenges of managing big data are amplified by data fragmentation, dispersion across diverse systems, and the limited interoperability of legacy electronic medical record systems. Integrating healthcare data into a cohesive big data framework is complicated due to schemas, formats, metadata, and standard variations. These obstacles hinder seamless data integration and complicate the utilization of big data in the healthcare sector [37]. Data fragmentation in healthcare can lead to complexities in integrating fragmented data sources. As healthcare organizations work to address these integration challenges, they may uncover additional fragmented data sources or inconsistencies, creating a reinforcing loop. This loop highlights the interdependency between data fragmentation and integration complexities, making data acquisition and management more challenging in the healthcare sector over time, as shown in R6 in Figure 4.

7) Big Data Visualization, Data Quality, and Accuracy and Effective Visualization (R7): Data quality is appropriate for use or meets user needs in quality management [42]. There is a consensus that data quality is contingent upon source data quality. In other words, the accuracy and reliability of data depend on the quality of its origin [34]. Data visualization is essential for extracting insights from vast datasets, enabling structured presentations. It aids data scientists in uncovering hidden patterns and supports real-time problem-solving. Business analysts employ it to identify improvement areas, comprehend consumer behavior drivers, and predict sales. Visualization-based tools address significant data challenges, opening avenues for business growth [21]. The reinforcing loop of data quality and accuracy and effective visualization indicate that data quality and accuracy are fundamental in healthcare big data, as they directly influence the reliability and credibility of insights and decisions. Effective visualization enhances understanding and facilitates informed healthcare decisions. They improve patient care, diagnosis accuracy, and operational efficiency while reducing risks and costs, as shown in R7 in Figure 4.

8) Effective Visualization, Data Drive-Storytelling, Decision Making and big Data Visualization (R8): In healthcare, tremendous connections between good visualization and data-driven storytelling allow for the effective transmission of data insights, the support of evidence-based decision-making, and the promotion of a better comprehension of complicated healthcare information. This mixture is essential for maximizing the potential of big data in healthcare and enhancing patient outcomes, as seen in R8 in Figure 4.

Clinical decision-making tools offer decisions supported by data, patient-specific evaluations, and/or suggestions [30]. Clinical judgments backed by data from health information systems can help decision-makers enhance patient safety, close knowledge and practice gaps, and achieve performance gains [7]. Care reminders, patient-specific recommendations, and prescribing help are essential computerized clinical decision support technologies that can lower medical and prescribing errors and guarantee requirements are reached [30]. Better clinical decision support is provided by the use of big data analytical tools on large healthcare data sets [16]. This reinforcing loop results in more impactful big data visualization in healthcare, where effective visualization enhances storytelling, which, in turn, influences informed decision-making. Each component strengthens the others in this iterative process, facilitating better healthcare decisions based on data-driven insights, as shown in R8 in Figure 4.

9) Data De-identification and Re-identification and Data Privacy and Security (B1): The HIPAA Privacy Rule [32] acknowledges the importance of sharing patient data within healthcare organizations. "De-identified data" refers to patient information from which all identifiable details, such as names, addresses, and Social Security numbers, have been removed. This process ensures that the data cannot be used to identify the patient reasonably. Re-identifying the patient from this de-identified data is only possible through the original data owners or legally authorized entities using specific methods [39]. For instance, a bioterrorism monitoring center, for example, can study de-identified data while maintaining patient privacy under normal conditions. This method of de-identification emphasizes monitoring for natural or human-induced disease outbreaks.

Additionally, if an attack is discovered, the Centers for Disease Control and Prevention (CDC) or other legally sanctioned organizations can instantly re-identify a patient using their private key and then take the necessary steps to contain the outbreak and improve the chances of survival for the infected individuals [39]. 2018, the Ministry of Health and Family Welfare introduced the Digital Information Security in Healthcare Act (DISHA). This legislation aimed to ensure privacy in the healthcare sector. DISHA enables regulation, secure storage, authorized transmission, and controlled access to healthcare data, including data linked to personally identifiable information [20]. As seen in B1 in Figure 4, The balancing loop indicates that the data de-identification serves as a privacy-enhancing safeguard, minimizing the risk of unauthorized access and privacy violations. Data privacy and security measures reinforce de-identification effectiveness and serve as barriers against re-identification attempts. The looming threat of re-identification underscores the imperative for robust privacy and security practices and meticulous de-identification protocols within the expansive domain of big data healthcare.

10) Big Data Acquisition and Data Ethics (B2): Let's begin with the definition of "Ethics involves the pursuit of goods that we are justified, and at times morally obligated, to seek. This includes the good of knowledge, which can potentially drive substantial advancements in healthcare [44]". Big Data in healthcare encompasses many data types, including biological, clinical, environmental, and lifestyle information. It is collected from individuals and large populations to understand their health and well-being, often spanning various time points [6]. Ethical considerations in data acquisition are essential for preserving privacy, protecting individuals' rights, and maintaining ethical standards in data-driven practices. However, these ethical constraints can lead to challenges like incomplete or less diverse datasets. Striking a balance between ethical concerns and the need for comprehensive data is a complex task that requires careful consideration of legal and ethical boundaries while aiming to acquire valuable and diverse data within those limits. It underscores the importance of responsible data acquisition practices in the moral and accountable use of big data as shown in B2 in Figure 4.

11) Data Storage and Data Management Costs (B3): Storage is a crucial component of handling big data in healthcare. As the volume of data in the healthcare industry grows, the demand for efficient and expansive storage solutions becomes paramount. Among the most promising technologies, cloud storage stands out. Cloud platforms offer scalability and efficiency, enabling seamless access to data, facilitating insight generation, and empowering scalable analytics solutions. Cloud computing, a robust and up-and-coming technology, accommodates vast data storage needs and supports extensive and intricate data processing. It eliminates the necessity of maintaining expensive computing hardware, software, and dedicated physical space, streamlining data management in healthcare [41]. For several reasons, using cloud technologies for big data analysis in healthcare makes sense. Big data analysis necessitates substantial financial outlays, necessitating a reliable and cheap infrastructure, which cloud solutions provide. Healthcare organizations deal with internal and external data sources, although significant amounts of data are obtained externally from public providers and third-party sources. We manage private patient data internally. Big data research in the healthcare field can be cost-efficient and effective thanks to cloud technologies that enable users to organize and analyze this diversity of data securely [41]. Hashem et al. Emphasizing that cloud computing is a viable platform for fulfilling the critical data storage requirements for conducting comprehensive big data analysis [26]. From Figure 4 in B3, expanding data storage can lead to increased data management costs, primarily due to infrastructure and maintenance expenses. However, an organization can counteract this cost escalation by optimizing data management practices, resulting in significant cost savings by efficiently utilizing existing storage resources. Striking the right balance between the costs of data storage expansion and the implementation of efficient data management strategies is essential for controlling overall expenses while ensuring data accessibility and reliability.

A. Applications of Big Data Analytics In Healthcare

Leveraging Big Data Analytics in Healthcare (BDAH) offers a promising solution to healthcare cost challenges. It can enhance disease diagnosis and treatment strategies. Several applications of BDAH in healthcare, such as disease diagnosis and treatment optimization, are discussed, highlighting the potential for improving healthcare efficiency and outcomes through data-driven insights.

1) Diagnosis: IBM's PureData Solution for Healthcare Analytics is a fully integrated software solution designed to assist hospitals and clinics in harnessing the power of healthcare analytics [31]. This software is engineered to provide swift outcomes while mitigating risks. Seattle's Children's Hospital uses big data analytics to enhance diagnosis and patient care. Their approach involves leveraging IBM's big data technology to diagnose approximately 350,000 patients. This system identifies patterns among patients, streamlining the diagnostic and treatment processes for improved precision. By aggregating data from ten different sources within the hospital and analyzing historical hospital records, it identifies areas requiring additional resources to expedite and enhance the accuracy of the diagnosis [31].

It consolidates essential information on applying big data analytics in medical contexts, focusing on detecting heart attacks and the unpredictable aspects of medical research, particularly utilizing IoT techniques and Hadoop [37]. This application serves as an end-user platform for online, specialized projects. The primary objective of this application is to empower users with the capability to access immediate guidance on heart disease through an intelligent online system [4].

2) Treatment: The diabetic dataset was effectively described with the assistance of Hive and R. Insights gained during this process can be employed to create predictive models. While the investigation phase has succeeded, the disclosed data has the potential for developing even more efficient predictive models [29]. Heart Disease Prognosis: Many nursing homes use a health information system to handle health information. The system extracts the hidden clues for a creative clinical analysis when it comprises a large amount of data. This investigation aims to develop an innovative approach for predicting heart illness that analyzes heart disease utilizing heart-related data. These 13 elements, together with clinical problems like sexuality, blood pressure, and cholesterol, are used to develop this method. Two additional factors, namely obesity and smoking, are utilized to get more appropriate results because these factors are considered to be significant risk factors for heart disease [15].

To investigate cutting-edge cancer and disease treatments, BDAH permits the tracking of patient results and outcomes while integrating deoxyribonucleic acid DNA information. Intel and Oregon Health and Science University collaborated to create the Collaborative Cancer Cloud platform, which enables hospitals to share vital patient data for more accurate cancer treatment. Doctors can obtain the DNA profiles of other cancer patients as a comparison tool for their own patients' genetic information. To help clinicians develop more effective treatment plans, BDAH plays a crucial role in evaluating patterns and spotting traits that patients have in common. Beyond cancer, this strategy applies to genetically based diseases and conditions like Alzheimer's, autism, and others [31]. A supercomputer, IBM's Watson Health, enables healthcare firms to create their applications, goods, and data analytics. To improve the precision of diagnosis and treatments, Watson Care Manager (WCM) inside this ecosystem interfaces with Apple's health applications to collect and communicate relevant data with healthcare professionals. Pediatric patients at Boston Children's Hospital benefit from using WCM, notably in heart health, individualized medicine, and critical care. Similarly, the cancer center at Columbia University uses Watson's capacity to read DNA and create individualized cancer treatment plans [24].

IV. Conclusion and Future Reseach

Integrating big data, EHRs, and HISs holds immense promise for reshaping healthcare. This integration can enhance patient care, foster collaboration among healthcare providers, advance research, and enable effective public health interventions through data analysis. However, this promising transformation also brings substantial risks. These risks encompass concerns regarding data privacy and security, data inconsistencies arising from human behavior, data fragmentation, issues related to data quality, re-identification risks, and complex challenges in data governance. Healthcare organizations must prioritize robust data privacy and security measures to mitigate these risks effectively. This includes implementing encryption techniques and adhering strictly to regulations like HIPAA. Addressing data inconsistencies necessitates investments in data quality management and governance practices to ensure data accuracy and reliability. Overcoming data fragmentation challenges requires promoting interoperability standards and standardization within healthcare systems.

Guaranteeing data quality and accuracy within extensive datasets mandates rigorous data cleaning and validation procedures. Adopting stringent data de-identification practices is indispensable to minimize re-identification risks. The article also introduces the concept of SD as seen through causal loops and provides a structured approach to understanding, decision-making, optimization, and continuous improvement within healthcare systems. It highlights the dynamic nature of healthcare operations and helps organizations navigate complexity for better patient care and resource management.

In conclusion, despite acknowledging the associated risks, the potential transformation that big data offers in healthcare is substantial. It underscores the critical significance of data security, quality, and governance in harnessing these benefits responsibly. Future research endeavors should concentrate on advancing data security, fostering interoperability, developing privacy-preserving analytics, establishing ethical data governance models, conducting impact assessments, embracing patient-centric approaches, encouraging global collaboration, and adapting regulations to unlock the full potential of big data in healthcare.

References

[1] Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H., and Saadi, M. Big data security and privacy in healthcare: A review. *Procedia Computer Science 113* (2017), 73–80.

[2] Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., Gehrke, J., Haas, L., Halevy, A., Han, J., et al. Challenges and opportunities with big data 2011-1.

[3] Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., and Taha, K. Efficient machine learning for big data: A review. *Big Data Research 2*, 3 (2015), 87–93.

[4] Alexander, C. A., and Wang, L. Big data analytics in heart attack prediction. *J Nurs Care 6*, 393 (2017), 2167–1168.

[5] ARCHENAA, J., AND ANITA, E. M. A survey of big data analytics in healthcare and government. *Procedia Computer Science 50* (2015), 408–413.

[6] AUFFRAY, C., BALLING, R., BARROSO, I., BENCZE, L., BENSON, M., BERGERON, J., BERNAL-DELGADO, E., BLOMBERG, N., BOCK, C., CONESA, A., ET AL. Making sense of big data in health research: towards an eu action plan. *Genome medicine 8*, 1 (2016), 1–13.

[7] BATES, D., MÄCHLER, M., BOLKER, B., AND WALKER, S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).

[8] BERTALANFFY, L. V. *General system theory: Foundations, development, applications*. G. Braziller, 1968.

[9] BOMMAREDDY, S., KHAN, J. A., ANAND, R., ET AL. A review on healthcare data privacy and security. *Networking Technologies in Smart Healthcare* (2022), 165–187.

[10] BRAILSFORD, S. C. System dynamics: What's in it for healthcare simulation modelers. In *2008 winter simulation conference* (2008), IEEE, pp. 1478–1483.

[11] CAO, Z., CHUANG, C.-H., KING, J.-K., AND LIN, C.-T. Multi-channel eeg recordings during a sustained-attention driving task. *Scientific data 6*, 1 (2019), 19.

[12] CAO, Z., LIN, C.-T., DING, W., CHEN, M.-H., LI, C.-T., AND SU, T.-P. Identifying ketamine responses in treatment-resistant depression using a wearable forehead eeg. *IEEE transactions on biomedical engineering 66*, 6 (2018), 1668–1679.

[13] CHAUDHURI, S., DAYAL, U., AND NARASAYYA, V. An overview of business intelligence technology. *Communications of the ACM 54*, 8 (2011), 88–98.

[14] CHEN, M., MAO, S., ZHANG, Y., LEUNG, V. C., ET AL. *Big data: related technologies, challenges and future prospects*, vol. 100. Springer, 2014.

[15] D'AGOSTINO, R. B., GRUNDY, S., SULLIVAN, L. M., WILSON, P., GROUP, C. R. P., ET AL. Validation of the framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *Jama 286*, 2 (2001), 180–187.

[16] DANG, D. A., AND MENDON, D. S. The value of big data in clinical decision making. *International Journal of Computer Science and Information Technologies 6*, 4 (2015), 3830–3835.

[17] DASH, S., SHAKYAWAR, S. K., SHARMA, M., AND KAUSHIK, S. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data 6*, 1 (2019), 1–25.

[18] DE MAURO, A., GRECO, M., AND GRIMALDI, M. A formal definition of big data based on its essential features. *Library review 65*, 3 (2016), 122–135.

[19] DOYLE-LINDRUD, S. The evolution of the electronic health record. *Number 2/April 2015 19*, 2 (2015), 153–154.

[20] FARAHANI, B., FIROUZI, F., CHANG, V., BADAROGLU, M., CONSTANT, N., AND MANKODIYA, K. Towards fog-driven iot ehealth: Promises and challenges of iot in medicine and healthcare. *Future generation computer systems 78* (2018), 659–676.

[21] FAROUKHI, A. Z., EL ALAOUI, I., GAHI, Y., AND AMINE, A. Big data monetization throughout big data value chain: a comprehensive review. *Journal of Big Data 7*, 1 (2020), 1–22.

[22] FLEISSNER, B., JASTI, K., ALES, J., AND THOMAS, R. The importance of data governance in healthcare. *White papers, Encore, available at: http://encorehealthresources. com/wp-content/uploads/2014/10/The-Importance-of-Data-Governance_FINAL-Oct-2014. pdf* (2014).

[23] FORRESTER, J. W. *The impact of feedback control concepts on the management sciences*. Foundation for Instrumentation Education and Research, 1960.

[24] GOEDERT, J. New uses, clients for ibm watson health. *Health Data Management* (2015).

[25] HAAS, L. Beauty and the beast: The theory and practice of information integration. In *Database Theory–ICDT 2007: 11th International Conference, Barcelona, Spain, January 10-12, 2007. Proceedings 11* (2006), Springer, pp. 28–43.

[26] HASHEM, I. A. T., YAQOOB, I., ANUAR, N. B., MOKHTAR, S., GANI, A., AND KHAN, S. U. The rise of "big data" on cloud computing: Review and open research issues. *Information systems 47* (2015), 98–115.

[27] HILL, K. How target figured out a teen girl was pregnant before her father did. *Forbes, Inc 7* (2012), 4–1.

[28] JAIN, P., GYANCHANDANI, M., AND KHARE, N. Big data privacy: a technological perspective and review. *Journal of Big Data 3* (2016), 1–25.

[29] JEFFREY, A. M., XIA, X., AND CRAIG, I. K. When to initiate hiv therapy: A control theoretic approach. *IEEE transactions on Biomedical Engineering 50*, 11 (2003), 1213–1220.

[30] KAWAMOTO, K., HOULIHAN, C. A., BALAS, E. A., AND LOBACH, D. F. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj 330*, 7494 (2005), 765.

[31] KRISHNAN, S. M. Application of analytics to big data in healthcare. In *2016 32nd Southern Biomedical Engineering Conference (SBEC)* (2016), IEEE, pp. 156–157.

[32] LANDI, W., AND RAO, R. B. Secure de-identification and re-identification. In *AMIA Annual Symposium Proceedings* (2003), vol. 2003, American Medical Informatics Association, p. 905.

[33] LANEY, D., ET AL. 3d data management: Controlling data volume, velocity and variety. *META group research note 6*, 70 (2001), 1.

[34] MAIER, M., SEREBRENIK, A., AND VANDERFEESTEN, I. Towards a big data reference architecture. *University of Eindhoven* (2013), 1–144.

[35] MANOGARAN, G., THOTA, C., LOPEZ, D., VIJAYAKUMAR, V., ABBAS, K. M., AND SUNDARSEKAR, R. Big data knowledge system in healthcare. *Internet of things and big data technologies for next generation healthcare* (2017), 133–157.

[36] MBONIHANKUYE, S., NKUNZIMANA, A., AND NDAGIJIMANA, A. Healthcare data security technology: Hipaa compliance. *Wireless communications and mobile computing 2019* (2019), 1–7.

[37] OLARONKE, I., AND OLUWASEUN, O. Big data in healthcare: Prospects, challenges and resolutions. In *2016 Future technologies conference (FTC)* (2016), IEEE, pp. 1152–1157.

[38] PARK, J., AND MACKAY, S. *Practical data acquisition for instrumentation and control systems*. Newnes, 2003.

[39] PARTS, C. 160 and 164; standards for privacy of individually identifiable health information: Final rule. *Federal Register 67* (2000).

[40] PORAMBAGE, P., YLIANTTILA, M., SCHMITT, C., KUMAR, P., GURTOV, A., AND VASILAKOS, A. V. The quest for privacy in the internet of things. *IEEE Cloud Computing 3*, 2 (2016), 36–45.

[41] SENTHILKUMAR, S., RAI, B. K., MESHRAM, A. A., GUNASEKARAN, A., AND CHANDRAKUMARMANGALAM, S. Big data in healthcare management: a review of literature. *American Journal of Theoretical and Applied Business 4*, 2 (2018), 57–69.

[42] SIDI, F., PANAHY, P. H. S., AFFENDEY, L. S., JABAR, M. A., IBRAHIM, H., AND MUSTAPHA, A. Data quality: A survey of data quality dimensions. In *2012 International Conference on Information Retrieval & Knowledge Management* (2012), IEEE, pp. 300–304.

[43] SILVA, B. M., RODRIGUES, J. J., CANELO, F., LOPES, I. M., AND LLORET, J. Towards a cooperative security system for mobile-health applications. *Electronic Commerce Research 19* (2019), 629–654.

[44] TASIOULAS, J. Big data, human rights and the ethics of scientific research. *accesible en https://www. abc. net. au/religion/big-data-human-rights-and-the-ethics-of scientific-research/10096298 (último acceso 17/10/19)* (2016).

[45] WALLACE, C. J., BIGELOW, S., XU, X., AND ELSTEIN, L. Collaborative practice: usability of text-based, electronic patient care guidelines. *CIN: Computers, Informatics, Nursing 25*, 1 (2007), 39–44.

[46] WANDERER, J. P., NELSON, S. E., EHRENFELD, J. M., MONAHAN, S., AND PARK, S. Clinical data visualization: the current state and future needs. *Journal of medical systems 40* (2016), 1–9.

[47] WANG, L., AND ALEXANDER, C. A. Big data in medical applications and health care. *American Medical Journal 6*, 1 (2015), 1.

[48] WU, X., ZHU, X., WU, G.-Q., AND DING, W. Data mining with big data. *IEEE transactions on knowledge and data engineering 26*, 1 (2013), 97–107.

[49] ZHANG, D. Inconsistencies in big data. In *2013 IEEE 12th International Conference on Cognitive Informatics and Cognitive Computing* (2013), IEEE, pp. 61–67.

[50] ZIEGLER, P., AND DITTRICH, K. R. Data integration—problems, approaches, and perspectives. In *Conceptual modelling in information systems engineering*. Springer, 2007, pp. 39–58.