

The Promise and Perils of Big Data in Healthcare: Evaluating Opportunities to Improve Quality and Access while Reducing Costs

Mei Ling

Department of Computer Science, Tsinghua University
mei.ling@tsinghua.edu.cn

Saad Ullah

China & School of Science
saad.ullah4992@gmail.com

Abstract

Big data analytics holds significant promise for transforming healthcare quality, access, and costs. Advanced analysis of vast data streams and datasets enables enhanced clinical decision-making, earlier public health interventions, accelerated research, and improved patient outcomes. However, there are also important challenges involving privacy, discrimination, efficacy, costs, and unintended consequences that must be addressed responsibly even while innovation advances. This paper examines the opportunities for leveraging big data across healthcare including population health initiatives, clinical support tools, medical discoveries, and personalized medicine. It also explores limitations and risks such as re-identification vulnerabilities, biased algorithms, infrastructure expenses, and nonlinear system effects. An integrated oversight framework is recommended that combines governance mechanisms across stakeholders to enable context-appropriate progress. Providers, patients, innovators, payers, and regulators each have roles in stewarding data use towards safe and ethical applications focused on health gains. Hybrid oversight strategies can foster breakthroughs by guiding evidence-based decision-making while monitoring impacts among priority groups. Ultimately realizing the promise of big data in healthcare requires policy innovation at pace with technical innovation. Collaborative development of flexible policies can accelerate life-saving discoveries through data while building public trust in data-driven interventions aimed at democratizing quality care.

Indexing terms: Big Data, Healthcare, Analytics, Quality, Access, Costs

Introduction

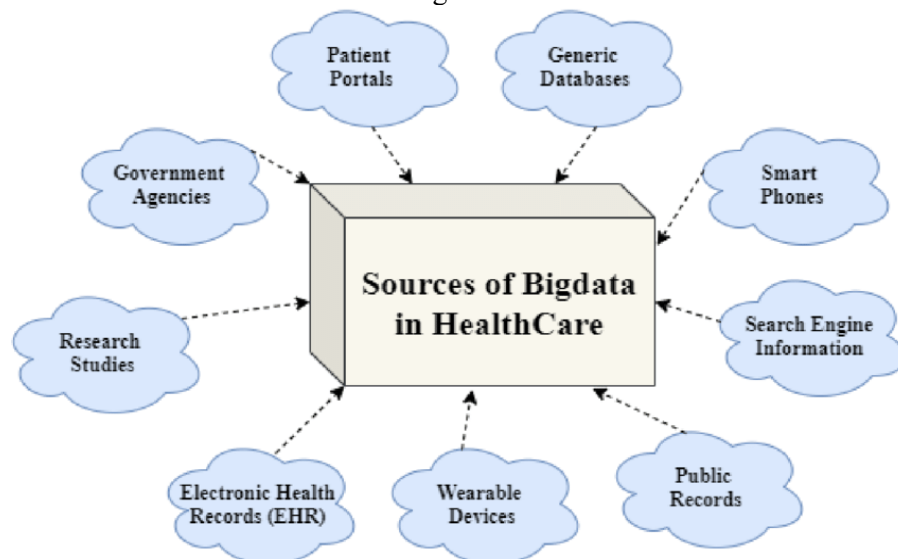
The healthcare industry is experiencing an exponential growth in data. Electronic health records (EHRs), medical imaging, genomics, mobile health applications, wearable devices, and remote monitoring tools are generating unprecedented volumes of health and medical information. It is estimated that the quantity of healthcare data doubles every two years and will reach 2,314 exabytes globally by 2025 (Raghupathi & Raghupathi, 2014; Statista, 2022). This proliferation of multidimensional health data holds tremendous promise for transforming healthcare delivery in terms of improving quality, expanding access, and reducing costs [1]. The surge in healthcare data is emblematic of the transformative impact of technological and administrative advancements on the healthcare landscape. The widespread integration of Electronic Health Records (EHRs) and computerized physician order entry systems within hospital settings has ushered in an era of automated documentation for clinical patient encounters. This automation facilitates the meticulous recording of detailed information encompassing symptoms, medical histories, assessments, interventions, and outcomes, creating a rich repository of data that can be harnessed for comprehensive analysis (Murdoch & Detsky, 2013). The efficiency achieved through these systems not only enhances the speed and accuracy of data capture but also lays the groundwork for a more nuanced understanding of patient health trajectories [2], [3].

The proliferation of medical devices further contributes to the exponential growth of healthcare data. High-resolution imaging scanners, genomics sequencers, and internet-connected wearables are at the vanguard of this data deluge, generating petabyte-scale datasets. These datasets, with their granular insights into diagnostic and physiologic indicators of health and disease, present unprecedented opportunities for personalized and data-driven healthcare approaches. The integration of genomics data, in particular, has opened new frontiers in precision medicine, allowing for tailored interventions based on an individual's unique genetic makeup. Concurrently, the evolution of health policies has played a pivotal role in shaping the landscape of healthcare data [4]. The emphasis on standardized quality measurement across healthcare facilities, coupled

with value-based payment reforms, has propelled the expansion of health reporting for entire populations. This has led to the creation of uniformly structured claims data, providing a standardized framework for assessing healthcare outcomes and resource utilization. The result is a wealth of information that not only facilitates performance evaluations of healthcare providers but also serves as a valuable resource for policymakers in their efforts to optimize healthcare delivery [5].

As the volume and complexity of healthcare data continues to escalate, so do the challenges and opportunities associated with its management and analysis. The need for robust data governance frameworks, interoperability standards, and advanced analytics tools is becoming increasingly paramount [6]. Moreover, the potential benefits of harnessing this vast trove of data extend beyond individual patient care to encompass broader public health initiatives, research endeavors, and the development of predictive models that can inform proactive healthcare interventions. In navigating this data-driven era, stakeholders must remain cognizant of ethical considerations, privacy concerns, and the imperative to strike a balance between innovation and safeguarding the integrity of sensitive health information.

Figure 1.



This proliferation of multimodal health data covering patients, encounters, transactions, outcomes, genomics, devices, and more has created a vast potential foundation for extracting value using analytics—the discovery and communication of meaningful health system patterns to inform decisions and predict key variables of interest (Groves et al., 2013). Analytics encompasses an array of quantitative and qualitative techniques including machine learning, statistics, behavioral science, clinical epidemiology, and health economics that can transform messy, disconnected datasets into critical insights to enhance efficiency, personalization, and automation at scale. The maturing capabilities and capacity for complex analytics combined with exponential health data growth lays the groundwork for a learning health system—leveraging real world evidence to systematically improve outcomes, innovation, and policies (Bates et al., 2014). Big data analytics applied effectively across the healthcare ecosystem—from public health surveillance to precision diagnostics and treatments—can provide the insights needed to accelerate improvements for health outcomes, equity and affordability [7].

Significant funding initiatives also reflect the promising potential impact of enhanced data sharing and analytics on medicine and health delivery. In 2015, the Obama administration launched the Precision Medicine Initiative—a \$215 million project supporting advances in personalized prevention and treatment informed by individual variability in genes, environments and lifestyles (Collins & Varmus, 2015). Core goals include generating longitudinal healthcare databases with genomic mappings and developing policies plus standards enabling responsible data access. More recently, the 21st Century Cures Act established over \$6 billion to accelerate therapeutic development and evidence generation leveraging real world data and analytics (Black et al., 2017). The scale of investment in advanced data infrastructure further underscores its transformative possibilities [6].

Realizing this promise to progress towards democratized access to precise, proactive and evidence-based care fueled by robust data sharing and analytics requires effectively navigating risks. Key policy, organizational and technical obstacles must be tackled to truly activate a learning health system powered by big data. Core challenges that demand focused mitigation efforts involve privacy concerns, algorithm biases, cost burdens, unproven analytics efficacy, system complexity plus implementation variability [8]. The following sections explore specific opportunities where big data analytics can drive healthcare improvements while also examining limitations that must be thoughtfully addressed across contexts to responsibly guide progress. Identifying balanced oversight and governance approaches represents an essential priority for enabling big data innovations across healthcare settings while assuring patient and community protections.

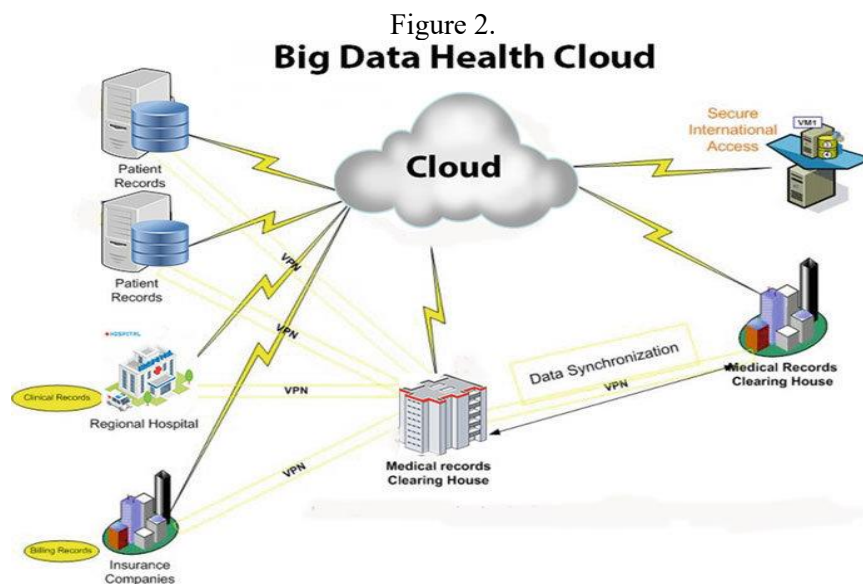
Navigating the Landscape of Healthcare Transformation through Data-driven Strategies

In the contemporary healthcare landscape, navigating the path of transformation is increasingly reliant on the judicious utilization of data-driven strategies. This journey involves a meticulous exploration of the profound impact that big data can have on reshaping healthcare systems. By leveraging advanced data analytics, organizations can strategically enhance the quality of healthcare delivery, improve accessibility to services, and concurrently achieve cost-effectiveness. The process entails discerning the intricate interplay between diverse datasets and implementing strategic initiatives that not only capitalize on the potential benefits but also address the inherent challenges. As healthcare continues its evolution, this paradigm of navigating through data-driven strategies stands as a crucial determinant in fostering a more efficient, responsive, and resilient healthcare ecosystem.

Opportunities for Quality and Access: In the realm of healthcare, the synergy between vast data streams and sophisticated analytics has ushered in a new era of possibilities. The sheer volume and diversity of healthcare data now available present a unique opportunity to discern intricate patterns, ultimately leading to more accurate predictions, personalized care strategies, disease prevention, and overall improvements in quality [9]. This convergence of data and analytics forms the foundation for precision medicine, a paradigm shift wherein interventions are precisely tailored to the unique characteristics of each individual. As the healthcare landscape becomes increasingly inundated with diverse data sets, the potential utility of this information expands correspondingly. Precision medicine hinges on the ability of analytics to seamlessly integrate and make sense of disparate sources of data, including patient records, diagnostic information, treatment protocols, and outcomes data. This integration is pivotal in optimizing decision-making processes across the entire spectrum of healthcare provision. From the broader scope of public health initiatives to the minutiae of clinical encounters, analytics can inform actions that not only enhance health outcomes and care quality but also contribute to substantial cost savings within the healthcare system.

The application of analytics in healthcare extends beyond the confines of traditional practices, permeating every facet of the industry [10]. Public health initiatives benefit from predictive analytics that identify trends and potential outbreaks, enabling timely interventions and resource allocation. At the clinical level, healthcare providers can leverage analytics to make informed decisions about treatment plans, utilizing historical patient data and real-time information to tailor interventions based on individual characteristics. This not only enhances the efficacy of treatments but also minimizes the risk of adverse reactions, aligning with the broader goals of precision medicine. Moreover, the value of analytics in healthcare is underscored by its capacity to address the longstanding challenge of disconnected information silos. By bridging gaps between patient data, diagnostic records, treatment histories, and outcomes information, analytics facilitates a holistic understanding of a patient's healthcare journey. This comprehensive view empowers healthcare professionals to make well-informed decisions, promoting a patient-centric approach that is crucial in the era of precision medicine [11]. The impact of analytics is not confined to the clinical setting alone; it extends to the realm of healthcare management and policy. Decision-makers can utilize analytics to optimize resource allocation, streamline operational processes, and formulate evidence-based policies. The result is a healthcare system that is not only more efficient but also capable of adapting to evolving challenges and demands.

Population Health: Furthermore, the integration of big data analytics in healthcare not only enhances population health management but also fosters personalized medicine by tailoring interventions based on individual health profiles. The granular insights derived from vast datasets enable healthcare providers to identify specific factors contributing to disease prevalence and progression. This precision allows for the development of targeted treatment plans and interventions, optimizing resource allocation and improving overall healthcare outcomes. Additionally, the use of predictive analytics enables proactive management of chronic conditions, predicting exacerbations before they occur and intervening promptly. The success of data-driven initiatives, such as the one implemented by the New York City Department of Health and Mental Hygiene, underscores the transformative potential of big data in healthcare. As healthcare systems continue to evolve, leveraging advanced analytics becomes imperative for optimizing patient care, reducing costs, and achieving more effective public health outcomes [12].



Clinical Decision Support: The utilization of big data in healthcare extends beyond clinical decision support systems to personalized medicine. Through the analysis of vast datasets encompassing genetic information, patient histories, and treatment outcomes, healthcare professionals can tailor interventions based on an individual's unique characteristics. This approach facilitates more precise diagnoses and treatment plans, optimizing therapeutic outcomes [13]. For instance, genomic data integrated with clinical records enables the identification of specific genetic markers associated with drug responses, allowing for the selection of medications that are more likely to be effective for a particular patient. This personalized approach not only enhances treatment efficacy but also minimizes the risk of adverse reactions, demonstrating the potential of big data to revolutionize patient care on an individualized level.

Research: In addition to enabling the discovery of correlations, vast datasets play a crucial role in advancing medical research through the derivation of new knowledge. The utilization of big data extends beyond mere correlation analysis, allowing for the identification of favorable responses to off-label prescriptions through the analysis of treatment data from a multitude of patients. This, in turn, unveils potential therapeutic indications that can be subject to further clinical testing. Furthermore, high-throughput experiments leverage massive biologic datasets to systematically manipulate gene variants among human cells, facilitating the determination of connections between specific gene variations and diseases [14]. Noteworthy initiatives such as the Human Genome Project and the Precision Medicine Initiative rely heavily on robust datasets containing genomic mappings, thereby contributing to the development of targeted therapies. The acceleration of research through big data methodologies not only enhances our understanding of various diseases but also holds the potential to unveil innovative interventions for addressing the most challenging medical conditions [7].

Access and Affordability: The implementation of big data techniques in healthcare facilitates streamlined operations and resource allocation. Through data-driven insights,

healthcare organizations can optimize their workflows, ensuring that staff, equipment, and facilities are utilized efficiently. Predictive modeling not only aids in preventing diseases but also assists in forecasting patient admissions, enabling hospitals to adjust staffing levels and allocate resources effectively. Additionally, the utilization of population health data allows for the identification of trends and patterns, enabling proactive measures to address emerging health issues [15]. The integration of clinical decision support systems further enhances efficiency by providing healthcare professionals with timely and accurate information, reducing diagnostic errors and unnecessary procedures. In essence, the efficiency gains derived from big data analytics in healthcare contribute to the overall improvement of healthcare delivery and resource utilization, fostering a more sustainable and effective healthcare system.

Table 1 summarizes the opportunities across the healthcare ecosystem made possible through effective application of analytics to big datasets. Collectively this amounts to an opportunity to reform the quality, reach, and cost structure of healthcare across interventional domains.

Table 1. Big data opportunities across the healthcare ecosystem

Ecosystem Domain	Opportunities
Public Health	Earlier disease outbreak recognition; targeted prevention initiatives
Health Systems	Care coordination & management; controlled costs for populations
Clinical Practice	Evidence-based decision support; reduced errors
Medical Research	Accelerated discoveries; precision medicine
Patients	Personalized care; improved satisfaction

Challenges for Implementation: The integration of big data analytics holds significant promise but is not devoid of challenges. Ethical considerations form a critical aspect of this landscape, as the responsible use of patient data is imperative to maintain privacy and confidentiality. Striking a balance between leveraging data for advancements in medical research and ensuring patient confidentiality requires robust ethical frameworks [16]. Additionally, organizational challenges arise in implementing these analytics solutions within the complex structures of healthcare institutions. Issues related to data governance, interoperability, and standardization need meticulous attention to facilitate seamless integration and collaboration across various healthcare entities. Furthermore, on the technical front, ensuring the security and accuracy of massive datasets is paramount. Cybersecurity measures must be stringent to safeguard against unauthorized access and potential breaches, which could compromise sensitive patient information. Addressing these ethical, organizational, and technical limitations is pivotal not only for unlocking the transformative potential of big data analytics in healthcare but also for instilling confidence, fostering trust, and safeguarding the well-being of individuals within the healthcare ecosystem.

Privacy and Security: The escalating volume of data in healthcare systems introduces complexities in ensuring data security. The interconnected nature of networked systems amplifies the potential access points for malicious attacks targeting sensitive health information. As technological advancements continue, the risk of unauthorized access and data breaches becomes more pronounced [17]. Moreover, patient apprehensions regarding the utilization of their medical and genetic data pose a challenge to the seamless functioning of healthcare systems. The reluctance to participate or share information can impede medical research and compromise the effectiveness of personalized healthcare approaches. To address these issues, healthcare institutions must implement robust security measures, transparent data usage policies, and comprehensive educational initiatives to inform patients about data protection protocols. Establishing trust through clear communication and stringent safeguards is crucial to foster patient confidence in sharing critical health-related information.

Inappropriate Analytics: Inappropriate analytics can further undermine the reliability of clinical insights. In cases where data governance is ineffective, the integration of incompatible data streams becomes a significant concern, compromising the overall quality of the analytics. The potential for advanced algorithms to detect correlations that lack true causal linkages introduces another layer of complexity. Such spurious predictor variables can contribute to overdiagnosis, placing patients at risk of unnecessary interventions and escalating medical costs [18]. To mitigate these risks, it is imperative to establish comprehensive guidelines that rigorously validate the

analytical validity and clinical utility of the data and algorithms employed in healthcare analytics. This ensures that clinicians can confidently rely on the generated insights to make informed and accurate decisions in patient care, ultimately improving the overall effectiveness and reliability of healthcare analytics systems.

Increased Costs and Inefficiencies: Although big data promises cost reductions from better efficiency, there are also expenses associated with data infrastructure and analytics. Substantial data stewardship is involved in consolidating, managing, and securing systems. Advanced analytics requires specialized data scientists in addition to considerable computing capacity. Upgrading health systems represents fixed infrastructure costs that may raise spending initially before longer-term savings materialize. If these data-management capabilities are not accounting for appropriately healthcare costs may increase.

Unintended Consequences: The complexity of interacting social, political, clinical, and economic forces within the healthcare system gives rise to nonlinear responses that elude prediction. As such, interventions enabled by big data analytics may set in motion unforeseen cascading effects. For instance, reducing hospital readmission rates to avoid Medicare penalties led some hospitals to only admit the healthiest patients, exacerbating disparities for minorities and poor populations. Avoiding unintended consequences requires careful monitoring along with nimble adjustment of data-guided policies considering contextual constraints and community needs [19].

Oversight for Implementation: The risks and complexities involved in leveraging big data require governance to ensure solutions are developed and deployed responsibly. Policies, incentives, structures, and processes are needed that balance innovation ambitions with ethical obligations. Accountability is necessary while also allowing flexibility for progress and learning given diverse health contexts. Guidelines must evolve iteratively by assessing implementation efforts. Strategies are required that are calibrated between strict regulation and unchecked experimentation enabling data-driven transformation while preventing patient harm.

As shown in Table 2, a blended oversight approach can help address the various big data obstacles across stakeholder groups. This establishes checks and balances for accountability while preserving space for progress. The multifaceted strategy combines transparency, collaboration, and responsible innovation.

Table 2

Stakeholder	Oversight Mechanisms
Providers	- Standards for procurement requiring plans for responsible data use, storage, sharing. -Incentives for evidence-based application of analytics. -Monitoring of population health impacts by priority groups
Patients	- Control over health data sharing enhanced through permissions and education. -Recourse processes for grievances related to privacy or profiling incidents
Innovators	- Voluntary ethics review boards and reporting processes prioritizing patient benefit. -Partnerships with patients and advocates to guide tool development
Payers	- Shared savings models or outcome-based contracts rewarding efficacy. -Rulemaking banning denial of coverage based solely on algorithmic insights
Regulators	- Flexible guidelines shaped collaboratively by industry alongside patient groups. -Iterative approach building oversight in response to implementation experience

By distributing responsibility across stakeholders from patient populations to innovators to providers and payers, big data solutions can be steered towards safe, ethical application. The combined inducements created through incentives, protections, and repercussions aim to accelerate discoveries while preventing harm. As evidence emerges regarding effects for individuals and communities, governance strategies can be adapted. This balanced approach is the most pragmatic way forward to foster big data innovations offering breakthrough improvements for health and healthcare.

Conclusion

In the contemporary healthcare landscape, the voluminous influx of data has emerged as a double-edged sword, offering unprecedented prospects for transforming various

facets of the industry while concurrently posing significant challenges. The potential advantages are multifaceted, encompassing enhanced patient care, improved accessibility, accelerated research endeavors, and streamlined cost management. Through the judicious use of data analytics and emerging technologies, healthcare providers can tailor treatments to individual patients, optimize resource allocation, and identify trends that contribute to more effective and efficient care delivery. However, the promise of these advancements is inextricably linked to the establishment of a robust oversight framework capable of navigating the intricate terrain of precision care [20]. A meticulous approach to data governance is imperative to ensure that the benefits derived from data analytics do not come at the expense of patient privacy, autonomy, and well-being [21]. As healthcare organizations increasingly rely on electronic health records (EHRs) and interconnected systems, the risk of privacy breaches looms large. Rigorous encryption protocols, stringent access controls, and ongoing cybersecurity measures are essential components of a comprehensive strategy to mitigate these risks and safeguard sensitive patient information.

Furthermore, the ethical considerations surrounding data utilization in healthcare necessitate vigilance against discrimination and bias. As algorithms and artificial intelligence become integral to clinical decision-making, there is a pressing need to address and rectify inherent biases in the data that may disproportionately impact certain demographic groups. Mitigating these biases requires a concerted effort to scrutinize algorithms, enhance transparency, and implement corrective measures to ensure equitable healthcare outcomes for diverse populations.

In the pursuit of innovation, there is a parallel imperative to guard against suboptimal efficacy and unintended consequences arising from the implementation of data-driven solutions. The dynamic nature of healthcare necessitates continuous monitoring and adaptation of algorithms to evolving medical knowledge and practices. Rigorous testing, validation, and ongoing evaluation of these systems are crucial to identify and rectify potential pitfalls, ensuring that the benefits of precision care are not compromised by unforeseen drawbacks [22]. Moreover, the economic implications of data-driven healthcare innovations demand careful consideration. While data analytics can contribute to cost management by optimizing resource allocation and reducing inefficiencies, the initial investment in technology, infrastructure, and training can be substantial. Striking a delicate balance between upfront costs and long-term benefits is essential to ensure the sustainability of data-driven healthcare initiatives.

Establishing comprehensive guidelines, creating incentivizing structures, and fostering collaborative partnerships are imperative to govern the responsible application of analytics in healthcare. This approach ensures that evidence-based decision-making processes become pervasive among clinicians and patients, fostering a paradigm shift towards personalized prevention and treatment strategies. The democratization of health through the tailoring of interventions is a promise that big data holds, promising efficient resource utilization [23]. For the transformative potential of data-driven healthcare to be fully realized, trust and transparency must remain at the forefront of considerations for all stakeholders involved in the healthcare ecosystem [24]. The synergy of data-enabled platforms can then coalesce to support individuals and communities, driving 21st-century innovations that prioritize the greater good. This requires a commitment to ethical practices, ongoing vigilance in addressing emerging challenges, and a dedication to ensuring that the benefits of data-driven healthcare are equitably distributed across diverse populations. In this way, the healthcare sector can harness the power of big data to propel itself into an era of unprecedented progress, where technological advancements align with human well-being.

References

- [1] J. Conn, "Future space: Harnessing the power of big data," *Mod. Healthc.*, vol. suppl, pp. 52, 54, 56–7, Aug. 2016.
- [2] P. M. Perry, "Harnessing the power of big data and data analysis to improve healthcare entities," *Healthc. Financ. Manage.*, vol. 70, no. 1, pp. 74–75, Jan. 2016.
- [3] M. Muniswamaiah, T. Agerwala, and C. Tappert, "Big data in cloud computing review and opportunities," *arXiv preprint arXiv:1912.10821*, 2019.

- [4] P. S. Bradley, "Implications of big data analytics on population health management," *Big Data*, vol. 1, no. 3, pp. 152–159, Sep. 2013.
- [5] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Approximate query processing for big data in heterogeneous databases," in *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 5765–5767.
- [6] E. Sultanow and A. M. Chircu, "Improving healthcare with data-driven track-and-trace systems," in *Strategic Data-Based Wisdom in the Big Data Era*, IGI Global, 2015, pp. 65–82.
- [7] Q. Huang, L. Wang, and Y. Yang, "Secure and Privacy-Preserving Data Sharing and Collaboration in Mobile Healthcare Social Networks of Smart Cities," *Security and Communication Networks*, vol. 2017, Aug. 2017.
- [8] N. Mehta, A. Pandit, and M. Kulkarni, "Elements of healthcare big data analytics," *Big data analytics in healthcare*, 2020.
- [9] M. Craven and C. D. Page, "Big data in healthcare: Opportunities and challenges," *Big Data*, vol. 3, no. 4, pp. 209–210, Dec. 2015.
- [10] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare," *Biomed Res. Int.*, vol. 2015, p. 370194, Jul. 2015.
- [11] C. Wickramage, T. Sahama, and G. N. Wickramanayake, "Information accountability of healthcare big data," in *2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, Colombo, Sri Lanka, 2015.
- [12] S. Nepal, R. Ranjan, and K.-K. R. Choo, "Trustworthy processing of healthcare big data in hybrid clouds," *IEEE Cloud Comput.*, vol. 2, no. 2, pp. 78–84, Mar. 2015.
- [13] A. Wehmeier and T. Baumann, "10. Big data – more risks than benefits for healthcare?," in *Big Data in Medical Science and Healthcare Management*, Berlin, München, Boston: DE GRUYTER, 2015.
- [14] A. W. Toga and I. D. Dinov, "Sharing big biomedical data," *J. Big Data*, vol. 2, no. 1, Jun. 2015.
- [15] E. Kolker and E. Kolker, "Healthcare analytics: Creating a prioritized improvement system with performance benchmarking," *Big Data*, vol. 2, no. 1, pp. 50–54, Mar. 2014.
- [16] A. E. Youssef, "A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments," *Int. J. Ambient Syst. Appl.*, vol. 2, no. 2, pp. 1–11, Jun. 2014.
- [17] G. G. T. Dantanarayana, T. Sahama, and G. N. Wikramanayake, "Quality of information for quality of life: Healthcare big data analytics," in *2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, Colombo, Sri Lanka, 2015.
- [18] S. Zillner, H. Oberkamp, C. Bretschneider, A. Zaveri, W. Faix, and S. Neururer, "Towards a technology roadmap for big data applications in the healthcare domain," in *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*, Redwood City, CA, USA, 2014.
- [19] R. Nambiar, R. Bhardwaj, A. Sethi, and R. Vargheese, "A look at challenges and opportunities of Big Data analytics in healthcare," in *2013 IEEE International Conference on Big Data*, Silicon Valley, CA, USA, 2013.
- [20] L. Fernandes, M. O'Connor, and V. Weaver, "Big data, bigger outcomes: Healthcare is embracing the big data movement, hoping to revolutionize HIM by distilling vast collection of data for specific analysis," *J. AHIMA*, vol. 83, no. 10, pp. 38–43; quiz 44, Oct. 2012.
- [21] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Federated query processing for big data in data science," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 6145–6147.
- [22] D. Levinson, "One on one. Levinson talks Stark law, big data. Interview of Daniel Levinson by Joe Carlson," *Mod. Healthc.*, vol. 42, no. 19, p. 16, May 2012.
- [23] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, "Context-aware query performance optimization for big data analytics in healthcare," in *2019 IEEE High Performance Extreme Computing Conference (HPEC-2019)*, 2019, pp. 1–7.
- [24] W. H. Inmon, M. Andrew Gettinger, and K. Krishnan, "Building the healthcare information factory: Healthcare information factory: Implementing textual

analytics,” in *Data Warehousing in the Age of Big Data*, Elsevier, 2013, pp. 289–332.