



Optimization of Big Data Analytics in Healthcare by Overcoming Security, Interoperability, and Cost Challenges

Muhammad Sohaib Anwar¹, Mazhar Abbas², Rabya Tauheed³, Asifa Maqbool⁴, Muhammad Ashraf⁵ & Muhammad Azhar⁶

¹Management Sciences Department, University of South Asia (USA), Lahore, 54000, Pakistan,

Email: sohaib.anwar@usa.edu.pk

²Cholistan Institute of Business Administration, CUVAS, Bahawalpur 63100, Pakistan,

Email: mazharabbas@cuvas.edu.pk

³Fashion Design Department, University of South Asia (USA), Lahore, 54000, Pakistan,

Email: rabya.tauheed@usa.edu.pk

⁴Muhammad Nawaz Sharif University of Engineering and Technology, Multan, Email: asifamaqbool@mnsuet.edu.pk

⁵Department of Management Sciences, COMSATS University Islamabad, Vehari 61100, Pakistan,

Email: muhammadashraf@cuivehari.edu.pk

⁶Faculty of Animal Production & Technology, Department of Poultry Sciences, CUVAS, Bahawalpur 63100, Pakistan,

Email: mazhar@cuvas.edu.pk

ARTICLE INFO

Article History:

Received: July 08, 2025
Revised: August 11, 2025
Accepted: August 28, 2025
Available Online: September 12, 2025

Keywords:

Big Data Analytics (BDA), Electronic Health Records (EHRs), Machine Learning (ML) and Artificial Intelligence (AI)

Corresponding Author:

Muhammad Sohaib Anwar

Email:

sohaib.anwar@usa.edu.pk

ABSTRACT

Big Data Analytics (BDA) is revolutionizing the delivery of health care services by enabling organizations extract values from huge volumes, varieties, and velocities of information from sources such as EHRs, Wearable Devices, and Medical Imaging. It optimizes decisions concerning clinical practices, creates individualized treatment plans, and optimizes patient's care. Advanced and newer technologies like Machine learning, and artificial intelligence have greatly boosted the functionality of predictive analytics especially in disease identification and prevention and more personal touch with patients. However, there are major challenges in incorporating BDA into the healthcare organizations. More. The issues facing implementation include data security and protection of their privacy, integration of the different systems, data quality, and issues to do with biases in algorithms, as well as the high costs of setting up the systems. These issues are even more significant at the present time because various healthcare organizations tended to implement BDA solutions to achieve the intended objectives within the healthcare domain such as enhanced patients' overall health and organizational performance. This article examines these modern issues, and enunciates some of the following major problem statements of BDA in the healthcare context. To this end, this paper presents approaches and practical solution models for enhancing BDA solutions concerning efficiency, integration, and security challenges. To our knowledge, this work will serve as an initial attempt at systematically summarizing the challenges and best practices for BDA integration in healthcare organizations to inform the development of future research and applications.



Introduction

Big Data Analytics (BDA) is essentially revolutionizing the healthcare structures as the key to process great quantity and complexity of the implemented information for the bettering of patients' outcomes and organisational and clinical experiences. BDA being in cognizance of the constant accumulation of data in EHRs, wearables and imaging arrays offers prognostic informatics, disease control and tailored therapies. For instance, BDA was employed in estimating admitting rates of emergency departments, resource requirements during disasters, and operating adaptability in the hospital material management [1][2][3].

The major advantage of BDA is the ability to analyze both the structured and unstructured information. These technologies also help in examination of clinical notes, images and other vital parameters from patients, for proper diagnosis and treatment [2][4]. These advancements are also supported by work on adopting interoperability standards to enable integration of different healthcare systems, in order to improve communication and improve decision-making [3].

However, BDA has potentials transformed in healthcare; some challenges exist with its implementation. Current topics include such challenges as data protection as users' sensitive healthcare information is getting exposed to hacking and subsequent data breaches more than ever, especially due to ransomware attacks targeting healthcare organizations. The issue of aggregating diverse data continues to be another major issue owing to the fact that most systems have no uniform methods of data formats or communication protocols. Furthermore, these biases are reflected in datasets that lead to more biases in the delivery of care through personalized as well as emerging technologies in analytics [4][5][6]. Infrastructure investments that include cables, training of personnel, and maintenance expenses are among implementation costs, which can be prohibitive for many health care facilities [6][7].

Healthcare management must connect and implement BDA for the purpose of maintaining and providing personalized, preemptive care for the patient. This means that there is more that can be done towards opening up the full potential of BDA in the healthcare system through working on issues surrounding data privacy as well as the interconnectivity and infrastructure of data systems [7].

Problems in Healthcare

Several factors exist that create challenges to the utilization of Big Data Analytics with challenges experienced by healthcare systems included the following. One emerging issue that should be of great consideration is privacy and security of data. In today's healthcare industry with the exponential data uniquely identifying each patient being electronically stored and communicated, healthcare systems' become vulnerable to cyber threats that may compromise a patient's privacy and also financial losses. Those data breaches harm patient trust and organization steady and thus data security remains relevant [8].

Another major problem is the one related to interoperability of the solutions. Healthcare facilities use different systems and interfaces leading to incompatible data structures and hence vertical information transfers are not easy. Despite the advent of standards like HL7 and more recently FHIR, most providers are still struggling to adopt and integrate these frameworks for BDA applications that would enhance efficiency [9].

Another factor that define BDA success in healthcare is data accuracy and completeness. Updated, integrated, and reliable data must be used in order to obtain accurate and reliable predictions and, therefore, make the correct decisions. The massive amount of unstructured data such as clinical

notes, and medical images render the preprocess, and quality assurance of data a challenging process [6].

Algorithms are also beginning to present issues of bias. If the models are built, with bias or not with a large focus on a specific specialty, the model may erroneously predict and participate in healthcare bias against the less represented in the datasets [7].

Lastly, the cost of implementing BDA remains a huge factor that contributes to its limit of use mostly in resource-poor health organizations. The capital cost of purchasing needed infrastructures, training people, and bearing the costs of BDA systems may be prohibitive for many healthcare organizations, thus limiting the adoption of these technologies [10].

Problem Statement

Big Data Analytics has three major concerns in the healthcare industry, which has to be met to optimize its use.

- Ensuring that individual's personal information, coming to health care providers is not leaking to third party individuals or malicious parties.
- Establishment of reference architectures that will support the sharing of data between the healthcare systems.
- Objectivity for the sake of quality and coherence in the input data so that good quality filter and prediction models are achieved.
- Reducing the risk of bias in achieving a predictive model that can help people from different backgrounds to receive equal treatment from health care providers.
- Identify ways on how general and sustainable solutions can be decreed in relation to the deployment and support of BDA systems especially in the healthcare niche which are usually characterized by limited resources.
- This article focuses on resolving the following problems:
 - Protecting sensitive healthcare data from breaches and unauthorized access.
 - Developing standardized data-sharing frameworks to enable seamless communication across different healthcare systems.
 - Ensuring the quality and consistency of data to produce accurate and reliable predictive models.
 - Minimizing bias in predictive models to ensure fair and equitable healthcare outcomes for all populations.
 - Finding cost-effective solutions for implementing and maintaining BDA systems, especially in resource-constrained healthcare environments.

Literature Review

However, latest studies have revealed that benefits and risks associated with Big DATA Analytics in healthcare are worthy of exploration. Zhang et al. (2022) investigated an application of artificial neural networks for the analysis of chronic diseases and the significance of identifying the potential threats by utilizing the models to give predictions [6]. Gupta et al. (2023) chose AI in personalized health care, stressing the importance of data conceal to conform to the regulation of privacy [7].

Typically, the ethical issues in applying Big Data in health care are described by O'Connor et al. (2022), specifically concerning the possibility of solidifying bias to training defective algorithms [8]. Using HL7 and FHIR, Jones and Patel (2023) looked at interoperability issues relating to healthcare and the role of standards in enhancing data exchange [9].

Nguyen and Chen (2023) examined related factors revealing that there are barriers to BDA implementation including; poor infrastructure and limited personnel [10]. Davis et al. (2023) described the growing danger of cyber threats targeting healthcare organisations and postulated that blockchain might improve data protection [11]. Kim et al. (2023) also showed how BDA can be used for the efficient allocation of the hospital resources, while Lee et al. (2023) used BDA for the analysis of genomic data in pursuing individualised medical treatment [12][13]. Williams et al. (2023) suggested cost-effective BDA strategies in low-resource settings while Chen et al. (2023) enhanced health care service delivery through use of patients' sentiments [14][15].

These works confirm the possibilities of using BDA to enhance the healthcare system, although it is necessary to realize that certain difficulties have to be overcome to achieve this goal.

Design and Consideration

To tackle the challenges identified in the healthcare sector, we can mathematically model the integration of Big Data Analytics in several steps:

Data Aggregation

Representing various data points from healthcare systems as a dataset:

$$D = \{d1, d2, \dots, dn\}$$

where $d1, d2, \dots, dn$ are individual data points.

Data Preprocessing

Cleaning the dataset to remove inconsistencies:

$$D_{clean} = Preprocess(D)$$

Feature Engineering

Extracting meaningful features for model building:

$$F = \{f1, f2, \dots, fm\}$$

where $f1, f2, \dots, fm$ represent the extracted features.

Model Building

Using logistic regression for predictive modeling:

$$P(y|X) = 1 / (1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m)})$$

where $\beta_0, \beta_1, \dots, \beta_m$ are model coefficients, and $X_1, X_2, \dots, X_m$ represent the features.

Model Evaluation

Calculating accuracy using the confusion matrix:

$$A = (TP + TN) / (TP + FP + TN + FN)$$

where $TP = \text{True Positives}$, $TN = \text{True Negatives}$, $FP = \text{False Positives}$, and $FN = \text{False Negatives}$.

Optimization

Improving the model using cross-validation and optimization techniques to maximize predictive accuracy.

Recommendation

To optimize the implementation of BDA in healthcare, the following parameters are used in the equations:

Parameter	Description	Example Values
β_0	Intercept term in logistic regression	0.45
$\beta_1, \beta_2, \dots, \beta_m$	Coefficients of model features	0.2, -0.1, 0.3, etc.
$X_1, X_2, \dots, X_m$	Features (e.g., age, gender, diagnosis)	65, Female, 2 (diabetes)
$P(y X)$	Probability of the outcome (e.g., disease prediction)	Calculated value based on model
Accuracy (A)	Model's performance metric	0.92
TP, TN, FP, FN	True Positives, True Negatives, False Positives, False Negatives	200, 500, 50, 20

Currently, the encryption and blockchain technology should become priorities for development of the healthcare systems in terms of data protection. Several approaches could be implemented in order to increase interoperability; among those, the use of an HL7 or FHIR framework can be used when it comes to sharing data. To minimize such risk, the models must be trained on diverse datasets that encompass a population of interests. Also, Big Data Analytics application in the medical field should make professionals complete their training to capture the optimal impacts of these technologies. Last but not least, increased joint funding between government and private sector can bring down the cost of BDA solutions, especially when adopted and implemented in the resource poor health systems.

Conclusion and Future Recommendations

Big Data Analytics has the capacity to revolutionize the healthcare domain through increasing efficiency, effectiveness and quality of the final service rendered. But it has drawbacks, like data privacy issues, the problem of compatibility with other information systems, and high costs of application implementation to make its advantages clear. It is assumed that strong confidentiality measures, definition of the general structure of exchange formats, and its moderate cost will become critical factors of the technology's successful adaption. More future research is needed for minimizing bias of the algorithms, for investigating secure data sharing approaches and for designing feasible revenue models for BDA implementation in healthcare.

References

- [1]. Zhang, H., Li, Y., & Wang, S. (2022). Application of Machine Learning in Predicting Chronic Diseases: Enhancing Early Diagnosis and Prevention. *Journal of Healthcare Informatics*, 14(3), 45-58.
- [2]. Gupta, R., Singh, A., & Verma, M. (2023). AI-Driven Personalized Healthcare: Transforming Patient Care with Data Privacy Measures. *Journal of Artificial Intelligence in Medicine*, 19(2), 134-142.

- [3]. O'Connor, D., & McGinnis, J. (2022). Ethical Concerns in Healthcare Big Data Analytics: Minimizing Bias in Machine Learning Models. *Healthcare Ethics Review*, 8(1), 25-40.
- [4]. Jones, M., & Patel, K. (2023). Interoperability Challenges in Healthcare: Overcoming Barriers with HL7 and FHIR Standards. *Health Information Systems Journal*, 11(4), 97-112.
- [5]. Nguyen, T., & Chen, F. (2023). Barriers to Big Data Analytics Adoption in Healthcare: Infrastructure and Training Challenges. *Healthcare Technology Innovations*, 5(3), 62-75.
- [6]. Davis, M., Patel, D., & Xu, Y. (2023). Cybersecurity in Healthcare: Using Blockchain to Protect Sensitive Data. *Journal of Cybersecurity in Healthcare*, 16(2), 111-120.
- [7]. Kim, J., Lee, H., & Jung, S. (2023). Predictive Analytics in Healthcare: Optimizing Hospital Resource Allocation Using Big Data. *Journal of Hospital Management*, 20(1), 79-95.
- [8]. Lee, S., & Yoon, K. (2023). Big Data in Genomic Medicine: Personalizing Treatments Using Analytics. *Journal of Genomic Medicine*, 15(3), 204-219.
- [9]. Williams, B., Smith, R., & Garcia, M. (2023). Cost-Effective Strategies for Big Data Analytics Implementation in Low-Resource Healthcare Settings. *Healthcare Economics Journal*, 12(4), 34-46.
- [10]. Chen, L., & Zhang, Y. (2023). Sentiment Analysis in Healthcare: Leveraging Patient Feedback for Service Improvement. *Journal of Healthcare Quality Improvement*, 9(2), 88-99.
- [11]. Zhang, T., & Lu, X. (2023). Predictive Modeling for Early Disease Detection: Harnessing Big Data to Improve Patient Outcomes. *Health Informatics Journal*, 21(5), 50-63.
- [12]. Brown, M., & Green, J. (2023). Data Security Measures for Healthcare Systems: A Review of Techniques and Frameworks. *International Journal of Health Information Security*, 7(4), 220-234.
- [13]. Zhang, S., & Lee, P. (2023). Evaluating Interoperability Frameworks in Healthcare Data Sharing: HL7 and FHIR Implementation Cases. *Journal of Medical Data Exchange*, 14(1), 45-60.
- [14]. Patel, R., & Sharma, D. (2023). Artificial Intelligence and Machine Learning in Personalized Healthcare: Challenges and Opportunities. *Journal of AI in Healthcare*, 18(2), 120-130.
- [15]. Wang, Y., & Liu, H. (2023). Big Data Applications in Healthcare: A Comprehensive Review and Future Directions. *International Journal of Healthcare Technology*, 22(3), 150-165.