



International Journal of Multidisciplinary Research and Growth Evaluation



International Journal of Multidisciplinary Research and Growth Evaluation

ISSN: 2582-7138

Received: 13-12-2020; Accepted: 16-01-2021

www.allmultidisciplinaryjournal.com

Volume 2; Issue 1; January-February 2021; Page No. 725-728

Building AI-Ready Data Platforms for Healthcare: Architecture, Challenges, and Best Practices

Syeda Kawsar

Independent Researcher, USA

Corresponding Author: Syeda Kawsar

DOI: <https://doi.org/10.54660/IJMRGE.2021.2.1.725-728>

Abstract

Due to the impressively continuous development of artificial intelligence (AI) across different fields, companies of various sectors have embraced the technology as it promises unprecedented benefits; one of the sectors is healthcare. AI is an exciting frontier in healthcare as it has the potential to revolutionize not only diagnostics or health record keeping but also create personalized medicine that will be capable of predicting future health trends. These innovations have the potential to transform patient care as well as organizational performances, but for their success, they depend on the

foundational data architectures that are able to efficiently manage the overwhelming challenges of healthcare data. In this paper, it discuss structure, issues, and good practices concerning data platforms healthcare organizations need to prepare for AI integration. More focus is placed on the approaches to adapting to integration challenges, data protection, and compliance with existing law. Using these suggestions, organizations need to develop frames to enhance AI's use in augmenting patient care quality.

Keywords: HealthCare, architecture, challenges, AI, data quality, security threat, costs, network security, data platforms

Introduction

AI increases innovative and patient care options and decreases the cost of operation while providing new methods and perspectives in medical data analysis. Right from diagnostics, treatment plans, and administration to smart systems, that can predict failures, sys AI is certain to transform the industry as stated by. But building better AI-based applications in the healthcare sector depends on the strength of the fundamental data platforms. Healthcare data is diverse, extensive, and comprehensive and is collected from different experimental systems. This information includes electronic health records, medical imaging, wearable, and genomic information. Some of these datasets must go through a complex process of preparation and organization to allow their utilization in AI analysis. However, interoperability and security issues, as well as the vast scale of the problem and the need to conform to legal requirements, add to the complexity of designing and deploying AI-ready data platforms. HIPAA and GDPR are just two examples of rules that strengthen the need to design correct and safe systems.

Architecture of AI-Ready Data for Health Care

1. Data Integration

There will always be the need to interlink different datasets, which is at the heart of an AI-ready data platform. Healthcare data originates from multiple sources:

1. Electronic Health Records (EHRs): Using these structured datasets, the important history of the patient can be retrieved, including diagnoses, medications and treatment plans.
2. Medical Imaging: It includes unstructured data like X-rays, MRIs, and CT scans, which necessitate vast storage and competently processed systems
3. Genomic Data: Large structured data sets that contain or present an explicit and vast map of a person's DNA, such as next-generation sequencing.
4. A robust integration layer supports:
5. Interoperability Standards: Standards that are used include HL7 FHIR (Fast Healthcare Interoperability Resources) that enable interoperability between these systems.

APIs: Application Programming Interfaces refer to means by which different systems and applications can share information.

2. Data Storage and Management

Storages are crucial for the volume and diversification of healthcare data populations. Key considerations include:

- **Hybrid Storage Models:** Any patient-specific information which requires a high level of security can be stored on the premises, while the routinely processed data can use cloud-based resources for their scalability
- **Data Lakehouse Architecture:** The use of both data lakes and data warehouses is considered the final possibility for both structured and unstructured data and supported real-time analyses.
- **Compression and Indexing:** This minimizes storage expenses through the use of high compression techniques while, on the other hand, indexing enables the data to be retrieved easily.

3. Scalability and Performance

Considering healthcare AI models, it is important to have scalable architectures that would be able to address an increasing amount of data or compute time. Strategies include:

- **Distributed Computing:** Infrastructures similar to Apache Hadoop and Apache Spark are in a position to way out the colossal number of dataset files explained a researcher ^[1].
- **Containerization:** There are application solutions that make it easier to implement solutions for scale; the most exciting of these tools are Docker and Kubernetes
- **Edge Computing:** Smart processing is useful for decreasing latency in the transmission of data as well as providing optimization for use in wearable devices and IoT.

4. Security and Privacy

In this field, security and privacy standards are rather strict because the transmitted information belongs to the healthcare sector. Best practices include:

- **Encryption:** This makes data confidentiality achievable through the enforcement of both transit and at-rest encryption.
- **Access Controls:** Management of access by roles restricts the visibility of data.
- **Audit Trails:** Supervisory and record-keeping activities provide more accountability for data access activities

5. Compliance

- **HIPAA (US):** Coordination of Measures to safeguard the identification of Protected Health Information (PHI).
- **GDPR (EU):** Regulating privacy information and data management
- **Local Regulations:** Compliance with regional legal prompts.

Challenges in Building AI-Ready Data Platforms

1. Data Quality and Standardization:

Problems with data include data disparity, data gaps, and dissimilar structures that impair AI's performance described ^[2]. Common issues include:

- **Missing Values:** Lack of completeness of EHR datasets is one of the reasons why analytics are affected.
- **Incompatible Formats:** Creating barriers for integration is the standardized variations of imaging like DICOM.

2. Interoperability

The systems may be very specialized and isolated and may not allow for intelligence sharing within the different sections of the system. Legacy systems compound these bottlenecks because they cannot integrate with many of these advanced technologies. Consequently, there are challenges in compiling the data for synthesis in order to offer information to healthcare providers. These are made of adopting middleware that lies between the old and new systems as part of the solution. SMART on FHIR and other similar open-source tools and initiatives present economically viable solutions to improve interstate compatibility. Also, rather than research sites having to physically transfer data to coordinating centres and back, cloud-based interoperability can facilitate real-time cross-institution shot data sharing. To make this a success, staff should be trained to use them and should also be motivated to keep using them.

3. Scalability Constraints

The increasing amount of data generated from health and sensor technology puts pressure on current structures. This challenge is particularly emerging as healthcare organizations want to bring in high-volume real-time data alongside their historical data. Such challenge can best be mitigated by the adoption of affordable cloud solutions, intelligent statistical analysis and a platform-based design paradigm. Also, edge computing reduces pressure on central servers by performing preliminary data processing and selecting the most optimal targets for loads distributed across systems. Such incremental scaling makes accurate cost estimations easy claimed a researcher as it considers the projections of the market demands making it sustainable and eliminating expenditures that are not needed.

4. Security Threats

Ransomware and phishing are some of the existing approaches that are vulnerable to healthcare data platforms. Consequently, due to the belonging of the healthcare sector to the shadiest and most vulnerable category, IT assets and the data contained within them are increasingly exposed to attack due to their high monetary value. Organizations that experience breaches examine staggering financial penalties, damaged reputation, as well as negative patient outcomes. To avoid such incidences, companies are advised to use different techniques such as authentication at different factors, threat detection in real-time and vulnerability tests frequently. Cybersecurity simulations can improve preparedness as well. With the use of AI based security solutions which predict and prevent threats, safety is well safeguarded. One way of protecting organizations against such risks is by creating awareness to universally reduce such likelihoods.

5. Cost

The creation of agile AI-suitable data assets entails the performance of significant investment in infrastructure professional employees, and compliance. Such investment is important but can be resource-demanding and thus challenging for organizations with constrained resources. So, using low-cost models that support interoperability and scalability standards decreases the cost of development and maintenance ^[3]. Firms are able to make their investments incremental, with basic building blocks as a foundation to

allow them to achieve their first goals while the fundamental infrastructure is laid down for later evolutionary stages.

Best Practices for Building AI-Ready Data Platforms

Inspired by the human brain, deep learning algorithms create neural networks capable of making intelligent decisions in complex situations with vast medical data and parameters^[4]. They are used for advanced medical image analysis, recognizing patterns in imaging data, and aiding in complex diagnoses. Computer vision in health care also enables computers to process, analyze, and make decisions based on visual data such as medical images, real-time video feeds, and patient photos. Natural Language Processing (NLP) in healthcare is a branch of artificial intelligence that focuses on enabling computers to understand, interpret, and manipulate human language as it is spoken or written. In healthcare, NLP systems are used to extract meaningful information from various types of unstructured text data, such as clinical notes, medical records, research reports, and even patient interactions^[5].

1. Prioritize Data Governance

- **Data Stewardship:** This makes for a more organized protocol to follow and gives an expectation on who is to be held accountable for overseeing data. They are responsible guardians of data and information and are required to ensure that data is correct, retrieval is correct, and use is correct.
- **Metadata Management:** Clear data lineage and good metadata practices give the necessary level of transparency for data origin and changes. This makes the data reliable and helpful in addressing compliance audits when needed.

2. Adopt Advanced Analytics and AI Tools

- **Machine Learning Frameworks:** Frameworks such as TensorFlow and PyTorch facilitate both the creation and training of Artificial intelligence-based ML models that fit particular niches. Such frameworks are capable of being scaled and tailored to accomplish challenging tasks in the field of data analysis.
- **Visualization Tools:** Tools like Power BI and Tableau build dashboards so that the stakeholders can engage with the data and find trends easily. Real-time analytics that are possible through streaming data processing tools enable real-time decision-making by passing the insights as the events unfold. Deploying AI models is automated, further cutting down the time taken to get insights into business operations.

3. Foster Interoperability

According to research, data sharing between different systems and organizations is highly dependent upon the data interoperation. Application of interoperability models such as HL7 FHIR helps simplify and manage data connections between diverse systems in an intricate system. It is important to note that the modernisation of these legacy systems is an important way to improve integration. As agencies implement traditional, interconnectible technologies, the sharing of information becomes more efficient.

4. Implement Layered Security

- **Network Security:** Anti-viruses, firewalls, intrusion detection systems, and other means stop unauthorized

users.

- **Application Security:** They exclude code review, penetration, and secure coding practices to minimize vulnerability in the software applications.
- **Technological developments** such as blockchain increase data reliability and authentication and largely decrease the possibility of fraud.

5. Continuous Monitoring and Auditing

Constant surveillance is necessary to maintain the reliability of platforms and data systems. Audits check that the organization is in adherence to the set laws and policies and, most of all, acknowledge possible weaknesses before they transform into major problems. Automatic monitoring programs, including the anomaly detection program, help to identify such patterns and address possible risks. Liu *et al.* utilized the FHIR Bulk Data API to create population-level exports from clinical systems, into a file format often referred to as “Flat-FHIR” represented in NDJSON-based data format. The FHIR Bulk Data API is designed to create a uniform capability for population-level exports from clinical systems, into a file format often referred to as “Flat-FHIR.” Leveraging the SMART backend services authentication and authorization profile, the approach enables healthcare providers and organizations to define and access cohorts from electronic health records and payor claims data with “push button” simplicity--a substantial advance over the current state, where each site of care needs highly skilled extraction transform and load (ETL) efforts^[6].

Conclusion

Not too long ago, AI appeared as a potentially revolutionary tool in healthcare with improvements in diagnosis, individualized treatments, prognosis and increased workflow productivity. But this only holds true if strong, elastic and compliant data platforms can be developed and implemented. Such platforms have to resist the specificities of the healthcare data they deal with, its heightened security requirements, data heterogeneity, and dispersal. There are fairly generic challenges, for example, the problems of compatibility, security, and cost, that need to be addressed to ensure that these platforms will be able to support the most advanced AI solutions. Data sharing and coordination is particularly important for data originating from such applications as EHRs, medical imaging, wearables and genomics datasets. The data formats need to be normalized and it is possible by integrating HL7 FHIR kind of frameworks available in the current context. At the same time, strong measures of protection from patient data have to be put in place to guarantee the latter's non-disclosure. Eventually, these approaches help establish and maintain data quality be in conformance with the prevailing laws, and make decisions on time.

References

1. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health*. 2020 May;17(9):3176. doi: 10.3390/ijerph17093176.
2. Misra NN, Dixit Y, Al-Mallahi A, Bhullar MS, Upadhyay R, Martynenko A. IoT, Big Data and Artificial Intelligence in Agriculture and Food Industry.

- IEEE Internet of Things Journal. 2020;9(9):1–1.
3. Agbehadji IE, Awuzie BO, Ngowi AB, Millham RC. Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing. *International Journal of Environmental Research and Public Health*. 2020 Jul;17(15):5330. doi: 10.3390/ijerph17155330.
 4. Kacheru G. The role of AI-powered telemedicine software in healthcare during the COVID-19 pandemic. *Turk J Comput Math Educ*. 2020;11(3):3054-60. doi:10.61841/turcomat.v11i3.14964.
 5. Bizzo BC, *et al*. Artificial intelligence and clinical decision support for radiologists and referring providers. *Journal of the American College of Radiology*. 2019.
 6. Kelly CJ, *et al*. Genetic improvement in cotton fiber elongation can impact yarn quality: Key challenges for delivering clinical impact with artificial intelligence. *Journal of Medical Artificial Intelligence*. 2019 Mar.
 7. Liu D, Sahu R, Ignatov V, Gottlieb D, Mandl K. High-performance computing on flat FHIR files created with the new SMART/HL7 bulk data access standard. *AMIA Annual Symposium Proceedings*. 2019;2019:592–6.