Fully Open Access   **Research Paper**

Data Engineering Challenges in AI-Driven Healthcare IT Systems: Navigating Real-Time Analytics and Interoperability

Sunil Kumar Mudusu
Church Mutual Insurance Company, - USA

Abstract

This paper studies the AI based healthcare IT system with focus on the inter-operability of the data, real time analytics and the security challenges. The literature review for qualitative data includes an increase in the adoption of the AI (78% in 2023) but the interoperability barriers still persist (65% – incompatibility due to different data standards). Findings point that data frameworks should be standardized; security should be enhanced and AI should be used in the healthcare IT optimization and optimization of patient outcomes.

Keywords: Data Engineering, IT, Analytics, Healthcare.

1   INTRODUCTION

In healthcare IT, AI and big data analysis has transformed in running the costs and improving the decision making and operational efficiency. But there are three large obstacles: interoperability, real time analytics and cybersecurity. In this study, literature is reviewed on AI adoption as well as limitations in data engineering including security threats. The narrower analysis of quantitative trends addresses the question by identifying barriers and suggesting ways of improvement. Contribution of these challenges towards the development of the data driven healthcare systems and secured, efficient patient care.

2   LITERATURE REVIEW

The Introduction of artificial intelligence (AI) in the healthcare IT systems has changed the rules of data analytics, making it possible to obtain real time insights and treating patients personally. Nevertheless, data engineering in AI based healthcare IT has provided outstanding challenges, in dealing with high dimensional biomedical data, attain interoperability and privacy in handling of sensitive patient information. The current literature review covers all the available research in these identified challenges and discusses the Big Data analytics, deep learning, interoperability frameworks, cybersecurity measures and the emerging technologies as the way for overcoming these barriers.

Big Data

The volume, velocity and variety of the healthcare data is growing at exponential rates from medical imaging, -omics data, sensor data, EHRs, and natural language text [1]. High dimensional, heterogeneous and poorly annotated datasets are difficult for traditional data mining and statistical learning algorithms

to deal with, and considerable feature engineering is a burden to extract the meaningful insight from the dataset.

End to end learning models for healthcare analytics have emerged as a model of deep learning process [1]. Although resolving the challenges of model interpretability, lacking integration of domain knowledge, and avoiding data bias, is desired. Big data in healthcare faces high challenges in storage, processes and visualization because of the tremendous data size, velocity and variety [2]. Biomedical data require high end computing solutions and robust data handling infrastructures to be able to derive value from it [3]. Seamless data preprocess, store and retrieve practices are needed to integrate AI driven analytics with the healthcare IT.

Challenges

Interoperability in healthcare IT systems has been a longstanding problem because it impedes easy data exchange between diverse systems and real time analytics [4]. Despite the continued adoption of EHR's after the HITECH Act significant work is left in data standardization. Structured data representation in biomedical ontologies such as ICD, SNOMED CT, RxNorm etc. facilitates NLP and clinical decision making [4].

But the real time interoperability comes with integrating multiple standards, dealing with the inconsistencies in the data and ensuring seamless communication between different system. In addition, data interoperability and real time analytics have gotten more complicated due to the growing use of the Internet of Medical Things (IoMT). For instance, IoMT enables a remote patient monitoring, epidemic detection as well as elderly healthcare services, while facing latency, fault tolerance and energy efficiency problems [5].

In order to overcome these challenges, we propose a fog-based architecture which processes the healthcare data nearer to the source by using edge computing [5]. Nevertheless, data engineering frameworks are complicated to understand and real time workload on heterogeneous data streams.

Security and Privacy

Data engineering in AI driven healthcare IT systems is a critical challenge caused by cybersecurity threat in AI. Healthcare has been digitalized; therefore, it has increased cybercrime and patients as well as providers are not willing to adopt Health Information Technologies (HITs) [6]. The implementation of the same is hindered by security gaps, usability constraints, and lack of standardization for secure data exchange across the healthcare institutions.

Because there exist intrusions to sensitive patients' data, a holistic cybersecurity is needed. With the advent of the expansion of IoT and 5G technologies, it has come to light the presence of new vulnerabilities that need robust threat detection mechanisms. However, given that IoT based healthcare system is prone to denial of service (DoS) attacks, Mirai botnets and man in the middle (MITM) attacks, the requirements of IoT based healthcare system are machine learning based anomaly detection models [7]. Data privacy regulations have to be taken very seriously and strict applied at the same time which in turn improves accuracy of threat detection models through deep feature engineering techniques.

Emerging Solutions

When Intelligent Process Automation (IPA) and data engineering come together, it has been demonstrated to enhance Healthcare IT efficiency and innovation [8]. This is made by the use of collaborative approach between data engineers, clinicians and IT professionals to improve data quality, continue to enhance and optimize AI driven analytics.

CPNP concludes case studies which can provide readers with an idea about how IPA has been transformative in helping healthcare workflows, when used effectively, to streamline and enhance patient outcomes [8]. Real time data processing in healthcare IT is promising to get revolutionized with the arrival of 5G technology. Federated learning is supported under reduced latency and increased bandwidth, which allow the training of AI models by federating the contribution of the data from distributed healthcare networks without compromising the privacy of the data [9].

In addition, network slicing permits multiple independent networks to run on the same physical structure [9]. Nevertheless, security risks brought by virtualization and multi-tenancy can be mitigated through formal verification of 5G networks [10].

Summary

Challenge	Description	Solutions
Data Complexity	High-dimensional [1,3,4]	Deep learning [1,2,3]
Interoperability Issues	Lack of standardized data [4]	Biomedical ontologies [4]
Analytics	Fault tolerance [5]	Fog computing [5]
Cybersecurity Risks	Cyber threats IT [6,7]	Threat detection [6,7]
Emerging Technologies	Scalable analytics [8,9]	Process Automation [8,9]

III.METHODOLOGY

Qualitative research methodology has been used for this literature review to analyse and aggregate the existing works on the data engineering challenges in the healthcare IT systems that are driven by AI. It conducts a review of journal articles and conference papers on interoperability, real-time analytics, data security and big data integration and provides the synthesis of best practice and innovation approaches. The thematic analysis is undertaken to identify the main streams, difficulties and suggested resolutions in the subject.

IV.RESULT AND DISCUSSION

AI Adoption in Healthcare and IT

Figure 1 clearly shows that the steady adoption of AI and big data in healthcare IT has grown from 2017 to 2023 and that this proves an increasing reliance on data driven decision making. Its acceleration starting from 2020 is in keeping with the deep learning and cloud-based solutions [2][5]. The most critical
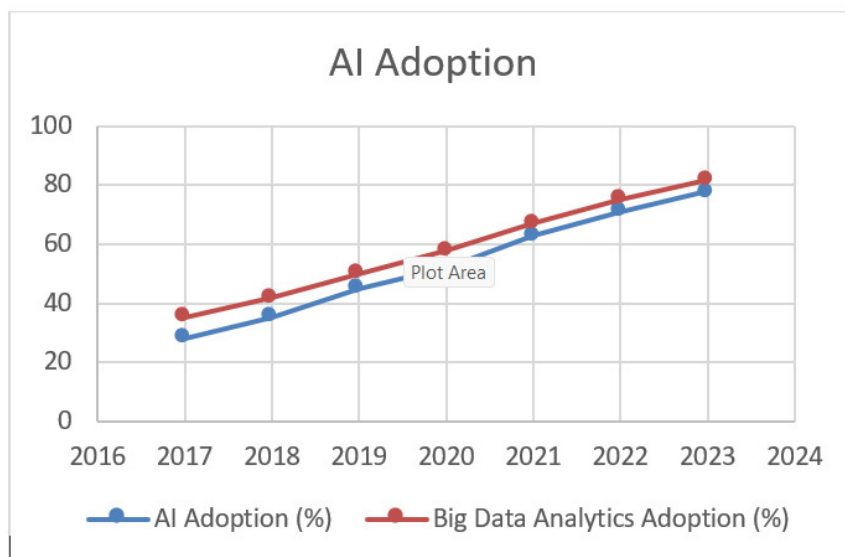

Fig. 1 AI adoption rates (Self-made)

challenge was said to be security and privacy issues (72%), incompatible data standards (65%) and lack of EHR integration (58%) [4]. These barriers indicate that interoperability will succeed only if standards and cross platform collaboration are successful (Figure 2).

   In 2017, the number of reported cybersecurity breaches increased from 112 to 400 in 2023, and the patient records affected soared from 24 million to 105 million [7]. This points to the increase in health systems' vulnerability to cyber-attack and the necessity of strengthened measures such as blockchain, hardening products with encryption, and artificial intelligence executive threat detection, as seen in Figure 3.
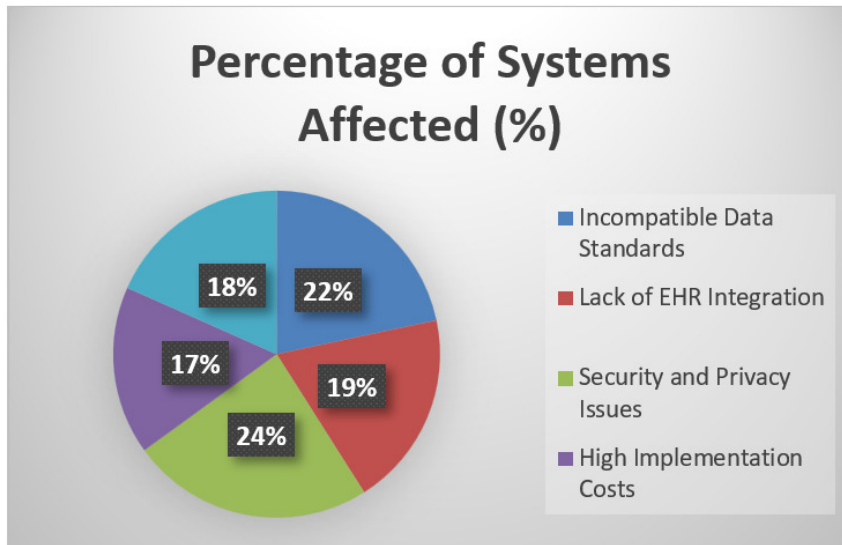
Percentage of Systems Affected (%)

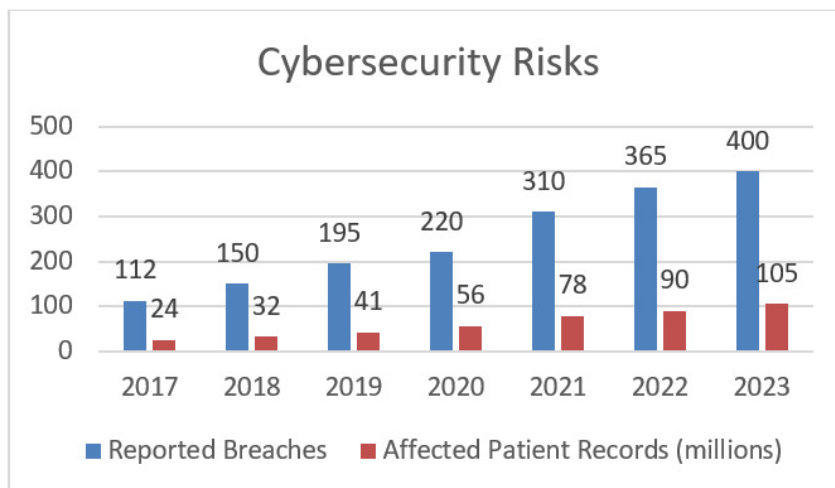**Fig. 2** Effect of AI on organisations (Self-made)

Fig. 3 Cybersecurity Incidents (Self-made)

V.CONCLUSION

The adoption of healthcare IT systems that run on AI is yet to take off due to high degrees of data interoperability, security and real time analytics. Despite that, adoption rates are on the rise, and yet cybersecurity threats and a lack of common data frameworks continue to block. The findings indicate emphasis on area of integrating standardized data, workforce training and AI advanced security solutions. Complement this, it is essential to address these issues for AI to enable optimality in healthcare and increase efficiency, accuracy and safety of the patient in a digitized healthcare environment.

References

[1] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236-1246, 2018. https://doi.org/10.1093/bib/bbx044

[2] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," *Journal of Big Data*, vol. 6, no. 1, pp. 1-25, 2019. https://doi.

`org/10.1186/s40537-019-0217-0`

[3] M. Adibuzzaman, P. DeLaurentis, J. Hill, and B. D. Benneyworth, "Big data in healthcare–the promises, challenges and opportunities from a research perspective: A case study with a model database," *AMIA Annual Symposium Proceedings*, vol. 2017, p. 384, Apr. 2018. `https://pmc.ncbi.nlm.nih.gov/articles/PMC5977694/pdf/2731403.pdf`

[4] Z. He, C. Tao, J. Bian, M. Dumontier, and W. R. Hogan, "Semantics-powered healthcare engineering and data analytics," *Journal of Healthcare Engineering*, 2017. `https://doi.org/10.1155/2017/7983473`

[5] G. Manogaran and D. Lopez, "A survey of big data architectures and machine learning algorithms in healthcare," *International Journal of Biomedical Engineering and Technology*, vol. 25, no. 2-4, pp. 182-211, 2017. `https://doi.org/10.1504/IJBET.2017.087722`

[6] R. Boina, A. Achanta, and S. Mandvikar, "Integrating data engineering with intelligent process automation for business efficiency," *International Journal of Science and Research (IJSR)*, vol. 12, no. 11, pp. 1736-1740, 2023. `https://www.researchgate.net/profile/Shreekant-Mandvikar/publication/375952622_Integrating_Data_Engineering_with_Intelligent_Process_Automation_for_Business_Efficiency/links/65656fcd3fa26f66f432b3f4/Integrating-Data-Engineering-with-Intelligent-Process-Automation-for-\Business-Efficiency.pdf`

[7] P. Natsiavas, J. Rasmussen, M. Voss-Knude, . Votis, L. Coppolino, P. Campegiani, and V. Koutkias, "Comprehensive user requirements engineering methodology for secure and interoperable health data exchange," *BMC Medical Informatics and Decision Making*, vol. 18, pp. 1-16, 2018. `https://doi.org/10.1186/s12911-018-0664-0`

[8] A. A. Alsulami, Q. Abu Al-Haija, A. Tayeb, and A. Alqahtani, "An intrusion detection and classification system for IoT traffic with improved data engineering," *Applied Sciences*, vol. 12, no. 23, p. 12336, 2022. `https://doi.org/10.3390/app122312336`

[9] V. S. Naresh, S. S. Pericherla, P. S. R. Murty, and S. Reddi, "Internet of Things in Healthcare: Architecture, Applications, Challenges, and Solutions," *Computer Systems Science & Engineering*, vol. 35, no. 6, 2020. `https://cdn.techscience.cn/uploads/attached/file/20201211/20201211004923_25523.pdf`

[10] D. Loghin, S. Cai, G. Chen, T. T. A. Dinh, F. Fan, Q. Lin, and Z. Zhang, "The disruptions of 5G on data-driven technologies and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1179-1198, 2020. `https://doi.org/10.1109/TKDE.2020.2967670`