Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240485

How Interoperability Can Enable Artificial Intelligence in Clinical Applications

Filip REHBURG[a,1], Adam GRAEFE[a], Miriam HÜBNER[a] and Sylvia THUN[a]
[a] *Berlin Institute of Health at Charité, Germany*
ORCiD ID: Filip Rehburg https://orcid.org/0009-0007-0457-5724

Abstract. This paper explores the critical role of Interoperability (IOP) in the integration of Artificial Intelligence (AI) for clinical applications. As AI gains prominence in medical analytics, its application in clinical practice faces challenges due to the lack of standardization in the medical sector. IOP, the ability of systems to exchange information seamlessly, emerges as a fundamental solution. Our paper discusses the indispensable nature of IOP throughout the Data Life Cycle, demonstrating how interoperable data can facilitate AI applications. The benefits of IOP encompass streamlined data entry for healthcare professionals, efficient data processing, enabling the sharing of data and algorithms for replication, and potentially increasing the significance of results obtained by medical data analytics via AI. Despite the challenges of IOP, its successful implementation promises substantial benefits for integrating AI into clinical practice, which could ultimately enhance patient outcomes and healthcare quality.

Keywords. Artificial Intelligence, Interoperability, Standardization, Digital Medicine

1. Introduction

Artificial Intelligence (AI) is emerging as a state-of-the-art technology for data analytics in the medical field [1-6]. It has demonstrated groundbreaking performance in various applications, from medical image analysis to diagnosis predictions and drug development [7]. This development is driven by a surge in digital data collection, providing the evidence necessary to train AI models. The technology relies on large, extensive datasets to produce reliable results. Thus, analyzing medical data from diverse sources is imperative to maximize the full potential of AI.

However, while studies have produced convincing results, adaptation of AI in clinical practice has been slow [7-10]. This is due to a lack of standardization in the medical sector. Creating comprehensive datasets requires combining data from various systems and locations, such as routine data, specialized documentation, molecular evidence, and outpatient data. However, data from different sources is often unstandardized, unstructured, and ambiguous, leading to incompatibility and imprecision. Restoring the integrity of a combined dataset requires significant effort, which hinders the research process. [11]

Interoperability (IOP) is the ability of two or more systems to collaborate and exchange information in a coordinated manner [12]. Interoperable data are data with

[1] Corresponding Author: Filip Rehburg; E-mail: filip.rehburg@charite.de.

unambiguously defined vocabularies, standardized formats, and agreed-upon protocols, enabling shared semantic and syntactic understanding. IOP is fundamental for developing and training AI algorithms and models in digital medicine by facilitating the joint processing of data from disparate sources.

Our paper provides an in-depth definition of IOP, dividing it into technical, syntactic, semantic, and organizational Interoperability. We further introduce the FAIR principles and the Data Life Cycle. Afterward, we discuss why IOP is indispensable for the application of AI in digital medicine. During this discussion, we focus on the different stages of the Data Life Cycle (DLC) and how IOP and the use of standardized data positively impact them. Finally, we address the challenges faced by IOP and give an outlook on the potential benefits of successful clinical AI applications.

2. Background and Methods

Interoperability (IOP) can be divided into technical, syntactic, semantic, and organizational IOP [13, 11]. Technical IOP is achieved by shared protocols such as HTTPS or REST for data exchange via the internet. Syntactic IOP determines data formats and structures, e.g., the widespread IOP standard Fast Healthcare Interoperability Resources (FHIR), which can be represented and exchanged via the JSON or XML formats. Another notable example is Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Semantic IOP specifies an unambiguous vocabulary of medical terms consisting of agreed-upon terminologies and ontologies. Examples of semantic IOP standards are the terminology SNOMED CT, the most comprehensive collection of terms for clinical data exchange [13], LOINC, a standard for disseminating laboratory test results and the rare disease ontology ORDO, which provides a vocabulary and defines relationships for terms in the field. Organizational IOP is defined as the cooperation of stakeholders, ensuring universal adaptation of standards. It is to be noted that IOP is only effective if all participating entities agree to the implementation of uniform standards and provide interfaces between standards when no unanimous agreement can be reached [11].

The FAIR principles (Findability, Accessibility, Interoperability, and Reusability) set criteria to enable the discovery and reuse of research data by machines and humans. In summary, they state that data must be enhanced by unique identifiers and detailed meta-data and be available via standard protocols in standardized formats [14].

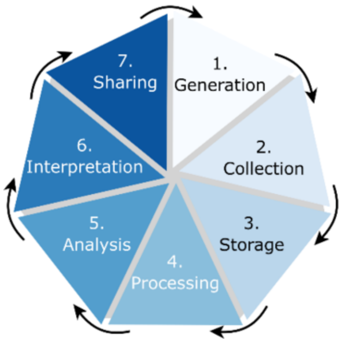

Figure 1. Data Life Cycle for Healthcare Data.

The Data Life Cycle (DLC) is a widespread concept to manage data flow in an organization from generation to destruction [15]. Data is vital for successfully implementing AI; thus, managing it throughout its lifespan is crucial. Over the years, definitions of the DLC, varying in detail and number of steps, have been published, ranging from four [16] or five [17] to seven [18] steps. We decided to include seven steps in a DLC for healthcare data based on the FAIR principles and data requirements for AI. As shown in Figure 1, it contains the steps of data generation, collection, storage, processing, analysis, interpretation, and sharing.

3. Results and Discussion

This discussion will follow data through the Data Life Cycle (DLC), demonstrating how IOP is essential at every step to prepare for the successful application of AI. The Section concludes by addressing the potential challenges of IOP.

Relevant healthcare data can come in various forms, including anamneses documented by physicians and molecular evidence such as genetic findings and laboratory results. Detailed and extensive information is crucial for thorough analysis via AI to support empirical medical decision-making. However, when combining data from such varied contexts, researchers often face heterogeneous terms, requiring significant effort and expertise to interpret their syntax and semantics.

The first step in the DLC is the generation of relevant healthcare data. IOP is critical in the subsequent data collection and storage stages. If healthcare data were collected and stored in standardized formats using unambiguous terminology, they could be used more effectively and efficiently in later life cycle phases. This is not the case in routine practice, where data collection happens in heterogeneous formats, provoking extensive preprocessing to ensure precision, integrity, and correctness when combining data from different sources. While interoperable data capture solutions exist in clinics, these often require medical professionals to document the same information doubly, e.g., for reimbursement and research purposes. Instead, IOP standards can be integrated into documentation applications, e.g., via terminology servers [19], so medical professionals can enter information in natural language and select from a list of terms, internally represented by codes. This can streamline data entry, allowing medical professionals to focus on their field of expertise. Simultaneously, it ensures that data is validated, precise, and interoperable, eliminating the need for redundant documentation.

During the processing phase, data is cleaned and otherwise prepared for analysis. At this stage, healthcare data is often in unstandardized formats, such as tables with local specifications or doctor's letters in natural language. Thus, processing requires the most significant time during the DLC. Local code systems must be mapped onto interoperable terminologies to combine data while maintaining medical meaning. During this process, ambiguities must be resolved, and the results must be validated to prevent the introduction of errors or inconsistencies. This laborious process inhibits research and care and could be avoided considerably by employing interoperable data. IOP standardizes data structure, format, and medical vocabulary, as well as ensures high data quality, facilitating analysis and reducing preprocessing.

It could be argued against the necessity of standardization since AI, in the form of Large Language Models (LLMs), can extract medical meaning from unstructured text [20]. However, this is a transitional solution since LLMs are susceptible to hallucination [20]. Further, their development, testing, and validation require significant effort. This effort could be invested in data analysis for knowledge generation if the data were recorded interoperably.

In medical studies, AI is emerging as a cutting-edge method for data analysis. Utilizing interoperable data sources can facilitate the development of AI algorithms by predefining the data structure. Thus, machine learning algorithms can be developed before the data is received or even collected, enabling parallelization. Furthermore, IOP can counteract the replication crisis of modern medicine [21]. When writing AI with data from interoperable sources instead of local specifications, analysis pipelines can more easily be shared to facilitate replication.

Moreover, IOP can increase the significance and impact of results obtained via AI. One factor is the structure IOP introduces into the data, which inherently encapsulates valuable information. Studies have shown that the success of AI depends on the representation of the data it is trained on [22]. In addition, algorithms based on interoperable data can include information from a broader range of locations; if necessary, this can involve federated learning. The resulting, larger training datasets can increase the performance of AI based on the law of large numbers [23, 24]. This could especially benefit the research into rare diseases, which is often held back by small sample sizes due to low prevalence in local populations [25].

Data sharing, the final step of the DLC, is crucial to empirical research, as stated by the FAIR principles. As part of these principles, IOP plays a fundamental role in enabling the replication of studies by making data and algorithms accessible and reusable. Data sharing contributes to the democratization of medicine by allowing for seamless information exchange between medical professionals and patients. Moreover, access to their data empowers patients to actively participate in medical decisions.

While IOP holds great potential for the application of AI, it does not come without challenges. The large healthcare sector has diverse stakeholders and a challenging environment for broad consensus. Because of financial barriers and resistance to changing practices, providing incentives for applying IOP standards is essential. Additionally, institutions must contend with regulatory hurdles; thus, unified legislation must be implemented to enforce IOP in the sector.

4. Conclusions

Although AI has shown promising results in medical studies, adaption in clinical practice has not kept up with expectations. Medical data is collected on most aspects of a patient's health, its reliable aggregation for AI applications in clinical practice is currently limited by a lack of IOP in the medical field. We demonstrated that implementing IOP can substantially facilitate the development of AI applications at every stage of the Data Life Cycle and support more extensive studies with more significant results. Thus, IOP is instrumental in unlocking the full performance of AI and holds the potential to bring AI into clinical practice.

The potential results on patient outcomes of applying AI in clinical practice are immense. Studies already produced highly accurate AIs that predict diagnoses, recommend suitable treatment, and support decision-making by physicians. Bringing these technologies into the clinic can shorten the path to patient diagnosis and increase treatment quality. To quantify the benefits of using interoperable data for the application of artificial intelligence in clinical applications, we propose a systematic review of the literature as future work.

References

[1] Sahu M, Gupta R, Ambasta RK, Kumar P. Artificial intelligence and machine learning in precision medicine: A paradigm shift in big data analysis. In: Progress in Molecular Biology and Translational Science. 2022. p. 57–100. https://doi.org/10.1016/bs.pmbts.2022.03.002

[2] Johnson KB, et al. Precision medicine, AI, and the future of personalized health care. Clinical and Translational Science. 2020 Oct 12;14(1):86–93. https://doi.org/10.1111/cts.12884

[3] Tong LP, Shi W, Isgut M, Zhong Y, Lais P, Gloster L, et al. Integrating Multi-omics Data with EHR for Precision Medicine Using Advanced Artificial Intelligence. IEEE Reviews in Biomedical Engineering. 2023 Jan 1;1–15. https://doi.org/10.1109/rbme.2023.3324264

[4] He X, Liu X, Zuo F, Shi H, Jing J. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. Seminars in Cancer Biology. 2023 Jan 1;88:187–200.

[5] Bonkhoff AK, Grefkes C. Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence. Brain. 2021 Dec 16;145(2):457–75. https://doi.org/10.1093/brain/awab439

[6] Briganti G, Moine OL. Artificial intelligence in Medicine: Today and tomorrow. Frontiers in Medicine. 2020 Feb 5;7. https://doi.org/10.3389/fmed.2020.00027

[7] Busnatu Ș, Niculescu AG, et al. Clinical Applications of Artificial Intelligence—An Updated Overview. Journal of Clinical Medicine. 2022 Apr 18;11(8):2265. https://doi.org/10.3390/jcm11082265

[8] Petersson L, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Services Research. 2022 Jul 1;22(1).

[9] Sharma M, et al. Artificial intelligence Applications in Health care Practice: Scoping review. Journal of Medical Internet Research. 2022 Oct 5;24(10):e40238. https://doi.org/10.2196/40238

[10] He J, et al. The practical implementation of artificial intelligence technologies in medicine. Nature Medicine. 2019 Jan 1;25(1):30–6. https://doi.org/10.1038/s41591-018-0307-0

[11] Lehne M, Saß J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. Npj Digital Medicine. 2019 Aug 20;2(1). https://doi.org/10.1038/s41746-019-0158-1

[12] Geraci A, et al. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. IEEE Std 610. 2008 Sep 23;1–217. https://doi.org/10.1109/ieeestd.1991.106963

[13] Benson T, Grieve G. Principles of health interoperability. Health information technology standards. 2021. https://doi.org/10.1007/978-3-030-56883-2

[14] Wilkinson MD, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016 Mar 15;3(1). https://doi.org/10.1038/sdata.2016.18

[15] Rahul K, Banyal RK. Data Life cycle Management in Big Data Analytics. Procedia Computer Science. 2020 Jan 1;173:364–71. https://doi.org/10.1016/j.procs.2020.06.042

[16] Zhang J, Symons J, Agapow P, Teo J, Paxton CA, Abdi J, et al. Best practices in the real-world data life cycle. PLOS Digital Health. 2022 Jan 18;1(1):e0000003. https://doi.org/10.1371/journal.pdig.0000003

[17] Labrinidis A, Jagadish HV. Challenges and opportunities with big data. Proceedings of the VLDB Endowment. 2012 Aug 1;5(12):2032–3. https://doi.org/10.14778/2367502.2367572

[18] Griffin PC, Khadake J, LeMay K, Lewis S, Orchard S, Pask AJ, et al. Best practice data life cycle approaches for the life sciences. F1000Research. 2018 Jun 4;6:1618. https://doi.org/10.12688/f1000research.12344.2

[19] Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. Journal of Biomedical Semantics [Internet]. 2018 Sep 17;9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/30223897/

[20] Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. Communications Medicine. 2023 Oct 10;3(1). https://doi.org/10.1038/s43856-023-00370-1

[21] Ioannidis JPA. Why most published research findings are false. PLOS Medicine. 2005 Aug 30;2(8):e124. https://doi.org/10.1371/journal.pmed.0020124

[22] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013 Aug 1;35(8):1798–828. https://doi.org/10.1109/tpami.2013.50

[23] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning. Springer series in statistics. 2009. https://doi.org/10.1007/978-0-387-84858-7

[24] Sun C, Shrivastava A, Singh S, Gupta A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. 2017 IEEE International Conference on Computer Vision (ICCV). 2017 Oct 1;843–52. https://doi.org/10.1109/iccv.2017.97

[25] Wojtara MS, Rana E, Rahman TKA, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. Clinical and Translational Science. 2023 Aug 30;16(11):2106–11. https://doi.org/10.1111/cts.13619