Research Article

Open Access · CrossMark

Data intelligence through integration in healthcare, research gaps and opportunities

Abstract

The healthcare sector consistently produces a large volume of data. Sources such as patient's medical history, EHRs, clinical trials, billing, wearables, social media, internet, and research provide useful data that can assist healthcare providers in gaining better insight about patient population as well as improving the patient outcomes and experience. Health data integration, the process of merging data from various sources creates opportunities for greater data intelligence. Data integration methods vary based on the quality, quantity, and capabilities of the integrating service and the needs of current and prospective users. The purpose of this literature review is to understand the state of academic research pertaining to data intelligence and data integration in healthcare. The paper explores the use of data intelligence on healthcare, focusing on the integration of artificial intelligence, machine learning, and other tools to create a unified and extensive healthcare data ecosystem. A systematic literature review was conducted to examine the data intelligence and integration in the healthcare sector. We identify main themes from the literature review as well as explore the research gaps and opportunities.

Keywords: health data intelligence, health data integration, clinical data integration, interoperability, medical data fusion, electronic medical records, electronic health records, artificial intelligence, machine learning

Mahalakshmi Nathan,[1] Dasantila Sherifi[2]
[1]Doctoral candidate in Health Informatics at Rutgers University, USA
[2]Assistant Professor and Health Information Management Program Director at Rutgers University, USA

Correspondence: Mahalakshmi Nathan, MS, MBA, is a doctoral candidate in Health Informatics at Rutgers University, USA

Introduction

Data intelligence, also known as business intelligence, has revolutionized healthcare by changing how medical services are provided, patient care is supplied, and healthcare administration is carried out.[1,2] At the core of this transformation is the ability to leverage vast amounts of data, ranging from EHRs and genomic data to real-time monitoring through wearable devices, at the individual and population level. Leveraging such data resources requires integration, the process of merging data from multiple sources to create a unified view for the user. Data intelligence through integration allows for a deeper understanding of patient health status, which in turn enables tailored treatment plans, predictive analytics for disease outcomes, and optimized healthcare delivery. The insights derived from data intelligence hold immense promise in enhancing the effectiveness and efficiency of medical services. Additionally, they can transform patient-centric approaches, resulting in more personalized and responsive healthcare that caters to individual needs.[1]

The advancement of artificial intelligence (AI) and Machine Learning (ML) algorithms has greatly enhanced the possibilities of data intelligence in healthcare, especially when it comes to analysing intricate datasets, recognizing patterns, and generating insights on a scale and speed that surpasses human capabilities. The incorporation of AI and big data analytics into healthcare systems is leading to an emergence of intelligent healthcare solutions that aim to identify health risks, improve disease prediction, detection, and prevention, treatment precision and outcomes, and resource allocation.[2] Research shows that hospitals, health systems, public health and other health entities in many countries are developing different methods and approaches to gather health data intelligence through data integration, sharing, and advanced analytics.[3,4] With growth of data integration and intelligence applications, it is important to explore related academic research for the purpose of identifying new technologies and methods that may possibly be replicated or considered as best practices, as well as for identifying gaps in data governance aspects.

Methods

Literature was systematically collected from digital libraries and databases such as PubMed, ACM Digital Library, IEEE Xplore, Springer Link, Elsevier Science Direct, Google Scholar, as well as specific publishers like Association for Computing Machinery (ACM), BioMed Central, Emerald Publishing, Frontiers, Hindawi, MDPI, SAGE Publications, Nature Publishing Group, Oxford University Press, Springer Nature, Springer Nature and others. The selected keywords included *data intelligence* OR *healthcare data integration* OR *machine learning in healthcare* OR *interoperability in healthcare data* OR *health data intelligence* OR *medical data fusion* OR *electronic health records integration* OR *clinical data integration* OR *health information exchange* OR *electronic medical record* OR *biomedical data integration* OR *AI in healthcare data analysis*. Inclusion and Exclusion criteria are presented in Table 1.

Table 1 Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Peer-reviewed articles, full text	Non-peer-reviewed articles, non-full-text
Articles published in 2019 onwards	Publications prior to 2019
Empirical studies, case studies, systematic reviews, and meta-analyses	Incomplete or preliminary studies
Articles focused on healthcare and incorporated elements of data intelligence, AI, M, clinical data integration, interoperability, data sharing, and multi-source data fusion in healthcare system	Articles focusing only on cloud computing and Internet of Things
Articles written in English language	Articles in languages other than English

Results

The use of search terms yielded 62 articles, which were reviewed and screening by one of the researchers. Reports assessed for eligibility were reviewed by both researchers to determine whether the inclusion requirements were being fulfilled. This process led to the extraction of 22 papers that met all the inclusion criteria and none of the exclusion criteria. Selection process is illustrated in Figure 1.
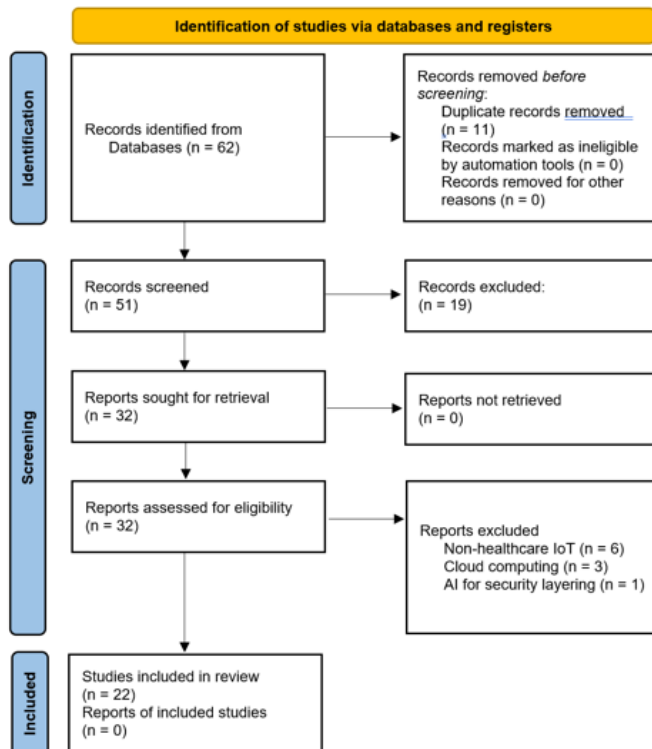

Figure 1 PRISMA Flow Diagram.

This literature review focused on healthcare data intelligence and data integration. Data intelligence is one of the main reasons for investing in the implementation of EHRs, as demonstrated by the meaningful use criteria over a decade ago[5] and health interoperability goals. According to a Global Market Insight Report, the "healthcare business intelligence market size was valued at around $7 billion in 2023 and is expected to grow at a CAGR of 10.4% between 2024 and 2032.[6] Literature review shows various approaches to integration and development of data intelligence tools in supporting clinical decision making, disease management and population health, not just in the United States but also in many other countries. Results of this review are organized into two large groups: studies that focused on integration of systems internally, within a healthcare organization and studies that focused on integration beyond one healthcare organization.

Studies on integration and data intelligence within a hospital or healthcare system

Li et al.[7] created a big data intelligence platform, a Hypertension DATAbase at Urumchi (UHDATA) to leverage hospital data, including electronic medical records of hypertension patients diagnosed since December 2004.[7] The platform updates dynamically using an electronic data collecting system and database synchronization technology, and it identifies accurately secondary hypertension. Parciak et al.[8] proposed an automated system for a maximum-care university hospital.[8] Their infrastructure integrated data from electronic health records, clinical trials, and biobanks. The resulting automated medical data integration infrastructure used ML and AI to process and harmonize data from diverse sources to make it findable, accessible, interoperable, and reusable (FAIR). It was shown that the automated medical data integration architecture improved health data FAIRness, simplified research and clinical decision-making, addressed the constraints of manual "FAIRification", and assured comprehensive data provenance.

Kazemi-Arpanahi et al.[9] used IT infrastructure of a hospital in Iran, as an integration framework and developed a FHIR based communication protocol to achieve interoperability among electrophysiology study-related information systems, hence enabling data transfer, purposeful review of the patients records, and unified reporting in EPS ablation.[9] In China, Lin et al.[10] created a big data intelligence platform for Sun Yat-sen University Cancer Centre.[10] Medical records of nasopharyngeal cancer patients for a period of ten years are brought together from 13 EHRs and other clinical information systems, and updated every seven days. Despite limitations such as data quality and missing data, the platform is being used for nasopharyngeal cancer research.

Boehm et al.[11] created a dataset with pre-treatment CT scans, H&E-stained diagnostic biopsies, and sequencing-inferred HRD status to examine 444 late-stage high-grade serous ovarian carcinoma (HGSOC) patients.[11] ML algorithms were used to enhance risk classification and estimate overall survival (OS) using various data sources, as well as identify quantitative variables linked to prognosis. In Finland, Isoviita et al. developed an open-source system which allowed for extraction, transformation, and loading of diverse datasets and use of cloud-based machine learning (CLOBNET).[12] They used clinical data to train the ML model and predict treatment response in HGSOC patients. The electronic health record data of 208 hospital patients was analysed, and logistic regression model proved to perform well in terms of identifying those with progressive disease or complete response to treatment. The model can be used to address other clinical questions pertaining to the patient population.

Studies focusing on integration and data intelligence beyond one organization

Guo et al.[13] combined data from four sources including county and state data and tested Cox proportional hazard models for colorectal, breast, and lung cancers under three data integration scenarios.[13] Linking datasets to add contextual-level variables to survival models increased model fit and performance for cancer outcomes.

Reda et al.[14] combined fitness and health data from wearable devices and wellness appliances with Linked Open Data and Web Semantic technologies.[14] They formed a common context-aware resource graph with data from different IoT fitness suppliers and combined other open projects and health ontologies into a unique integrated dashboard portal that is also available to consumers. Bridging data silos, device interoperability, and format heterogeneity, allowed for logical reasoning and the analysis of patient behaviour and lifestyle, enabling improved therapeutic decision-making.

Mandl et al.[15] examined interoperable systems and standardized application programming interface (API) effects.[15] Open APIs allow approximately 200 million patients at 15,000 care providers to access and analyse EHR data. The authors draw attention to the increasing need for reliable healthcare data, pointing out gaps in data quality, cross-sector interoperability, use of AI solutions, costly technological expertise, and impact on healthcare inequities.

AlZubi et al.[16] developed an intelligent health system to predict diabetes in smart health cities using big data analytics and Bluetooth technology.[16] They used seven ML models on the PIMA Indian database (including remote sensor), assessed algorithm training, tested for accuracy, and identified the best performing models. Their best performing models, AdaBoost classifier and deep learning algorithm predict diabetes risk by using several data sources, underlining the importance of AI and big data integration in urban healthcare delivery.

Jensen utilized innovative data integration techniques for the Australian Institute of Health and Welfare (AIHW).[17] Data intelligence derived from AIHW has helped identify vulnerable populations and new insights into dementia, suicide, disability and utilization of health services.

Tang developed a framework that combines electronic health records, wearable device data, and environmental data to generate an accurate representation of a person's health.[18] Four Bayesian inference algorithms were applied to hypothyroidism and other related big data and BAN classifier was found to have the best effect in classifying hypothyroidism data. Such applications make it possible to intelligently process, clean, and categorize complex health data, which in turn makes data intelligence more efficient.

Liu et al.[19] developed a DL-based data integration network to detect liver fibrosis in individuals with chronic Hepatitis B.[19] The network uses DL methods to analyse complicated biological data, clinical, biochemical, and imaging information with the goal of increasing the accuracy of liver fibrosis diagnosis, minimizing the need for invasive biopsies, and making therapy interventions more rapid.

Mirzaei et al.[20] used ML to integrate healthcare data from several sources, with the goal of developing and evaluating models that map variables from separate HISB databases into a uniform schema.[20] They selected four HISB-related online databases and conducted two dataset mapping experiments: intra- and inter-database. Through a number of experiments, they identified models that were capable of categorizing over 90-96% of variables depending on the number of datasets used. Accurate and efficient mapping of data affects data intelligence.

Syed et al.[21] evaluated prostate and lung cancer treatment models by using data from 40 radiation facilities as well as an external dataset from Virginia Commonwealth University.[21] The goal was to standardize radiotherapy structure names among healthcare facilities by using multi-view data integration techniques. By using ML they combined geometrical data of structures, related physician text data, and radiotherapy plans, and tested intermediate and late integration methods. The intermediate integration method outperformed in prostate structures and resulted in fewer false positives for organs at risk and planning target volume structures. This enhanced the precision of cancer diagnosis.

Liu et al.[22] developed a cloud-based digital twin healthcare (CloudDTH) model and incorporated large quantities of data from the hospital, family doctors, other clinics, imaging tests, health insurance, pharmacy, home health services, wearable sensors, medical equipment, and community service staff.[22] By deploying simulation and integration with IoT into the CloudDTH architecture they were able to create a real-time monitoring and crisis warning system for elderly patients, prediction of risk, and personalized treatment plans.

Lee at al.[23] developed a multimodal longitudinal data integration framework (MildInt) that was used for deep learning.[23] Clinical records, imaging and biomarker data were analysed and DL methods were used to generate feature representations and train classifiers for disease prediction and progression monitoring. MildInt was then used to predict the development of Mild Cognitive Impairment to Alzheimer's Disease. More broadly, the framework accommodated heterogenous data integration and was capable of processing time series and multimodal data.

Zhao et al.[24] employed ML and DL methods to analyse genetic and longitudinal EHR data of 109,490 people, in order to predict 10-year cardiovascular disease (CVD).[24] Upon comparison with the existing Pooled Cohort Risk Equation for cardiovascular disease risk prediction from American College of Cardiology and American Heart Association (ACC/AHA), they found that their models outperformed the ACC/AHA tool. The research also discovered that combining genetic characteristics with longitudinal EHR data enhanced the prediction accuracy of CVD events.

Zoega et al.[26] shared details about the MedIntel Data Platform created in Australia.[26] This platform has integrated seven existing data sets including administrative data from the Medicare Consumer Directory (MCD); prescribed medications available in the national program Pharmaceutical Benefits Scheme (PBS); medical records available in the Medicare Benefits Schedule (MBS) dataset; mortality data from the national health index (NDI); records of hospitalizations, clinics, and nursing home admissions from the Admitted Patient Data Collection (APDC) dataset; emergency department visit records from the Emergency Department Data Collection (EDDC) dataset; and cancer registry data. Integrating 20 years of medical and health outcomes data for the designated population creates opportunities for better data intelligence insights into the medical needs of the population and better planning of health services. Authors provide details on data content and thresholds used for include the data in the MedIntel Data Platform, as well as patient matching efforts.

Wegner et al.[27] introduced a COVID-19-specific Common Data Model. This model integrates 11 COVID-19 datasets from different regions.[27] The CDM contains 4639 COVID-19 data factors, including age, sex, diagnosis, and disease-specific characteristics like Anosmia and Dyspnea. The COVID-19 CDM currently has 4895 ideas and 4639 typical SARS-CoV-2 data variables. Each data variable has information for standardization. The COVID-19 pandemic demanded worldwide data integration for collaborative study. The lack of standardized and interoperable data was identified as a major barrier to better data intelligence.

Lian et al.[28] investigated the application of hierarchical data fusion methods in intelligent medical monitoring, with the goal of improving the accuracy and reliability of health monitoring systems via the integration of data from different sensors and devices, and medication radio frequency tags.[28] The intelligent telemedicine system can monitor and gather patient data in real-time, decreasing clinical medical staff effort in identifying specific medical services. Fusion algorithms were assessed for medical imaging measures including colour preservation, sharpness, and texture. The fusion approach outperformed previous methods in equipment operation monitoring and picture clarity.

Zhang et al.[29] also developed a data fusion model to bring together medical data from multiple sources for data mining purposes, while maintaining data privacy.[29] Their experiments show that the privacy-free data fusion and mining (PDFM) can effectively search for comparable medical information among the data available from IoH. Data mining techniques were used to uncover patterns and insights within the integrated data, aiming to improve patient care, disease prediction, and treatment outcomes.

All of the articles selected for this review discuss or propose integration methods and/or models for predicting disease, suggesting treatment, or monitoring patient outcomes. Descriptions of the integration approaches, platforms, and/or the predictive models are shared; however, some are more detailed than others, which makes replication or practical use difficult. As mentioned in ten studies, AI, ML, DL is being used in developing predictive models for large heterogenous databases, facilitating the process of data integration and even securing data. Data standards or common data modelling are discussed only in seven studies, which brings questions pertaining to data quality. Integration with IoT and IoH is discussed in at least six studies and concerns about data governance are brought up. Cloud based data fusion from multiple resources is discussed in three of the studies, thus highlighting the potential for achieving greater integration and more insightful data intelligence without necessarily investing heavily in internal data platforms. List of articles is presented in Table 2.

Table 2 List of Articles

S. No.	Author Name	Year	Contribution	Findings
1	Li, N. et al.[7]	2024	Disease-Specific Data Platform	Ensures decisions are based on current, comprehensive data for complex health conditions.
2	Parciak, M. et al.[8]	2023	Data Standardization and Automation	Makes comprehensive, high-quality data available for clinical decision-making and research.
3	Kazemi-Arpanahi, H. et al.[9]	2020	Interoperability and Reporting	Ensures consistent and comprehensive data exchange across systems for EPS procedures.
4	Lin, L. et al.[10]	2019	Real-time Data Updating	Supports real-world studies with continuously updated EHR data, improving data quality for clinical decisions.
5	Boehm, K. M. et al.[11]	2022	Integrates multimodal data for risk stratification in ovarian cancer using ML models.	Enhanced risk stratification, potentially leading to personalized treatment strategies
6	Isoviita, V. M. et al.[12]	2019	Suggests an open-source platform for integrating healthcare datasets and applying ML algorithms.	Facilitated efficient healthcare data integration and analysis, promoting collaborative research efforts
7	Guo, Y. et al.[13]	2020	Predictive Modelling	Improves accuracy of prognostic assessments, enabling personalized treatment plans.
8	Reda, R. et al.[14]	2022	Lifestyle Data Integration	Incorporates patient behavior and fitness data into clinical decisions for personalized care.
9	Mandl, K. D. et al.[15]	2024	Highlights the need for an interoperable data ecosystem to fully leverage AI in healthcare.	Identified key components of a digital data ecosystem that supports AI integration
10	AlZu'bi, S. et al.[16]	2023	Integrates various data sources for diabetes risk prediction using big data analytics.	Achieved high accuracy in diabetes prediction, demonstrating the system's potential for real-world application
11	Jensen, L. R.[17]	2022	Holistic Health Insights	Provides insights into public health and individual well-being for informed care strategies.
12	Tang, H.[18]	2022	Proposes a framework for integrating and analyzing health data from diverse sources using ML techniques.	Improved data processing and classification accuracy, facilitating better healthcare decisions
13	Liu, Z. et al.[19]	2022	Utilizes deep learning for integrating clinical, biochemical, and imaging data for liver fibrosis diagnosis.	Achieved high diagnostic accuracy, outperforming traditional methods
14	Mirzaei, A. et al.[20]	2022	Applies ML methods for integrating healthcare data from multiple HISB databases.	Demonstrated effective integration of diverse healthcare databases, enabling more comprehensive health information analysis
15	Syed, K. et al.[21]	2021	Combines textual and geometric information from radiotherapy plans to standardize structure names.	Achieved high accuracy in standardizing structure names, improving data consistency for radiotherapy

Table 2 Continued....

16	Liu, Y. et al.[22]	2019	Leverages cloud computing and digital twins for real-time health data processing and prediction.	Enhanced healthcare service delivery for the elderly through real-time data analysis and personalized care plans
17	Lee, G. et al.[23]	2019	Integrates multimodal longitudinal data using deep learning for disease prediction and monitoring.	Improved prediction accuracy for various health outcomes, demonstrating the framework's effectiveness
18	Zhao, J. et al.[24]	2019	Utilizes ML and deep learning to analyze EHR and genetic data for CVD risk prediction.	The best AUROC achieved was 0.79, outperforming the baseline approach of the AHA Pooled Cohort Risk equations with an AUROC of 0.73.
19	Zoega, H. et al.[25]	2024	Shares the success of a data platform created by the integration of the existing seven data sets.	Medical, administrative, prescription, cancer registry, mortality, hospital admissions, outpatient visits and nursing home admission data sets were integrated into one platform for better data intelligence.
20	Wegner, P. et al.[26]	2022	Global Health Data Integration	Supports global collaborative research and informed responses to health crises like pandemics.
21	Lian, W. et al.[27]	2020	Hierarchical Data Fusion	Enhances monitoring accuracy for early detection and tailored treatment plans.
22	Zhang, Q. et al.[28]	2020	Data Mining	Uncovers patterns in integrated data to improve disease prediction and treatment outcomes.

Research gaps and opportunities

This review examined the impact that data intelligence has on the healthcare industry, with a particular focus on the significance of data integration in improving patient outcomes and provision of healthcare. We explored the multifaceted approach to clinical data integration, which is pivotal in consolidating patient information from disparate sources, thereby enabling a holistic view of patient health and facilitating personalized treatment strategies. The discussion also highlighted patient-oriented integration, which places the patient at the centre of healthcare delivery, ensuring that data from various patient touchpoints are seamlessly connected to deliver coordinated care. We examined ontology-based data integration, a method that employs structured vocabularies and relationships to harmonize heterogeneous data, thus improving data quality and interoperability across healthcare systems. The integration of AI and ML was identified as a game-changer, with these technologies offering predictive insights and decision support tools that can lead to breakthroughs in diagnosis and treatment plans. The review provides important insights not only into the various approaches that are being used but also into aspects that are not discussed much, such as data quality, ethical considerations. Data intelligence applications are associated with a range of issues and ethical considerations. As organizations increasingly depend on data for decision-making and innovation, it is crucial to acknowledge these challenges and use research as a tool to address them.[30]

Data quality

Data accuracy, completeness, consistency, validity, timeliness, and error rates with patient matching are not addressed in all studies. This is consistent with findings from Perkins.[31] Insufficient data quality can result in inaccurate information which may lead to adverse outcomes.[32] The impact of data quality and consistency on the effectiveness of data intelligence in healthcare is not fully understood. Further studies are needed to establish methodologies for ensuring better data governance, especially when it comes to the accuracy, completeness, and consistency of healthcare data. There is also a gap in the ability to integrate and analyse healthcare data in real-time, which is crucial for acute care settings and immediate clinical decision-making. Research into real-time data processing and analytics is needed. The implementation of low-latency learning in the context of healthcare IoT offers significant opportunities for real-time analysis. It will be essential to develop algorithms capable of adjusting to the ever-changing conditions of IoT devices and networks in order to provide healthcare services in real-time. Data intelligence is certainly valuable but only when it is reliable and timely.

Transparency and explainability

A considerable number of complicated ML models, particularly DNNs, are commonly labelled "black boxes" on account of the convoluted nature of their decision-making procedures. The absence of transparency presents considerable obstacles. For example, when referring to the "clinical data", which data are being extracted from EHR – notes from physicians or other clinicians, coded data, or other. Also, some of the studies describe integration of up to 20 years' worth of medical data but not the common data modelling aspects. Given the various medical record formats that have existed during that period, it is important to address issues of misalignment, data modelling and data mapping. The ability to provide a clear explanation of the reasoning behind a decision is critical, especially in situations that have a direct impact on the lives of individuals. Transparency is fundamental to establishing trust and ensuring accountability,[33] as well as in helping with replication and achieving operational efficiencies. The demand for explainability increases in parallel with the complexity of AI models. Future investigations may focus on the construction of models that are transparent and interpretable, thereby building confidence and comprehension among medical professionals.

Algorithmic accountability

This remains a challenge and it is not addressed in all of the studies reviewed. When automated systems make decisions, attribution might

be difficult to determine. For instance, who bears responsibility if a medical device used by itself results in death? Clearly defined lines of duty and accountability are essential. It is necessary to establish ethical frameworks in order to properly handle complex situations.[34] Data and algorithmic biases have the potential to result in inaccurate medical treatment of particular patient populations or individuals within a hospital setting. This is particularly critical in sectors such as health, finance, employment, and criminal justice, where inaccurate assessments can have serious consequences.[35] The ethical imperative is to guarantee impartiality in applications utilizing data intelligence. It is imperative for developers to make concerted efforts to eradicate bias from datasets and algorithms in order to avoid discriminatory results.

Scalability of data integration solutions

A number of the articles contained within this issue revolve around the federated learning framework and its implementations within the healthcare sector. Subsequent investigations may delve into complex optimization methodologies in an effort to reduce computational demands, enhance model convergence, and further strengthen privacy-preserving mechanisms. As healthcare data continues to grow exponentially, there is a need for scalable data integration solutions that can handle large volumes of data without compromising performance. Clinical language models trained in the clinical domain exist but they are comparatively small in terms of the potential parameters that would be needed to capture unstructured data from EHRs.[36] Research into new technologies and data integration architectures that can support scalability is necessary.

Overreliance on automation

Making fatal mistakes could result from relying too much on automated systems without human oversight. From an ethical prospective, it is crucial to maintain a careful balance between automation and human involvement. When making crucial decisions, human judgment is still essential.

Patient-centred data integration

While there is a focus on EHR, PHR, and EMR, there is a need for more research on patient-centred data integration that includes patient-generated health data from wearables and other personal devices. Additionally, the integration of genomics, electronic health records, and imaging data, continues to be challenging. The investigation of a unified framework capable of efficiently merging multimodal data will facilitate the development of more all-encompassing healthcare analytics.

Last but not least, ensuring **data security and access** amidst the presence of vast quantities of data in a legal environment that is complex is a challenging endeavour. However, adherence to such regulations is a moral obligation.[37] Violations can result in severe effects for organizations, such as financial losses, harm to reputation, and even the loss of life.[39] Only a few studies highlighted in this review address data security aspects. Data security and privacy maybe well preserved in the research summarized in this paper, however, sharing more details on how that is achieved is valuable for future integration and data intelligence efforts. Additionally, the ethical implications of using AI and ML in healthcare, particularly in terms of data privacy and consent, are not fully explored. Specific research in this direction would be helpful in supporting regulations and ethical guidelines and practices for data intelligence in healthcare.

Applications of data intelligence have a lot of promise, but they also come with a lot of gaps and responsibility. Embracing the power

of data for the greater good while reducing potential harm requires addressing technical challenges and ethical concerns. The Joint Commission suggests a number of steps to enhance ethical soundness and efficacy of data-driven efforts, including establishing the ground truth of data, EHR data accuracy, bias, quality of AI tools, and patient relationships.[38]

Limitations

The findings of this review have limitations. Our research primarily relied on articles sourced from PubMed, which, while being a comprehensive database for medical and life sciences literature, may not encompass all relevant studies on the topic, especially those published outside of its scope or in other specialized databases. Moreover, while efforts were made to include a broad range of publication dates, our study may have overlooked recent advancements that could have enriched our analysis. Additionally, our focus on data integration may have led us to overlook the broader context of data management in healthcare, including issues related to data governance, security, which are all critical to the successful integration of data.

Conclusion

This review underscored that data intelligence through integration within healthcare is not just a technological upgrade but a paradigm shift towards a more data-driven and patient-centric approach to healthcare delivery. Addition of patient-generated data, such as information from wearable devices, and genomic data into integration projects can enhance individualized treatments. The insights from data intelligence are instrumental in breaking down silos, optimizing care pathways and ultimately leading to a more informed and effective healthcare ecosystem.

Despite the progress made, the landscape is still marred by several unresolved challenges. The reliability of clinical decisions and analytical outcomes are significantly affected by the quality of integrated data. While rich with insights, the variety of data types and sources add another layer of complexity, in terms of standardization, storage, and interpretation. Ensuring interoperability is crucial since the effectiveness of data integration relies on the ability to exchange and use information seamlessly across various healthcare systems and applications. Furthermore, the capacity to process data in real time is increasingly vital in a healthcare context where timely interventions can be lifesaving.

The integration of such diverse data sources is not only technically demanding but also necessitates careful consideration of privacy and ethical standards. The ethical deployment of AI and ML in healthcare also presents a complex array of considerations. These technologies hold immense potential for predictive analytics, diagnostic accuracy, and treatment personalization, but they also raise questions about privacy, consent, and the potential for algorithmic bias. Ensuring that AI and ML are used responsibly and equitably is crucial to maintain trust and deliver equitable care across the healthcare spectrum.

While data intelligence is set to revolutionize healthcare, the journey is fraught with technical, ethical, and operational hurdles. Bridging these research gaps is not merely an academic exercise but a pressing imperative to ensure that the full potential of data intelligence is harnessed to serve the collective interests of patients, healthcare providers, and researchers. The future of healthcare depends on our ability to navigate these challenges thoughtfully and to forge a path that leverages data intelligence to create a more effective, efficient, and patient-centred healthcare system.

References

1. Manogaran G, Thota C, Lopez D, et al. Big data security intelligence for healthcare industry 4.0. *Cybersecurity for Industry 4.0*. 2017;4:103–126.

2. Bajwa J, Munir U, Nori A, et al. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Health J*. 2021;8(2):e188–e194.

3. Budd J, Miller BS, Manning E, et al. Digital technologies in the public-health response to COVID-19. *Nature medicine*. 2020;26:1183–1192.

4. World Health Organization. *Global action plan on the public health response to dementia 2017-2025. World Health Organization*. 2017.

5. Ferranti JM, Langman MK, Tanaka D, et al. Bridging the gap: leveraging business intelligence tools in support of patient safety and financial effectiveness. *Journal of the American Medical Informatics Association*. 2010;17(2):136–143.

6. Global Market Insights: Healthcare Business Intelligence Market. 2024.

7. Li N, Zhu Q, Dang Y, et al. Development and Implementation of a Dynamically Updated Big Data Intelligence Platform Using Electronic Medical Records for Secondary Hypertension. *Reviews in Cardiovascular Medicine*. 2024;25:104.

8. Parciak M, Suhr M, Schmidt C, et al. FAIRness through automation: development of an automated medical data integration infrastructure for FAIR health data in a maximum care university hospital. *BMC Medical Informatics and Decision Making*. 2023;23:94.

9. Kazemi-Arpanahi H, Shanbehzadeh M, Mirbagheri E, et al. Data integration in cardiac electrophysiology ablation toward achieving proper interoperability in health information systems. *Journal of Education and Health Promotion*. 2020;9:262.

10. Lin L, Liang W, Li CF, et al. Development and implementation of a dynamically updated big data intelligence platform from electronic health records for nasopharyngeal carcinoma research. *The British Journal of Radiology*. 2019;92:1102.

11. Boehm KM, Aherne EA, Ellenson L, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nature cancer*. 2022;3:723–733.

12. Isoviita VM, Salminen L, Azar J, et al. Open source infrastructure for health care data integration and machine learning analyses. *JCO Clinical Cancer Informatics*. 2019;3:1–16.

13. Guo Y, Bian J, Modave F, et al. Assessing the effect of data integration on predictive ability of cancer survival models. *Health Informatics Journal*. 2020;26:8–20.

14. Reda R, Piccinini F, Martinelli G, et al. Heterogeneous self-tracked health and fitness data integration and sharing according to a linked open data approach. *Computing 104*. 2022;104:835–857.

15. Mandl KD, Gottlieb D, Mandel JC. Integration of AI in healthcare requires an interoperable digital data ecosystem. *Nature Medicine*. 2024:1–4.

16. AlZu'bi S, Elbes M, Mughaid A, et al. Diabetes monitoring system in smart health cities based on big data intelligence. *Future Internet*. 2023;15:85.

17. Jensen, LR. Using Data Integration to Improve Health and Welfare Insights. *Int J Environ Res Public Health*. 2022;19:836.

18. Tang H. Intelligent Processing and Classification of Multisource Health Big Data from the Perspective of Physical and Medical Integration. Scientific Programming. 2022.

19. Liu Z, Wen H, Zhu Z, et al. Diagnosis of significant liver fibrosis in patients with chronic hepatitis B using a deep learning-based data integration network. *Hepatology International*. 2022;16:526–536.

20. Mirzaei A, Aslani P, Schneider CR. Healthcare data integration using machine learning: A case study evaluation with health information-seeking behavior databases. *Research in Social and Administrative Pharmacy*. 2022;18(12):4144–4149.

21. Syed K, Sleeman IV WC, Hagan M, et al. Multi-view data integration methods for radiotherapy structure name standardization. *Cancers*. 2021;13:1796.

22. Liu Y, Zhang L, Yang Y, et al. A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE access*. 2019;7:49088–49101.

23. Lee G, Kang B, Nho K, et al. MildInt: deep learning-based multimodal longitudinal data integration framework. *Frontiers in genetics*. 2019;10:617.

24. Zhao J, Feng Q, Wu P, et al. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Scientific*. 2019;9:717.

25. Zoega H, Falster MO, Gillies MB, et al. The Medicines Intelligence Data Platform: A Population-Based Data Resource From New South Wales, Australia. *PDS*. 2024;33(8):5887.

26. Wegner P, Jose GM, Lage-Rupprecht V, et al. Common data model for COVID-19 datasets. *Bioinformatics*. 2022;38:5466–5468.

27. Lian W, Xue T, Lu Y, et al. Research on hierarchical data fusion of intelligent medical monitoring. *IEEE Access*. 2020;8:38355–38367.

28. Zhang Q, Lian B, Cao P, et al. Multi-source medical data integration and mining for healthcare services. *IEEE Access*. 2020;8:165010–165017.

29. Benke K, Benke, G. Artificial intelligence and big data in public health. *International journal of environmental research and public health*. 2018;15:2796.

30. Perkins SW, Muste JC, Alam T, et al. Improving Clinical Documentation with Artificial Intelligence: A Systematic Review. Perspectives in HIM, Summer. 2024.

31. Yaqoob I, Salah K, Jayaraman R, et al. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. *Neural Computing and Applications*. 2022;34:1–16.

32. Alam MN, Kaur M, Kabir MS. Explainable AI in Healthcare: Enhancing transparency and trust upon legal and ethical consideration. *Int Res J Eng Technol*. 2023;10:1–9.

33. Martínez-García M, Hernández-Lemus E. Data integration challenges for machine learning in precision medicine. *Frontiers in Medicine*. 2022;8:784455.

34. Norori N, Hu Q, Aellen FM, et al. Addressing bias in big data and AI for health care: A call for open science. *Patterns*. 2021;2:2021.

35. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *NPJ digital medicine*. 2022;5:194.

36. Shah V, Shukla S: Data Distribution into Distributed Systems, Integration, and Advancing Machine Learning. *Revista Española de Documentación Científica*. 2017;11:1–17.

37. Thapa C, Camtepe S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Computers in biology and medicine*. 2021;129:104130.

38. Ross P, Spates K. Considering the safety and quality of artificial intelligence in health care. *Joint Commission Journal on Quality and Patient Safety*. 2020;46(10):596.