Digital health data infrastructure and analytics with focus on the polish regional centre*

Dawid Pawuś[1,2,*], Helena Ibrahim[1,2], Karol Fabianek[1,3] and Grzegorz Mehlich[1]

[1] *Departament of IT, Opole University Hospital, Witosa 26 Street, Opole, 46-020, Poland*

[2] *University of Technology, Prószkowska 76 Street, 45-758 Opole, Faculty of Electrical Engineering, Automatic Control and Informatics, Poland*

[3] *University of Technology, Prószkowska 76 Street, 45-758 Opole, Faculty of Informatics, Poland*

Abstract

The digital transformation of healthcare has led to the rapid expansion of electronic medical documentation (EMD), hospital information systems (HIS), and medical data warehouses (MDWs). These developments, coupled with the emergence of artificial intelligence (AI) and machine learning (ML), offer unprecedented opportunities for clinical decision support, precision medicine, and population health management. This review synthesises literature from 2020 to 2025 and includes grey sources to: (i) describe the evolving landscape of electronic health data infrastructures; (ii) evaluate the architecture and functionality of MDWs; (iii) assess AI and ML applications across clinical and research domains; and (iv) present strategic digital health initiatives globally and in Poland. Special attention is given to the ongoing implementation of the Regional Digital Medicine Centre (RDMC) at the University Clinical Hospital in Opole, a component of Poland's nationwide healthcare digitisation programme. We examine the centre's modular architecture, data governance models, and AI integration strategies, and situate it within broader European efforts like the European Health Data Space. Our findings underline the critical importance of data standardisation, secure infrastructures, and collaborative frameworks to realise the full potential of digital medicine.

Keywords

Electronic Health Record; Health Information System; Medical Data Warehouse; HL7 FHIR; AI; ML

1. Introduction

Over the past decade, healthcare systems worldwide have transitioned from paper-based charts to fully digital electronic medical documentation (EMD). According to recent estimates, more than 95% of hospitals in some high-income countries now use certified electronic health records (EHR) platforms [1, 2]. Simultaneously, hospital information systems (HIS) generate petabyte-scale clinical, operational, and administrative data that can inform clinical practice, health-service planning, and translational research. The COVID-19 pandemic further underscored the necessity of timely, high-quality data to support rapid evidence generation.

*ITTAP'2025: 5th International Workshop on Information Technologies: Theoretical and Applied Problems, October 22–24, 2025, Ternopil, Ukraine, Opole, Poland

[1*] Corresponding author.

[†] These authors contributed equally.

dawid.pawus@usk.opole.pl (D. Pawuś); helena.ibrahim@usk.opole.pl (H. Ibrahim); karol.fabianek@usk.opole.pl (K. Fabianek); grzegorz.mehlich@usk.opole.pl (G. Mehlich)

0000-0003-3308-3474 (D. Pawuś)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-01-02

However, raw EMD alone is insufficient. Data must be integrated, harmonised, and curated in medical data warehouses (MDWs) capable of supporting real-time analytics and secondary use. As data volumes grow, so too does interest in artificial intelligence (AI) and machine-learning (ML) techniques that promise to uncover latent patterns, enable predictive modelling, and automate routine clinical tasks. Yet deploying such techniques at scale remains challenging due to data heterogeneity, privacy regulations, and sociotechnical barriers. This review article addresses these challenges by synthesising contemporary literature (January 2020 – April 2025) and examining emblematic national projects. We place particular emphasis on the Regional Center for Digital Medicine (RDMC) initiative, implemented by, among others, the University Clinical Hospital in Opole, funded by the Medical Research Agency (ABM), and compare it with international exemplars such as the European Health Data Space (EHDS), NIH All of Us, UK Biobank, and Canada's CanDIG and Canadian Precision Health Initiative (CPHI).

For clarity, all abbreviations used in this paper are summarized below:

- EMD – Electronic Medical Documentation;
- EHR – Electronic Health Record;
- HIS – Hospital Information System;
- MDW – Medical Data Warehouse;
- AI – Artificial Intelligence;
- ML – Machine Learning;
- FHIR – Fast Healthcare Interoperability Resources;
- FAIR – Findable, Accessible, Interoperable, Reusable;
- OMOP CDM – Observational Medical Outcomes Partnership Common Data Model;
- RDMC – Regional Digital Medicine Centre.

Where possible, key technical terms have been briefly defined upon first use to ensure accessibility for a broader audience.

2. Review Strategy and Thematic Framework

This narrative review employed a structured and thematic approach to synthesizing recent advancements in the field of electronic health data, intelligent systems, and medical data infrastructure. The objective was to identify, categorize, and critically assess key developments shaping data-driven healthcare between 2020 and 2025. A comprehensive literature search was performed across the databases Google Scholar, PubMed, IEEE Xplore, and Scopus, using Boolean keyword combinations including: "electronic health record", "electronic medical documentation", "health information system", "FHIR", "FAIR data", "medical data warehouse", "machine learning in healthcare", "big data analytics", and "precision health". The inclusion criteria focused on peer-reviewed publications written in English, published between 2020 and 2025. In addition, selected grey literature—such as governmental policy documents, technical white papers, and institutional reports—was analyzed to capture insights into ongoing national and international digital health initiatives.

From an initial pool of over 100 publications, a final set of 54 high-quality sources was selected for full-text review. Studies were included based on relevance to one or more of the following four thematic categories:

- Electronic Medical Documentation and Health Information Systems (EMD & HIS)
- Medical Data Warehousing (MDW) and Infrastructure Integration
- Big Data Analytics and Intelligent Applications in Medicine
- Strategic Initiatives and Governance Models in Digital Health

Qualitative synthesis was performed by mapping findings across these categories to identify technological trends, research gaps, and emerging best practices. Attention was also paid to methodological rigor, scalability, and ethical considerations present in the reviewed works. To support transparency and thematic clarity, Table 1 presents a categorization of the referenced literature by research focus, type of innovation, and domain relevance (see also Fig. 1).

Table 1. Thematic Classification of Reviewed Literature

Thematic Area	Research Focus	Representative References	Key Contributions
EMD & HIS Infrastructure	Standards (FHIR, SMART), interoperability, FAIR principles	[3–9], [5], [6], [13]	Adoption of HL7 FHIR, FAIR data principles, hybrid models integrating OMOP and FHIR, synthetic data for privacy
	Data quality and governance, NLP, blockchain, metadata traceability	[10–13]	Structured improvement initiatives, NLP for unstructured data, blockchain-based governance
	Data security in IoT/IIoT, lightweight encryption, cloud safety	[14–19]	LEAIoT, hybrid AES-ECDH encryption, blockchain with smart contracts, NuCypher re-encryption
Medical Data Warehousing	Data integration, OLAP support, semantic modeling, hybrid warehousing	[20–23], [26]	MDW architectures integrating EHRs and imaging, rare disease detection, patient similarity modeling
	FAIR-aligned infrastructure, hospital-based MDW applications	[22], [24–25], [27]	Fact constellation schema, MDWs in pandemic response, French governance framework
	Genomic and omics data integration, secure modeling for precision medicine	[28], [29]	Multi-level data models, scalable MDWs for genomic-phenotypic linkage
Big Data Analytics in	BDA in clinical decision-making, real-	[30–32], [36]	Empirical evidence from Poland, data streaming in

Healthcare	time data, national surveys		clinical workflows
	Integration of EHRs, imaging, genomics, wearable data	[33], [34], [37]	Predictive modeling, clinical informatics, patient-centered analytics
	Public health datasets, mining techniques, equity in low-resource settings	[34], [38], [40]	Use of SEER, NHANES, MIMIC for risk stratification; challenges in adoption and infrastructure gaps
Intelligent Systems & AI Applications	Clinical Decision Support Systems (CDSS), AI interpretability, DDx tools	[13], [41–46]	Diagnostic support, ethical integration, evaluation of IsabelHealth and Memem7 systems
	ML/DL for imaging, genomics, NLP, drug discovery	[47–54]	AI for image classification, synthetic data generation, precision therapeutics, pharmacovigilance
	Regulatory and ethical challenges, explainable AI, clinical validation	[44], [45], [54]	Algorithm transparency, validation frameworks, clinician trust, patient-centered AI

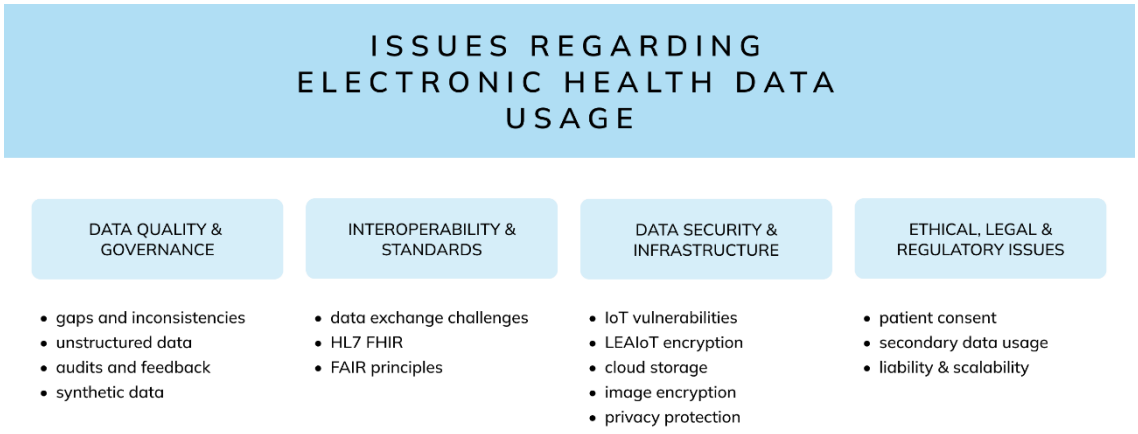

Figure 1: Important Issues Regarding the Use of Electronic Health Data.

Figure 1 illustrates important issues regarding the use of electronic health data. These will be discussed in detail in the following sections. We will then present the gathered knowledge in the context of the RDMC project.

3. Electronic Medical Documentation Landscape

3.1. Definitions and Standards

Electronic Medical Documentation (EMD) encompasses Electronic Medical Records (EMR)—institution-specific digital charts—and broader Electronic Health Records that follow patients across providers. The HL7 FHIR standard has emerged as the de facto technical specification for representing and exchanging granular clinical resources, while initiatives such as SMART on FHIR enable plug-and-play apps within EHR ecosystems [3, 4].

3.2. Interoperability and FAIR Principles

Interoperability remains a persistent hurdle. The FAIR data principles (Findable, Accessible, Interoperable, Reusable) [5, 6] advocate machine-actionable metadata, persistent identifiers, and open vocabularies. Hybrid approaches combining FHIR resources with OMOP CDM tables are increasingly common to support both transactional workflows and analytical queries [7–9].

3.3. Data Quality and Governance

High-quality data is essential for the effective use of electronic medical documentation in analytics, clinical decision-making, and research. However, EMD often contains gaps, inconsistencies, and unstructured content that limit its usability.

Authors of [10] developed a quality improvement program using audits and feedback cycles to assess EMR documentation completeness across clinical disciplines, enabling targeted interventions and real-time compliance monitoring. Similarly, technological advancements such as natural language processing (NLP) and blockchain can enhance data structure, interoperability, and traceability. Yet, as [11] note, barriers like regulatory complexity, cost, and limited scalability persist. Public datasets like MIMIC-IV [12] illustrate the research potential of real-world EHRs but also highlight common challenges such as noisy, sparse, or biased data. Modular database design and rigorous de-identification are critical to support secondary use while protecting patient privacy. Synthetic data offers an emerging solution. As authors of [13] argue, it can enhance privacy, expand small datasets, and support simulation-based research. However, its clinical utility depends on fidelity to real-world distributions, careful bias mitigation, and regulatory safeguards like differential privacy and data custody frameworks.

Robust data governance—spanning quality monitoring, ethical oversight, and technological safeguards—is essential to unlock the full potential of EMD in healthcare innovation.

3.4. Data Security

As healthcare increasingly relies on digital infrastructures, ensuring the security and integrity of medical data is critical. The reviewed literature presents advanced strategies to address the challenges posed by sensitive data handling, especially within IoT, cloud, and IIoT-based systems. IoT-based healthcare systems offer efficient data collection and monitoring but expose vulnerabilities in wireless communication and lightweight devices. To mitigate this, a Lightweight Encryption Algorithm for IoT (LEAIoT) was proposed, significantly reducing hardware usage and improving encryption speeds by over 96% [14].

In real-time medical imaging, cyberattacks threaten data privacy during wireless transfers. A fast image encryption method using novel chaotic maps (LRQ and QRQ) demonstrates high resistance to attacks and efficient processing, encrypting 128×128 images in ~0.03s [15]. Cloud-based storage is essential for large-scale medical data, but traditional encryption is often inefficient or unsupported. An alternative method combining data fragmentation with NoSQL databases offers lower latency than AES, balancing security and performance on public clouds [16]. The integration of blockchain with IIoT in healthcare (BHIIoT) introduces a secure distributed architecture using NuCypher re-encryption and multi-proof mechanisms (PoW and PoS). This approach enhances authentication, traceability, and resilience against data tampering [17].

The Internet of Medical Things (IoMT) also benefits from blockchain-based solutions. By embedding smart contracts into the IoMT ecosystem, sensitive data can be managed with tamper-proof, decentralized access control, enhancing privacy and data integrity [18]. Finally, secure cloud sharing of digitized medical records demands robust encryption. A hybrid scheme using AES and elliptic curve Diffie–Hellman (ECDH) ensures end-to-end security, protecting patient-centric data even in semi-trusted cloud environments [19].

Despite diverse approaches, all studies emphasize the need for lightweight, efficient, and scalable solutions to protect medical data in increasingly complex healthcare infrastructures.

4. Medical Data Warehousing

4.1. Architecture and Integration

The evolution of healthcare data infrastructures has positioned medical data warehouses (MDWs) as central components in the transformation toward data-driven healthcare systems. These repositories aggregate large-scale, heterogeneous datasets from various sources — including electronic health records (EHRs), imaging systems, genomic databases, and administrative records — to enable advanced analytics, clinical research, and decision support [20].

Medical data warehouses enhance data accessibility, support interoperability, and provide structured frameworks for analytical processing through OLAP tools, AI models, and multidimensional queries. In recent years, several frameworks and implementations have demonstrated the capacity of MDWs to facilitate early disease detection, optimize treatment planning, and improve healthcare delivery [21, 22]. For instance, deep learning methods applied to medical warehouses have successfully uncovered predictive patterns in clinical datasets, revealing insights that are otherwise hidden in unstructured or high-dimensional data [20]. In cancer diagnosis, particularly breast cancer, specialized MDWs have been developed to integrate demographic, radiological, and textual data, supporting both clinical and investigative decisions [23]. Similarly, by mapping DICOM metadata to clinical data warehouse structures, hybrid query systems now allow combined analysis of clinical and imaging data, providing richer insights for researchers and clinicians alike [21]. These developments underscore the role of semantic interoperability and multi-level data modeling in modern MDW architectures.

Automated infrastructures also contribute to FAIR data principles (Findable, Accessible, Interoperable, Reusable), enabling more efficient data integration, pseudonymization, and provenance tracking in hospital environments [22]. In pandemic scenarios, such as COVID-19, MDWs with bottom-up construction methodologies and fact constellation schema models have

supported rapid, multidimensional querying for population-level monitoring and policy response [24]. Targeted applications of MDWs are also evident in disease-specific domains, such as HIV/AIDS, where data warehousing allows longitudinal tracking of patient histories and facilitates evidence-based interventions through OLAP analysis [25]. In rare diseases like ciliopathies, patient-patient similarity models leveraged on MDWs have shown promise in diagnostic screening, identifying high-risk individuals from large, unbalanced datasets using medical concept embeddings and ranking strategies [26].

Despite these advances, multiple challenges remain. Data heterogeneity, fragmentation across healthcare systems, and inconsistent terminologies complicate data integration [20, 24, 25]. Ethical and legal considerations, especially concerning patient privacy, consent, and governance models, remain unresolved in many institutional settings [27]. As highlighted in the French case study, sustainable governance, transparency in data transformation processes, and technical standardization are essential to fully realize the benefits of MDWs in clinical research and practice [27]. Finally, the integration of omics data with clinical repositories is emerging as a critical direction for precision medicine. Secure, scalable MDWs that incorporate genomic and phenotypic data enable more granular disease modeling and therapeutic targeting, though they also amplify the technical and ethical complexities of healthcare analytics [28, 29].

In conclusion, medical data warehousing is a cornerstone technology in the ongoing digital transformation of healthcare. Its success depends on interdisciplinary collaboration, robust technical infrastructure, adherence to privacy standards, and alignment with clinical and research objectives.

4.2. Medical Big Data and Analytical Use Cases

A critical pillar of innovation in contemporary healthcare is the integration of big data analytics (BDA), particularly within the broader landscape of digital transformation and Industry 4.0. The convergence of Internet of Health Things, cyber–physical systems, and machine learning in healthcare systems has enabled value-added and cost-efficient service delivery. One study [30] comprehensively reviewed the role of big data in Industry 4.0, underlining its importance in enhancing resource management, clinical operations, and outcome evaluation across healthcare institutions. The findings emphasize how big data not only transforms information flow but also optimizes processes at multiple levels of healthcare operations.

The adoption of BDA is not confined to theoretical benefits; empirical studies reveal tangible progress in medical facilities, particularly in Poland. A national survey [31] involving 217 institutions demonstrated that structured and unstructured data are increasingly integrated into clinical and administrative decision-making. Data sources range from databases and sensors to documents and emails, indicating a systemic shift toward data-driven healthcare. This aligns with broader trends identified in a review by [32], which underscores how big data enhances clinical decision support, resource allocation, and precision medicine. Importantly, this transformation is not without its challenges, with data quality and interoperability remaining key concerns.

Healthcare analytics continues to evolve as a tool for evidence-based decision-making, drawing from a wide spectrum of data including EHRs, imaging, genomics, and patient-generated health data. According to [33], these diverse sources support both quantitative and qualitative analyses that drive outcome-focused clinical decisions. Similarly, [32] highlights the intersection of genomics and big data as a frontier for personalized medicine, where machine

learning algorithms enable individualized treatment strategies by analyzing massive volumes of omics data. These developments not only facilitate the design of robust predictive models but also advance the field of computational medicine.

Data mining methods have gained prominence in extracting value from large-scale public health databases. A review by [34] illustrates how resources like SEER, NHANES, and MIMIC have enabled the construction of predictive models and clinical decision-support tools. Despite the heterogeneous nature of such datasets, data mining techniques have shown remarkable potential in evaluating patient risk and disease progression, thereby enhancing the overall utility of clinical big data.

Another dimension of data-driven healthcare involves Electronic Health Records (EHRs), whose adoption remains uneven across countries. A bibliometric analysis [35] traced global trends in EHR research, noting the United States' leadership in publication volume. The study identifies technological and policy-driven disparities in adoption rates, with recurring themes including health information technology, e-health, and the technology acceptance model.

Recent attention has also shifted toward real-time data streaming, which supports proactive healthcare delivery. A review by [36] explores how real-time monitoring and predictive analytics allow clinicians to anticipate complications and intervene early. Use cases involving wearable technologies and streamlining clinical trials exemplify how data streaming enhances personalized care, although concerns regarding security, infrastructure, and interoperability remain prevalent.

A patient-centric perspective on big data is further elaborated in [37], where data analytics is portrayed as a transformative force in tailoring healthcare delivery. By leveraging EHRs and wearable devices, clinicians can employ predictive models and machine learning algorithms to offer individualized care plans and optimize operational workflows. This transition, however, necessitates ethical oversight, particularly regarding data privacy and security.

Finally, the broader implications of digital disruption in healthcare are captured in [38], which highlights both opportunities and managerial challenges associated with large-scale data analysis. Drawing from case studies in the UAE and interviews with stakeholders across global institutions, the study reveals a gap between the rapid growth of medical data and the healthcare system's ability to effectively utilize it. Recommendations focus on enhancing infrastructure, training, and digital integration to fully leverage the potential of big data analytics.

The integration of big data technologies into healthcare has become a transformative force in modern medicine. Medical informatics, situated at the intersection of healthcare and information technology, is playing a critical role in improving clinical outcomes, optimizing healthcare delivery, and reducing operational costs. The increasing volume, velocity, and variety of healthcare data — from electronic health records, imaging systems, and wearable devices to genomic data and unstructured clinical notes — has created new opportunities and challenges for the medical field [30].

Big data analytics enables predictive modeling, early disease detection, and personalized treatment planning. It facilitates evidence-based decision-making by offering real-time insights drawn from vast datasets [39]. Moreover, it supports population health monitoring, drug discovery, and resource allocation in hospitals and public health institutions [24, 34]. In particular, five core subfields benefit from big data analytics: medical image analysis,

bioinformatics, clinical informatics, public health informatics, and medical signal analytics — each with specific tools, data repositories, and analytical workflows [33].

However, the implementation of big data solutions in healthcare comes with significant technical and ethical challenges. Issues such as data heterogeneity, lack of standardization, and interoperability barriers between systems hinder seamless data exchange and integration [40]. Furthermore, the security and privacy of patient information remain paramount, especially when data is distributed across multiple platforms and accessed remotely [30]. Despite their potential, many healthcare systems — especially in rural or low-resource settings — face barriers to adopting medical informatics solutions. These include poor infrastructure, workforce shortages, and limited financial investment [34]. Nevertheless, technologies such as telemedicine, mobile health applications, and cloud-based EHRs are gradually bridging these gaps and improving healthcare equity in underserved areas [34]. Finally, the discipline of biomedical informatics continues to evolve as both a scientific and applied field. It not only addresses technical innovations but also engages with broader economic, ethical, and social considerations that shape healthcare systems [40].

5. Intelligent Systems in Clinical Practice

5.1. Decision Support Systems in Medicine

A dynamic and rapidly evolving domain within clinical practice is the development and implementation of intelligent systems designed to support medical decision-making. Clinical Decision Support Systems (CDSSs) play a central role in this landscape by offering patient-specific, context-aware recommendations that enhance diagnostic accuracy, treatment planning, and overall care delivery [41]. The integration of artificial intelligence into CDSSs has shown notable effectiveness in improving clinical outcomes, reducing medication errors, and supporting evidence-based decisions in diverse healthcare settings [42, 43].

Studies emphasize the importance of designing CDSSs that align with human, technological, and organizational contexts, particularly in primary care, where system usability and workflow integration are crucial [13]. Equally important are considerations related to professional identity and clinician autonomy, as CDSSs may be perceived either as supportive tools or as threats to clinical expertise, depending on implementation strategies and user engagement [44]. Interpretability of AI-based CDSSs remains a core concern, prompting ongoing research into transparent algorithmic models and effective ways to present AI-driven recommendations in a medically meaningful and comprehensible manner [45].

Ethical aspects of CDSS use, particularly in the context of healthcare resource allocation, are gaining attention. AI-driven systems must balance efficiency with fairness, and transparency in algorithm design is essential to uphold trust, equity, and patient-centered values in clinical environments [46]. Collectively, these studies highlight not only the transformative potential of CDSSs but also the need for thoughtful integration strategies, user-centric design, and robust ethical frameworks to ensure their successful adoption in medical practice.

CDSSs include differential diagnosis generators (DDx), which support clinicians by suggesting possible diagnoses based on patient data. A feasibility study assessed two such tools —IsabelHealth and Memem7—across three datasets: real-world cases from Afghanistan, cases from a professional medical forum, and complex cases from the New England Journal of

Medicine. Both tools showed similar overall accuracy in identifying expert diagnoses, though performance varied by dataset. Memem7 performed better on Afghan cases, while IsabelHealth was superior in forum cases. Only 27% of cases showed agreement between both tools, yet qualitative analysis indicated they often offered complementary insights. The study suggests that combining DDx tools may enhance diagnostic accuracy and that real-world evaluation is essential for understanding CDSS effectiveness in diverse clinical settings. In critical care, clinical informatics and CDSS have proven particularly valuable, helping clinicians manage massive streams of real-time data and make important decisions under pressure [35].

5.2. Other Machine Learning Solutions in Medical Data

Artificial intelligence (AI) and machine learning (ML) are increasingly transforming healthcare by enabling advanced data analysis, improving diagnostics, and supporting precision medicine. Their applications span image recognition, disease prediction, drug discovery, and clinical decision-making [47, 48]. AI adoption in clinical practice is growing, particularly through supervised and unsupervised ML techniques. These methods enhance diagnostic accuracy, enable patient stratification, and uncover hidden patterns in complex datasets. Non-generative models, including decision trees, SVMs, and neural networks, remain widely used due to their reliability and interpretability [49].

In geriatric care, ML supports individualized treatment planning by accounting for multimorbidity and physiological variability, though successful implementation depends on clinician trust and regulatory approval [50]. Deep learning (DL) further expands ML capabilities, offering improved performance in areas like imaging, genomics, and EHR analysis [48, 51]. Natural language processing enables the extraction of insights from unstructured text such as clinical notes and patient reviews. Techniques like sentiment analysis and topic modeling are proving valuable in pharmacovigilance and patient feedback analysis [52].

AI also accelerates precision medicine, particularly in genomics and personalized care. Algorithms identify biomarkers, predict therapy responses, and optimize treatments based on patient-specific data [51, 53]. Synthetic data is emerging as a solution to privacy and data scarcity, though concerns remain regarding validity and regulation [54]. AI-assisted drug discovery is another growing domain, with ML models reducing time and cost by predicting drug-target interactions and generating novel compounds [53].

Despite progress, challenges such as data quality, model explainability, bias, and clinical validation persist. Continued research and careful integration are essential to ensure safe and effective use of AI in healthcare.

In future iterations, AI modules developed within the RDMC will undergo structured evaluation based on clinical performance metrics such as sensitivity, specificity, ROC-AUC, and calibration reliability. The assessment framework follows guidelines from the SPIRIT-AI and CONSORT-AI extensions, ensuring transparency and clinical validity prior to deployment in hospital workflows.

6. Global Strategic Initiatives

The global healthcare landscape is being transformed by ambitious large-scale programmes that demonstrate how coordinated governance and harmonised data models can unlock the

tremendous value of Electronic Medical Data and Medical Data Warehouses (MDWs). These initiatives represent a paradigm shift toward precision medicine, evidence-based healthcare, and international scientific collaboration.

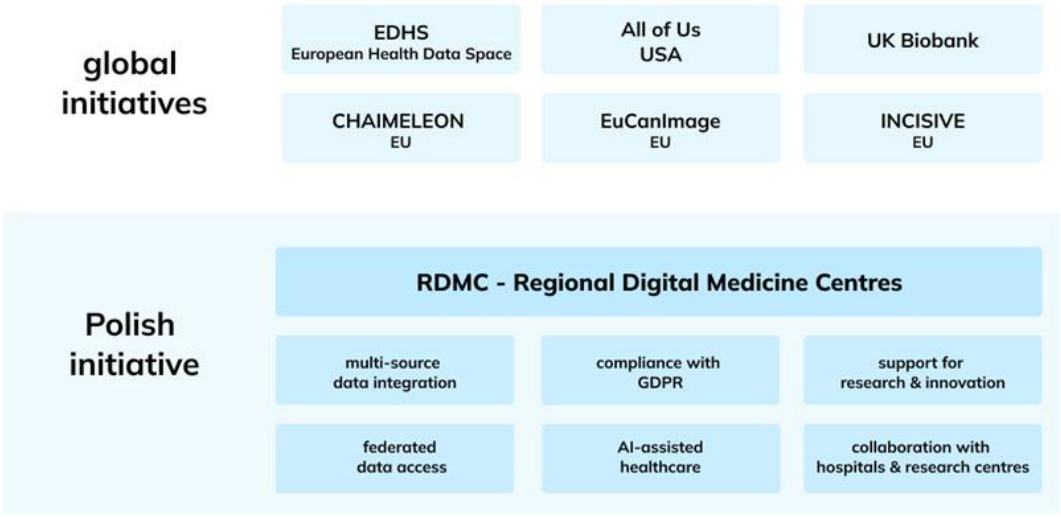

Figure 2: Overview of global and Polish digital health data initiatives.

Figure 2 presents a comparison between major global health data initiatives and the Polish project Regional Digital Medicine Centres (RDMC). The global initiatives include programs such as the European Health Data Space (EDHS), All of Us (USA), UK Biobank, CHAIMELEON (EU), EuCanImage (EU), and INCISIVE (EU). The Polish RDMC initiative focuses on multi-source data integration, compliance with GDPR, federated data access, AI-assisted healthcare, support for research and innovation, and collaboration with hospitals and research centres. Table 2 below summarises a concise comparative analysis between global exemplars and the RDMC initiative, emphasising architectural, organisational, and interoperability aspects.

Table 2. Summary of global examples and RDMC initiatives

Initiative	**Scope**	**Architecture**	**Standards & Interoperability**	**Governance Focus**	**RDMC Relevance**
EHDS (EU)	EU-wide	Federated, multi-layer	HL7 FHIR, GDPR compliance	Legal & ethical interoperability	Aligned – RDMC follows FAIR and GDPR
All of Us (US)	National cohort	Centralised with cloud APIs	FHIR, OMOP	Participant consent, diversity	Similar governance for research data
UK Biobank	Cohort-based	Central data repository	DICOM, OMOP	Data access policy, ethics	Reference model for

(UK)				board	national datasets
RDMC (Poland)	Federated regional centres	Modular, service-oriented	HL7 FHIR, OMOP, DICOM	GDPR, ISO 27001, FAIR	Combines features of all above

This comparative summary highlights RDMC's federated and modular design as its main differentiator, allowing Poland to integrate with both national and European infrastructures while retaining institutional autonomy.

6.1. Regional Digital Medicine Centres in Poland: Building a Federated Infrastructure for Clinical Data and Research Innovation

In 2023, Poland launched an ambitious nationwide program to modernize medical data infrastructure by establishing Regional Digital Medicine Centres (RDMCs). Funded by the Medical Research Agency (MRA), these centres were strategically created within academic hospitals and clinical research institutions to support the country's digital transformation in healthcare and biomedical research.

The RDMCs form a federated network designed to standardize the collection, integration, and analysis of diverse medical data. This includes electronic health records, diagnostic imaging, laboratory results, genomic and other omics data, as well as biobank metadata. The initiative focuses not only on digitizing data but also on ensuring interoperability across systems using international standards like FHIR, OMOP, DICOM, and HL7 CDA.

Each RDMC serves as a data hub, equipped with infrastructure for secure storage, anonymization, and processing of sensitive health data. The architecture is modular and scalable, enabling the integration of artificial intelligence and machine learning tools for real-time and retrospective analyses. One of the key features is a feasibility engine that allows non-technical users to query large datasets, identify eligible patient cohorts, and design data-driven clinical studies.

A significant emphasis is placed on data quality, standardization, and compliance with European data protection laws, including the General Data Protection Regulation (GDPR). RDMCs follow strict protocols for pseudonymization at the local level and full anonymization for inter-centre or national data exchange. The system also includes robust cybersecurity measures, disaster recovery planning, and user accountability mechanisms.

Each centre collaborates closely with a central coordinating unit—the future Digital Medicine Head Office—ensuring strategic alignment, knowledge sharing, and governance across the network. Moreover, the RDMCs are expected to contribute to pan-European initiatives such as the European Health Data Space and the European Open Science Cloud, by adhering to FAIR data principles (Findable, Accessible, Interoperable, Reusable).

Beyond data infrastructure, RDMCs support the development of AI-based clinical decision tools, including predictive models for disease progression, drug safety algorithms, and diagnostic aids. These innovations are powered by integrated datasets derived from routine care and clinical trials. Ultimately, the RDMC program lays the foundation for a national digital

ecosystem that enhances research capabilities, accelerates clinical innovation, and improves patient outcomes through data-driven healthcare [3, 4].

6.2. The European Health Data Space (EHDS): A New Framework for Health Data Access and Reuse in the EU

The European Health Data Space (EHDS), formally established under Regulation (EU) 2025/327, represents a transformative step toward a unified European ecosystem for the governance, access, exchange, and secondary use of electronic health data. As one of the central pillars of the European Health Union, EHDS aims to empower individuals, foster innovation, and strengthen cross-border healthcare delivery and biomedical research within the European Union.

The EHDS establishes a harmonized legal, technical, and governance framework with the following main goals:

- Primary Use: To grant EU citizens immediate, free, and interoperable access to their electronic health records (EHRs) across all Member States, enabling continuity of care regardless of location. Health professionals, with consent, will be able to access patients' medical records—including summaries, prescriptions, diagnostics, and discharge reports—across borders via the MyHealth@EU infrastructure.
- Secondary Use: To facilitate the secure and privacy-compliant reuse of health data—in anonymized or pseudonymized formats—for purposes such as research, innovation, policy-making, regulatory assessment, public health, and crisis preparedness. This will be implemented through a decentralized data infrastructure called HealthData@EU, which links national health data access bodies.
- Single Market Enablement: To support a common market for EHR systems, wellness apps, and AI-based tools through mandatory certification for interoperability and cybersecurity, ensuring a level playing field and stimulating digital health innovation across the EU.

The EHDS complements the General Data Protection Regulation and the NIS 2 Directive, while introducing specific provisions for health data. It requires that all Member States:

- Establish digital health authorities and data access bodies,
- Participate in EU-wide data infrastructures,
- Implement interoperability standards and security requirements,
- Enable patients to access, share, amend, or restrict their health data.

Two harmonized software components—interoperability and logging modules—will become mandatory in all certified EHR systems to ensure technical uniformity and traceability.

The EHDS explicitly prohibits secondary data use for commercial profiling, insurance risk scoring, or employment-related decisions. Researchers and companies must request access via national authorities and use secure processing environments. Results of permitted data uses must be published, ensuring transparency. The regulation also supports the opt-out mechanism for individuals in some Member States, respecting national discretion.

EHDS is expected to unlock an estimated €50 billion in annual economic value through improved data reuse, while addressing longstanding challenges in data fragmentation, access inequities, and cross-border care delivery. It also enhances EU resilience against health crises and cyber threats by mandating robust infrastructure, data traceability, and governance [55].

6.3. The United States' All of Us Research Program

The All of Us Research Program, launched by the U.S. National Institutes of Health (NIH), is a landmark initiative designed to accelerate precision medicine by building one of the world's most diverse longitudinal health research cohorts. With a target enrollment of over one million participants, the program collects a wide range of health-related data—including electronic health records, genomic data, survey responses, biometric measurements, and wearable sensor data—to enable research across a broad spectrum of diseases and health determinants.

A distinguishing feature of All of Us is its focus on diversity and inclusion, particularly among populations historically underrepresented in biomedical research. The program operates under a tiered data access policy, emphasizing strong governance, participant privacy, and ethical use. Through its cloud-based Researcher Workbench, qualified scientists can access curated datasets to investigate disease risk factors, treatment response variability, and health equity challenges.

All of Us is a cornerstone of the U.S. Precision Medicine Initiative and is supported through federal appropriations, including the 21st Century Cures Act. Despite recent budget constraints, the program remains committed to expanding data depth—particularly in genomics and pediatrics—and enabling secure, scalable research that reflects the full diversity of the U.S. population [56, 57].

6.4. The United Kingdom's UK Biobank

UK Biobank is a large-scale prospective cohort study that began participant recruitment in 2006, following several years of preparatory work initiated in 2003. Between 2006 and 2010, it enrolled approximately 500,000 individuals aged 40 to 69 years from across the United Kingdom. The resource was established to enable research into the genetic, environmental, and lifestyle determinants of a wide range of diseases of middle and older age. Participants provided extensive baseline data, including detailed questionnaires, physical and cognitive measurements, biological samples (blood, urine, saliva), and consent for long-term linkage to their health records. Over time, the dataset has been continuously enriched with repeat assessments, electronic health record updates, and molecular data layers, including genome-wide genotyping, whole exome sequencing ($n > 470,000$), and whole genome sequencing ($n \approx 500,000$).

UK Biobank hosts the world's largest multimodal imaging sub-study, comprising MRI, DXA, and ultrasound scans in over 100,000 participants, with ongoing repeat imaging. Additionally, a major proteomics initiative is underway, aiming to quantify up to 5,400 proteins in blood samples from 600,000 timepoints. To support international research access, UK Biobank has developed a secure cloud-based Research Analysis Platform (UKB-RAP), currently used by more than 21,000 approved researchers in over 60 countries. The resource is governed by a robust ethics and access framework, with a strong emphasis on data protection and participant privacy.

UK Biobank is funded by a consortium of UK public agencies, charitable foundations, and industry partners, and continues to be a foundational platform for population-scale precision medicine research globally [58].

7. Regional Digital Medicine Centre (RDMC) at the University Clinical Hospital in Opole: A Process-Oriented Approach to Data-Driven Biomedical Infrastructure

The Regional Digital Medicine Centre (RDMC), currently being implemented at the University Clinical Hospital in Opole, is a flagship initiative within the national Polish programme for digital transformation in healthcare, funded by the Medical Research Agency (ABM). The centre is under active development, with its informatics and systems engineering focus dedicated to building a high-performance, modular infrastructure designed to enable secure, interoperable, and research-ready integration of clinical and omics data. At the core of the evolving RDMC is a multi-tiered, service-oriented architecture intended to combine hospital-based operational systems with dedicated analytical environments for research purposes. The planned infrastructure comprises (see Fig. 3):

- Data warehouse subsystem, integrating structured and unstructured data from source systems such as HIS (Hospital Information System), PACS (imaging), LIS (laboratory), and other telemetry or monitoring devices.
- High-availability storage solutions, scalable to accommodate multi-petabyte datasets, including omics data, medical images, and histopathology slides.
- Data ingestion and harmonisation pipelines, designed to ensure syntactic and semantic interoperability using international standards (e.g., HL7 FHIR for clinical data exchange, DICOM for imaging, and OMOP CDM for research harmonisation).

Data flows are planned to be orchestrated through a unified integration layer that supports ETL (Extract, Transform, Load) processes, with embedded quality control mechanisms, audit logging, and data provenance tracking. To ensure robust data privacy and regulatory compliance, the platform is being built with a multi-layer security model that includes:

- Anonymisation and pseudonymisation pipelines following GDPR and ISO 27001-compliant methodologies;
- Role-based access control (RBAC) and federated identity management across consortium entities;
- Encrypted channels for internal and inter-institutional data exchange;
- Full audit trails, with data usage and modification logs available for governance review.

Each data domain (clinical, imaging, omics) is being made accessible through modular interfaces, with fine-grained access permissions tailored to research scenarios and user clearance levels. The RDMC is being developed following a process-centric and modular approach, structured into logical domains with clearly defined interfaces and responsibilities:

- Data Acquisition – Automated extraction from clinical subsystems, backed by adapters tailored to individual system APIs and formats.
- Data Standardisation & Curation – Harmonisation uses controlled vocabularies (e.g., SNOMED CT, LOINC) and validation against reference schemas. Nonetheless, full semantic interoperability remains a work in progress, as SNOMED CT adoption in Poland is partial and many institutions rely on ICD-10-PL or locally defined terminologies. The RDMC therefore integrates mapping layers to translate national and internal codes to international standards where feasible. The project also collaborates with national initiatives to promote consistent terminology adoption.
- Data Federation – Integration of external datasets from consortium partners (e.g., biobank, genomics centre) through interoperable APIs and shared data contracts.
- Research Enablement – Provision of analytics-ready datasets in isolated research environments, equipped with computational resources and containerised AI toolkits.
- Clinical Decision Support Integration – Development of physician-assistant AI modules to support hypothesis generation, pattern recognition, and early risk prediction based on real-world data.

However, a critical challenge in the Polish context lies in the heterogeneity of legacy hospital systems. Many local HIS installations, particularly outside major academic centres, do not yet expose modern APIs or HL7 interfaces. To mitigate this, the RDMC employs a hybrid approach that combines:

- semi-automated data ingestion workflows for non-standard systems;
- middleware adapters translating proprietary formats to FHIR-compliant messages;
- manual data validation layers where automation is infeasible.

This pragmatic design acknowledges current infrastructural limitations while ensuring gradual alignment with national interoperability goals.

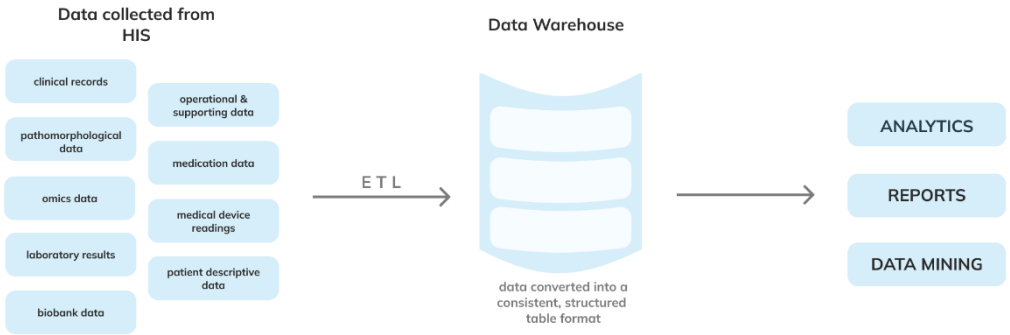

Figure 3: Data flow from source to analytics.

The platform is being engineered to ensure full traceability and repeatability, with each data transformation step documented and reversible. This design principle underpins transparency for regulatory audits and reproducibility of scientific results. A key innovation in the project is the progressive integration of AI modules for diagnostic and prognostic tool development. Planned system capabilities include:

- A dedicated machine learning environment, supporting training and validation of predictive models using de-identified clinical and molecular data.
- A clinical assistant module, leveraging natural language processing (NLP), structured query interfaces, and inference engines to support clinician decision-making at the point of care.
- Reusable pipelines for deep phenotyping, feeding into disease stratification and patient subtyping frameworks.

These components are being developed as microservices and deployed in containerised environments (e.g., Docker, Kubernetes), enabling high scalability and system resilience across both clinical and research settings. All data within the RDMC will follow a well-defined lifecycle:

- Acquisition, Harmonisation, Curation, Analysis, Archiving, with complete data lineage preserved throughout the process, ensuring traceability for audit and clinical review.

The platform is being designed to be fully interoperable with national and international infrastructures, enabling multi-centre studies, federated learning, and generation of real-world evidence.

The RDMC in Opole is not only a regional infrastructure project but is conceived as a strategic node within the emerging national network of Digital Medicine Centres. It is set to contribute significantly to the development of a federated biomedical research ecosystem in Poland. Once fully implemented, its capabilities will support:

- Designing and simulating clinical trials using real-world data cohorts;
- Enabling personalised diagnostics and treatment stratification;
- Supporting regulatory-grade analytics for innovation in medical software and AI tools.

The implementation of the RDMC adheres to the FAIR data principles and aligns with European efforts such as the European Health Data Space, ensuring sustainability, expandability, and relevance on both national and international levels.

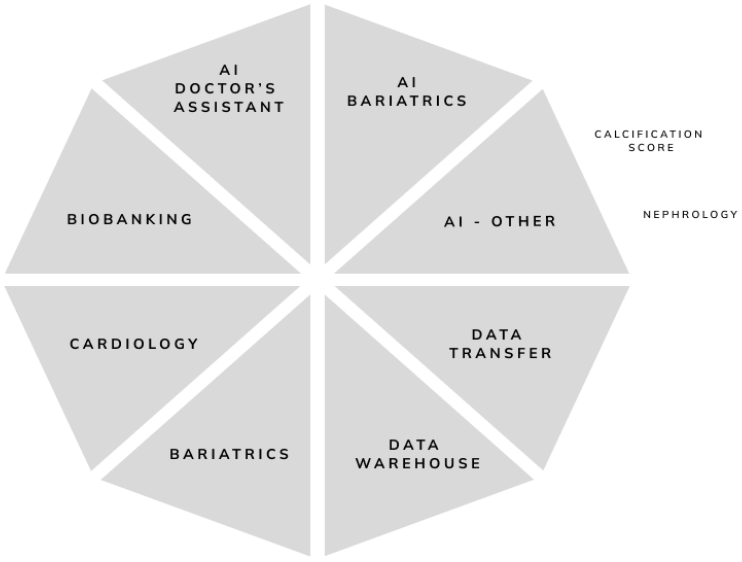

Figure 4: Teams responsible for specific functional areas within the Regional Digital Medicine Center.

As part of the ongoing implementation, the project is organised into dedicated interdisciplinary teams, each responsible for a specific functional area within the Regional Digital Medicine Centre (see Fig. 4):

- Team 1 – AI Clinical Assistant: Focuses on developing AI-based physician support tools to improve descriptive data collection and clinical decision-making in real-world situations.
- Team 2 – AI in Bariatrics: Aims to apply machine learning models to support obesity treatment pathways and patient stratification.
- Team 3 – AI – Other Applications: Develops predictive models in two key areas—coronary artery calcification indexing and nephrology risk profiling.
- Team 4 – Data Transfer: Works on designing and implementing secure, efficient, and interoperable data exchange mechanisms within and across institutions.
- Team 5 – Data Warehouse: Responsible for building the centralised data integration and storage infrastructure that supports analytical and research needs.
- Team 6 – Bariatrics: Develops clinical pathways and research frameworks for metabolic and bariatric medicine within the digital infrastructure.
- Team 7 – Cardiology: Focuses on digital support for cardiology-related data analytics, risk modelling, and clinical workflow integration.
- Team 8 – Biobanking: Designs processes for biological sample collection, annotation, and integration into the broader data ecosystem.

The project is carried out by a national research and clinical consortium composed of leading institutions in healthcare and biomedical science:

- University Clinical Hospital in Opole – project leader;
- Institute of Human Genetics, Polish Academy of Sciences – consortium member;
- Łukasiewicz Research Network – PORT Polish Center for Technology Development – consortium member;
- Institute of Bioorganic Chemistry, Polish Academy of Sciences / Poznań Supercomputing and Networking Center (PSNC) – consortium member;
- University of Opole – consortium member.

This collaborative framework brings together clinical expertise, advanced data infrastructure, genomics, bioinformatics, and AI research capabilities, enabling a comprehensive and sustainable approach to digital transformation in healthcare.

8. Conclusions

The digital transformation of healthcare, accelerated by the widespread adoption of electronic medical documentation, big data analytics, and artificial intelligence, represents a profound shift in how medical knowledge is generated, shared, and applied. As this review has shown, the development and implementation of secure, scalable, and interoperable health data infrastructures are crucial to achieving the promises of precision medicine, population-level analytics, and real-time clinical decision support.

Across global contexts—from the UK Biobank to the NIH's All of Us program and Canada's CanDIG/CPHI initiatives—key success factors include robust governance, adherence to data standards (e.g., HL7 FHIR, OMOP, FAIR), and strong institutional collaboration. These examples highlight the importance of aligning technical innovation with ethical oversight, inclusivity, and regulatory compliance.

In Poland, the Regional Digital Medicine Centres (RDMCs) initiative reflects this alignment, offering a federated, modular model for integrating clinical, omics, and biobank data across multiple domains. The implementation currently underway at the University Clinical Hospital in Opole provides a compelling case study in how such infrastructure can be developed incrementally, with clear responsibilities distributed across thematic teams, and with emphasis on secure data exchange, AI readiness, and research enablement. The project's architecture—including automated pipelines, multi-level data security, and containerised analytics environments—sets a strong foundation for national-scale collaboration and international interoperability.

As healthcare systems move towards more connected and intelligent ecosystems, sustained investment in technical capacity, human expertise, and ethical frameworks will be essential. Future efforts should prioritise reproducibility, scalability, and equity in digital health deployment. If implemented effectively, initiatives like RDMC have the potential to not only improve healthcare delivery at the institutional level, but also to shape international standards for how health data is governed, analysed, and translated into clinical impact.

9. Acknowledgment

This work was financed by the Medical Research Agency in Poland under grant agreements No. 2023/ABM/02/00004-00

10. Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Klimanek, P. (n.d.). Electronic medical documentation - EMR and EHR systems, from https://synappsehealth.com/en/articles/i/electronic-medical-documentation-emr-and-ehr-systems/

[2] Tajammul Pangarkar. (2025, January 14). EHR Industry Statistics 2025 By Digital Record Technology, from https://media.market.us/ehr-industry-statistics/

[3] Regional Digital Medicine Centres. (n.d.), from https://eosc.eu/use-case/regional-digital-medicine-centres/

[4] Regional Digital Medicine Centres (RDMCs). (n.d.). *WHAT ARE REGIONAL DIGITAL MEDICINE CENTRES (RDMCs)?*, from https://abm.gov.pl/en/polish-clinical-trials-network/regional-digital-medicine-cent/270,Regional-Digital-Medicine-Centres-RDMCs.html

[5] Inau, E. T.; Sack, J.; Waltemath, D.; Zeleke, A. A. (2021). Initiatives, concepts, and implementation practices of FAIR (findable, accessible, interoperable, and reusable) data principles in health data stewardship practice: protocol for a scoping review, *JMIR Research Protocols*, Vol. 10, No. 2, e22505

[6] Inau, E. T.; Sack, J.; Waltemath, D.; Zeleke, A. A. (2023). Initiatives, concepts, and implementation practices of the findable, accessible, interoperable, and reusable data principles in health data stewardship: scoping review, *Journal of Medical Internet Research*, Vol. 25, e45013

[7] Rinaldi, E.; Thun, S. (2021). From OpenEHR to FHIR and OMOP data model for microbiology findings, *Public Health and Informatics*, IOS Press, 402–406

[8] Henke, E.; Peng, Y.; Reinecke, I.; Zoch, M.; Sedlmayr, M.; Bathelt, F. (2023). An extract-transform-load process design for the Incremental Loading of German Real-World Data based on FHIR and OMOP CDM: Algorithm Development and Validation, *JMIR Medical Informatics*, Vol. 11, e47310

[9] Xiao, G.; Pfaff, E.; Prud'hommeaux, E.; Booth, D.; Sharma, D. K.; Huo, N.; Yu, Y.; Zong, N.; Ruddy, K. J.; Chute, C. G.; others. (2022). FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common data Model, *Journal of Biomedical Informatics*, Vol. 134, 104201

[10] Jedwab, R. M.; Franco, M.; Owen, D.; Ingram, A.; Redley, B.; Dobroff, N. (2022). Improving the quality of electronic medical record documentation: development of a compliance and quality program, *Applied Clinical Informatics*, Vol. 13, No. 04, 836–844

[11] Negro-Calduch, E.; Azzopardi-Muscat, N.; Krishnamurthy, R. S.; Novillo-Ortiz, D. (2021). Technological progress in electronic health record system optimization: Systematic review of systematic literature reviews, *International Journal of Medical Informatics*, Vol. 152, 104507

[12] Johnson, A. E.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Horng, S.; Pollard, T. J.; Hao, S.; Moody, B.; Gow, B.; others. (2023). MIMIC-IV, a freely accessible electronic health record dataset, *Scientific Data*, Vol. 10, No. 1, 1

[13] Giuffrè, M.; Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: innovation, application, and privacy, *NPJ Digital Medicine*, Vol. 6, No. 1, 186

[14] Kanulla, L. K.; Gokulkumari, G.; Krishna, M. V.; Rajamani, S. K. (2023). IoT Based Smart Medical Data Security System, *International Conference on Intelligent Computing and Networking*, Springer, 131–142
[15] Sarosh, P.; Parah, S. A.; Malik, B. A.; Hijji, M.; Muhammad, K. (2022). Real-time medical data security solution for smart healthcare, *IEEE Transactions on Industrial Informatics*, Vol. 19, No. 7, 8137–8147
[16] Santos, N.; Younis, W.; Ghita, B.; Masala, G. (2021). Enhancing medical data security on public cloud, *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, IEEE, 103–108
[17] Khan, A. A.; Bourouis, S.; Kamruzzaman, M.; Hadjouni, M.; Shaikh, Z. A.; Laghari, A. A.; Elmannai, H.; Dhahbi, S. (2023). Data security in healthcare industrial internet of things with blockchain, *IEEE Sensors Journal*, Vol. 23, No. 20, 25144–25151
[18] Chatterjee, P.; Das, D.; Banerjee, S.; Ghosh, U.; Mpembele, A. B.; Rogers, T. (2023). An approach towards the security management for sensitive medical data in the iomt ecosystem, *Proceedings of the Twenty-Fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 400–405
[19] Kumar, K. P.; Prathap, B. R.; Thiruthuvanathan, M. M.; Murthy, H.; Pillai, V. J. (2024). Secure approach to sharing digitized medical data in a cloud environment, *Data Science and Management*, Vol. 7, No. 2, 108–118
[20] Mishra, N.; Samantaray, S. K. (2023). Review on Knowledge-Centric Healthcare Data Analysis Case Using Deep Neural Network for Medical Data Warehousing Application, *Digital Twins and Healthcare: Trends, Techniques, and Challenges*, IGI Global Scientific Publishing, 193–214
[21] Kaspar, M.; Liman, L.; Morbach, C.; Dietrich, G.; Seidlmayer, L. K.; Puppe, F.; Störk, S. (2023). Querying a clinical data warehouse for combinations of clinical and imaging data, *Journal of Digital Imaging*, Vol. 36, No. 2, 715–724
[22] Parciak, M.; Suhr, M.; Schmidt, C.; Bönisch, C.; Löhnhardt, B.; Kesztyüs, D.; Kesztyüs, T. (2023). 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*, Vol. 23, No. 1, 94
[23] Amara, N.; Lamouchi, O.; Gattoufi, S. (2021). Implementation of a Medical Data Warehouse Framework to Support Decisions, *Advances in Computer Vision and Computational Biology: Proceedings from IPCV'20, HIMS'20, BIOCOMP'20, and BIOENG'20*, Springer, 521–536
[24] Turcan, G.; Peker, S. (2022). A multidimensional data warehouse design to combat the health pandemics, *Journal of Data, Information and Management*, Vol. 4, No. 3, 371–386
[25] Mehmood, A. (2024). An Integrated Data Warehouse to Identify HIV/AIDS Prevalence, *2024 Horizons of Information Technology and Engineering (HITE)*, IEEE, 1–6
[26] Chen, X.; Faviez, C.; Vincent, M.; Briseño-Roa, L.; Faour, H.; Annereau, J.-P.; Lyonnet, S.; Zaidan, M.; Saunier, S.; Garcelon, N.; others. (2022). Patient-patient similarity-based screening of a clinical data warehouse to support ciliopathy diagnosis, *Frontiers in Pharmacology*, Vol. 13, 786710
[27] Doutreligne, M.; Degremont, A.; Jachiet, P.-A.; Lamer, A.; Tannier, X. (2023). Good practices for clinical data warehouse implementation: A case study in France, *PLOS Digital Health*, Vol. 2, No. 7, e0000298
[28] Singh, S. K.; Dhama, A. S.; Kaur, J.; Sharma, N.; Verma, P.; Singh, H. (2024). Omics and clinical data integration and data warehousing, *Integrative Omics*, Elsevier, 225–236
[29] Arnold, C. G.; Sonn, B.; Meyers, F. J.; Vest, A.; Puls, R.; Zirkler, E.; Edelmann, M.; Brooks, I. M.; Monte, A. A. (2023). Accessing and utilizing clinical and genomic data from an

electronic health record data warehouse, *Translational Medicine Communications*, Vol. 8, No. 1, 7
[30] Awrahman, B. J.; Aziz Fatah, C.; Hamaamin, M. Y. (2022). A review of the role and challenges of big data in healthcare informatics and analytics, *Computational Intelligence and Neuroscience*, Vol. 2022, No. 1, 5317760
[31] Batko, K.; Ślęzak, A. (2022). The use of Big Data Analytics in healthcare, *Journal of Big Data*, Vol. 9, No. 1, 3
[32] Arowoogun, J. O.; Babawarun, O.; Chidi, R.; Adeniyi, A.; Okolo, C. (2024). A comprehensive review of data analytics in healthcare management: Leveraging big data for decision-making, *World Journal of Advanced Research and Reviews*, Vol. 21, No. 2, 1810–1821
[33] Tandon, R.; Harnden, A.; Brannan, G. D. (2025). Healthcare Analytics, *StatPearls [Internet]*, StatPearls Publishing
[34] Abdul, S.; Adeghe, E. P.; Adegoke, B. O.; Adegoke, A. A.; Udedeh, E. H. (2024). A review of the challenges and opportunities in implementing health informatics in rural healthcare settings, *International Medical Science Research Journal*, Vol. 4, No. 5, 606–631
[35] Nadkarni, G. N.; Sakhuja, A. (2023). Clinical informatics in critical care medicine, *The Yale Journal of Biology and Medicine*, Vol. 96, No. 3, 397
[36] Shukla, S. (2023). Real-time monitoring and predictive analytics in healthcare: harnessing the power of data streaming, *International Journal of Computer Applications*, Vol. 185, No. 8, 32–37
[37] Ibeh, C. V.; Elufioye, O. A.; Olorunsogo, T.; Asuzu, O. F.; Nduubuisi, N. L.; Daraojimba, A. I. (2024). Data analytics in healthcare: A review of patient-centric approaches and healthcare delivery, *World Journal of Advanced Research and Reviews*, Vol. 21, No. 02, 1750–1760
[38] El Khatib, M.; Hamidi, S.; Al Ameeri, I.; Al Zaabi, H.; Al Marqab, R. (2022). Digital disruption and big data in healthcare-opportunities and challenges, *ClinicoEconomics and Outcomes Research*, 563–574
[39] Rehman, A.; Naz, S.; Razzak, I. (2022). Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities, *Multimedia Systems*, Vol. 28, No. 4, 1339–1371
[40] Shortliffe, E. H.; Chiang, M. F. (2021). Biomedical informatics: The science and the pragmatics, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, Springer, 3–44
[41] Musen, M. A.; Middleton, B.; Greenes, R. A. (2021). Clinical decision-support systems, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, Springer, 795–840
[42] Meunier, P.-Y.; Raynaud, C.; Guimaraes, E.; Gueyffier, F.; Letrilliart, L. (2023). Barriers and facilitators to the use of clinical decision support systems in primary care: a mixed-methods systematic review, *The Annals of Family Medicine*, Vol. 21, No. 1, 57–69
[43] Ouanes, K.; Farhah, N. (2024). Effectiveness of artificial intelligence (AI) in clinical decision support systems and care delivery, *Journal of Medical Systems*, Vol. 48, No. 1, 74
[44] Ackerhans, S.; Huynh, T.; Kaiser, C.; Schultz, C. (2024). Exploring the role of professional identity in the implementation of clinical decision support systems—a narrative review, *Implementation Science*, Vol. 19, No. 1, 11
[45] Xu, Q.; Xie, W.; Liao, B.; Hu, C.; Qin, L.; Yang, Z.; Xiong, H.; Lyu, Y.; Zhou, Y.; Luo, A. (2023). Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: A systematic review, *Journal of Healthcare Engineering*, Vol. 2023, No. 1, 9919269

[46] Elgin, C. Y.; Elgin, C. (2024). Ethical implications of AI-driven clinical decision support systems on healthcare resource allocation: a qualitative study of healthcare professionals' perspectives, *BMC Medical Ethics*, Vol. 25, No. 1, 148
[47] Haug, C. J.; Drazen, J. M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023, *New England Journal of Medicine*, Vol. 388, No. 13, 1201–1208
[48] Chakraborty, C.; Bhattacharya, M.; Pal, S.; Lee, S.-S. (2024). From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare, *Current Research in Biotechnology*, Vol. 7, 100164
[49] Pantanowitz, L.; Pearce, T.; Abukhiran, I.; Hanna, M.; Wheeler, S.; Soong, T. R.; Tafti, A. P.; Pantanowitz, J.; Lu, M. Y.; Mahmood, F.; others. (2025). Nongenerative artificial intelligence in medicine: advancements and applications in supervised and unsupervised machine learning, *Modern Pathology*, Vol. 38, No. 3, 100680
[50] Woodman, R. J.; Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future, *Aging Clinical and Experimental Research*, Vol. 35, No. 11, 2363–2397
[51] Recharla, M.; Chakilam, C.; Kannan, S.; Nuka, S. T.; Suura, S. R. (2025). Harnessing AI and Machine Learning for Precision Medicine: Advancements in Genomic Research, Disease Detection, and Personalized Healthcare, *American Journal of Psychiatric Rehabilitation*, Vol. 28, No. 1, 112–123
[52] Harrison, C. J.; Sidey-Gibbons, C. J. (2021). Machine learning in medicine: a practical introduction to natural language processing, *BMC Medical Research Methodology*, Vol. 21, No. 1, 158
[53] Raparthi, M.; Gayam, S. R.; Nimmagadda, V. S. P.; Sahu, M. K.; Putha, S.; Pattyam, S. P.; Kondapaka, K. K.; Kasaraneni, B. P.; Thuniki, P.; Kuna, S. S. (2022). AI assisted drug discovery: Emphasizing its role in accelerating precision medicine initiatives and improving treatment outcomes, *Human-Computer Interaction*, Vol. 2, No. 2
[54] Chen, R. J.; Lu, M. Y.; Chen, T. Y.; Williamson, D. F.; Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare, *Nature Biomedical Engineering*, Vol. 5, No. 6, 493–497
[55] The European Health Data Space (EHDS). (n.d.), from https://www.european-health-data-space.com/
[56] Welcome to the All of Us Research Hub. (n.d.), from https://www.researchallofus.org/
[57] All of Us Research Program Overview. (n.d.), from https://allofus.nih.gov/article/program-overview
[58] UK Biobank. (n.d.), from https://www.ukbiobank.ac.uk/discoveries-and-impact/major-achievements/