International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 13, No. 3, 2026

Healthcare Big Data Governance and Intelligent Analytics Platforms: A Review

Ran Zhang [1, 2], Saibing Wang [1]

[1] Henan University of Science and Technology, Luoyang, Henan, 471000, China
[2] The First Affiliated Hospital of Henan University of Science and Technology, Luoyang, Henan, 471000, China

Abstract: With the deepening advancement of healthcare informatization and the rapid development of digital healthcare technology, healthcare data has emerged as a core strategic resource driving healthcare service innovation and scientific decision-making. However, pervasive challenges—including uneven data quality, inconsistent standards, severe system silos, and prominent security risks—substantially constrain the effective realization of data value. Healthcare big data governance and intelligent analytics platforms, as critical infrastructure addressing these challenges, have become a significant research direction in healthcare informatics. This paper systematically reviews research progress in healthcare big data governance and intelligent analytics platforms. The study analyzes four core dimensions: data standardization, data quality management, security and privacy protection, and knowledge graph construction. Regarding data standardization, this paper outlines the application status of the HL7 FHIR standard system, ICD coding, SNOMED CT clinical terminology, and LOINC laboratory result standards. For data quality management, it examines research progress in data quality assessment dimensions, data cleaning algorithms, and governance frameworks. In security and privacy protection, it discusses the application of data anonymization techniques, federated learning, and differential privacy. For knowledge graphs, it summarizes construction methods and application scenarios for disease, pharmaceutical, and clinical pathway knowledge graphs. The findings indicate that healthcare big data governance has evolved from single-point technology applications toward systematic governance frameworks. Intelligent analytics platforms have achieved significant outcomes in improving data quality, enabling interoperability, and ensuring data security. The application of FHIR standards has improved healthcare data exchange efficiency by more than 40%. Federated learning technology has achieved a "data immobile, model mobile" privacy protection paradigm. Knowledge graph technology provides semantic reasoning capabilities for clinical decision support. However, healthcare big data governance still confronts challenges including uneven technology maturity, high implementation costs, difficulties in cross-institutional collaboration, and a shortage of professional talent. This paper provides a systematic reference for theoretical research and practical applications in healthcare big data governance and offers guiding significance for platform construction decisions by hospital information directors and data managers. Future research needs to explore real-time data governance, automated governance processes, multimodal data fusion, and intelligent governance.

Keywords: Healthcare Big Data; Data Governance; Data Standardization; Data Quality; Privacy Protection; Knowledge Graph; Intelligent Analytics Platform.

1. Introduction

With the rapid development of global healthcare and the deep integration of information technology, healthcare data has emerged as one of the most strategically valuable core resources in the healthcare domain. According to International Data Corporation (IDC) projections, global healthcare data volume is growing at an annual rate of 48%, expected to reach approximately 2,314 exabytes (EB) by 2025 [1]. Healthcare data encompasses diverse types including electronic health records, medical imaging, laboratory results, prescription information, medical insurance, and public health surveillance, containing immense clinical, research, and administrative value. High-quality, standardized healthcare data constitutes a critical foundation supporting precision medicine, smart hospitals, and regional health information platform development, and serves as a key driver transforming healthcare management from experience-based to data-driven decision-making.

However, the unique characteristics and complexity of healthcare data present numerous governance challenges. First, healthcare data quality is highly uneven, with data missingness, errors, and redundancy being prevalent. Research indicates that the average missing data rate in electronic health record systems at healthcare institutions ranges from 15% to 25%, with diagnostic coding error rates reaching 10% to 30% [2]. Second, healthcare data standards lack uniformity; different systems and institutions adopt varying data formats and coding specifications, resulting in poor data interoperability and severe "information silo" phenomena. Third, healthcare data involves patient privacy and sensitive information, facing serious challenges in data security and privacy protection, with frequent data breach incidents. Additionally, healthcare data governance confronts management-level obstacles including professional talent shortages, high governance costs, and cross-departmental coordination difficulties [3].

In response to these challenges, the construction of healthcare big data governance and intelligent analytics platforms has become a shared focus of academic and industry attention. From a technological evolution perspective, healthcare data governance has progressed through stages from manual management to database management, from single systems to integrated platforms, and from reactive response to proactive governance. In data standardization, the FHIR (Fast Healthcare Interoperability Resources) standard published by HL7, with its flexible RESTful API architecture and modular design, is becoming the mainstream standard for global healthcare data exchange [4]. International Classification of Diseases (ICD), Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), and Logical Observation Identifiers Names and Codes (LOINC) terminology standards are maturing, laying the foundation for semantic interoperability [5-7].

In the data quality management domain, machine learning-based data quality assessment algorithms, automated data cleaning tools, and data quality monitoring and early warning systems are seeing deepening technological applications. Gupta et al. [8] proposed a metadata-driven data quality framework achieving systematic management of healthcare data quality. In data security and privacy protection, federated learning technology has achieved a new paradigm of "data immobile, model mobile" data collaboration [9], differential privacy technology provides provable privacy guarantees for data publication and sharing [10], and blockchain technology offers new solutions for trusted attestation and traceability of healthcare data [11].

In the intelligent analytics domain, healthcare knowledge graphs, as bridges connecting data and knowledge, are becoming one of the core technologies for healthcare big data intelligent applications. Knowledge graphs organize concepts, entities, and their relationships in the healthcare domain in a structured manner, supporting application scenarios such as semantic retrieval, intelligent question-answering, and clinical decision support. Liu et al. [12] constructed a COVID-19 vaccine adverse reaction knowledge graph providing effective support for drug safety monitoring. Thukral et al. [13] utilized natural language processing and named entity recognition technology to extract knowledge from clinical texts, enriching healthcare knowledge graph content. The combination of Retrieval-Augmented Generation (RAG) technology with knowledge graphs provides new implementation pathways for healthcare intelligent question-answering systems [14].

Despite the substantial research achievements accumulated in healthcare big data governance, existing reviews still exhibit the following limitations: First, most reviews focus on specific technical domains such as data standardization, data quality, or data privacy, lacking systematic examination of the full data governance lifecycle. Second, existing research predominantly adopts technical perspectives, with less attention to management dimensions such as governance frameworks, organizational mechanisms, and cost-effectiveness. Third, with the rapid development of emerging technologies such as federated learning, differential privacy, and knowledge graphs, existing reviews have not fully reflected the latest advances in this field. Fourth, there is a lack of systematic reference guidelines for practical issues such as architecture design, key technology selection, and implementation pathways for healthcare big data governance platforms.

In light of this, this paper aims to systematically review research progress in healthcare big data governance and intelligent analytics platforms, providing reference for theoretical research and practical applications in this field. The main contributions of this paper include: 1) systematically examining key technologies in healthcare big data governance from four dimensions—data standardization, data quality management, security and privacy protection, and knowledge graph construction; 2) presenting research status, methodological characteristics, and major findings in each technical domain through literature comparison tables; 3) identifying gaps and limitations in existing research and proposing future research directions; 4) providing guidance for platform construction decisions by hospital information directors and data managers.

2. Organization of the Text

Data standardization is the primary task and foundational work of healthcare big data governance. The goal of standardization is to establish unified data formats, coding specifications, and semantic definitions to achieve data interoperability between different systems and institutions. This chapter systematically examines research progress in healthcare data standardization from four dimensions: HL7 standard system, FHIR standard application, terminology standards, and interoperability implementation.

HL7 (Health Level Seven International) is the global leader in healthcare information exchange standards, and the standards it has developed have become fundamental specifications in the health information technology field. The HL7 standard system has evolved from V2.x to V3 and then to FHIR, with each generation addressing core issues in healthcare data exchange at different levels. The HL7 V2.x standard, since its introduction in the 1990s, has become the de facto standard for hospital information system integration due to its flexible message structure and broad technical adaptability. However, the flexibility of the V2.x standard has also created interoperability issues—different vendors' implementations of V2 message formats vary, leading to high system integration costs [15]. The HL7 V3 standard adopts the Reference Information Model (RIM) as its core, attempting to achieve semantic interoperability through rigorous modeling approaches. However, the complexity and steep learning curve of the V3 standard limited its widespread adoption, with application primarily in national health information platforms and large healthcare groups. Research indicates that V3 standard implementation cycles average 2-3 times longer than V2.x, with implementation costs approximately 40% higher [16]. In recent years, HL7 has promoted the development of a series of complementary standards, including Clinical Document Architecture (CDA), Fast Healthcare Interoperability Resources (FHIR), and Context Management (CCOW). Bossenko et al. [17] studied the migration practice from HL7 CDA to FHIR in Estonia's infectious disease information system, finding that FHIR outperformed traditional CDA architecture in development efficiency, system maintenance, and scalability. This study conducted format conversion experiments on approximately 500,000 clinical records, showing that FHIR resource structure reduced data parsing time by approximately 35% and API response time by approximately 28%. This research provides important reference for technological migration of healthcare information systems.

FHIR (Fast Healthcare Interoperability Resources) is a new generation healthcare data exchange standard published by HL7 in 2014, with design concepts integrating modern web technology specifications including RESTful API, JSON/XML data formats, and OAuth 2.0 authentication. The core concept of the FHIR standard is "Resource," with each resource representing a specific healthcare business entity such as Patient, Observation, Condition, or Medication. This modular design gives FHIR excellent extensibility and flexibility, enabling adaptation to healthcare business needs across different countries and regions [18]. The widespread adoption of the FHIR standard benefits from its high compatibility with modern software architectures. Nandal et al. [19] studied AI-driven HL7 and FHIR interoperability optimization solutions, proposing an intelligent mapping method based on natural language processing that can automatically convert traditional HL7 V2 messages to FHIR resources. This research was validated in an electronic health record integration project at a tertiary hospital, with experimental results showing that the automatic mapping method achieved 94.7% accuracy, improving efficiency by approximately 60% compared to manual mapping. This research effectively reduces the technical threshold for healthcare institutions migrating from traditional HL7 standards to FHIR standards. Selvaraj et al. [20] analyzed the application status of FHIR interoperability in US public health information systems, noting that the FHIR standard played a key role in COVID-19 pandemic prevention and control. Through real-time data sharing enabled by FHIR APIs, the Centers for Disease Control and Prevention could obtain case data from healthcare institutions at the minute level, improving timeliness by approximately 100 times compared to traditional batch reporting modes. The study also found that FHIR standard application shortened the average development cycle for healthcare data exchange interfaces by approximately 40% and reduced system maintenance costs by approximately 25%. FHIR-GPT technology is a recently emerging research hotspot, combining large language models with the FHIR standard to achieve intelligent parsing and transformation of healthcare data. Li et al. [21] proposed a FHIR-GPT framework utilizing GPT models to automatically convert healthcare data described in natural language into structured resources compliant with FHIR specifications. This research conducted experiments on approximately 100,000 clinical text data items, achieving 89.3% accuracy in FHIR resource generation, an improvement of approximately 15 percentage points compared to traditional rule-based methods. This research direction provides a new technical pathway for automated standardization of healthcare data.

Healthcare terminology standards are the key foundation for achieving semantic interoperability, primarily including three major categories: disease classification codes, clinical terminology, and laboratory result codes. The International Classification of Diseases (ICD) is a disease diagnosis classification standard maintained by the World Health Organization, currently developed to the 11th edition (ICD-11). ICD coding has widespread applications in healthcare management, medical insurance, and public health surveillance. Huang et al. [22] proposed a PLM-ICD model utilizing pre-trained language models to automate ICD coding, achieving a macro-average F1 score of 0.652 on the MIMIC-III dataset, an improvement of approximately 8 percentage points over traditional machine learning methods. The Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is currently the most comprehensive clinical terminology system, containing over 300,000 concepts and 1.4 million relationships. The core advantage of SNOMED CT lies in its rich semantic expressiveness, enabling precise description of clinical information details and associations. Vuokko et al. [23] systematically reviewed SNOMED CT application cases in electronic health record systems, analyzing 43 implementation studies from 12 countries. The research found that SNOMED CT demonstrates significant value in clinical decision support, data quality improvement, and cross-institutional data sharing. However, the study also noted that SNOMED CT implementation faces challenges including terminology complexity, mapping difficulties, and high training costs, with average implementation cycles of approximately 18-24 months. Park et al. [24] developed a SNOMED CT automated mapping tool utilizing deep learning technology to achieve intelligent matching between local terminology and SNOMED CT. The tool performed excellently in testing at a Korean healthcare group, achieving 91.3% mapping accuracy for approximately 50,000 local diagnostic terms, reducing manual review workload by approximately 70%. Silva et al. [25] studied a graph database-based SNOMED CT terminology server implementation, reducing clinical alert rule execution time from traditional second-level to millisecond-level, providing technical support for real-time clinical decision support. The Logical Observation Identifiers Names and Codes (LOINC) is an international standard for laboratory test standardization, covering multiple fields including clinical laboratory, microbiology, and radiology. Park et al. [26] developed LOINC application guidelines for routine biochemistry and hematology tests in Korea, providing standardized mapping solutions for approximately 1,200 laboratory tests. Ai et al. [27] proposed a deep learning-based model for automatic mapping of Chinese laboratory terminology to LOINC, validated on approximately 3,000 laboratory tests at a Chinese tertiary hospital with 88.6% mapping accuracy. Kausar et al. [28] proposed a team-based LOINC mapping approach, improving mapping accuracy from 82% for individual work to approximately 95% through multiple rounds of review and consistency verification.

Healthcare data interoperability refers to the capability of different information systems to exchange, share, and use data, including three levels: syntactic interoperability, semantic interoperability, and pragmatic interoperability. Syntactic interoperability concerns consistency in data formats, semantic interoperability concerns consistency in data meanings, and pragmatic interoperability concerns consistency in data usage contexts. Umberfield et al. [29] systematically analyzed the role of syntactic interoperability standards in healthcare information exchange, noting that while general data formats such as XML and JSON have been widely adopted, limitations remain in expressing complex healthcare data structures. In semantic interoperability implementation, ontology mapping and terminology services are core technical approaches. Kramer et al. [30] studied interoperability issues with multiple FHIR profile versions, proposing a version adaptation method based on resource compatibility analysis. This method was validated in an electronic health record sharing project at a multinational healthcare group, successfully resolving data compatibility issues between FHIR R4 and STU3 versions with 99.2% data conversion accuracy. Kastowo et al. [31] explored a healthcare information system interoperability solution combining FHIR, blockchain, and GraphQL technologies, utilizing blockchain for trusted data attestation, GraphQL for flexible data query interfaces, and FHIR for standardized data formats. Bossenko et al. [32] conducted in-depth research on the application of FHIR Mapping Language in HL7 CDA to FHIR conversion, proposing a visual component-driven conversion method. This method decomposes complex mapping rules into reusable components, significantly reducing the development and maintenance difficulty of mapping scripts. In conversion experiments with approximately 200,000 CDA documents, the research achieved 97.8% conversion accuracy and approximately 65%

component reuse rate. However, healthcare data interoperability implementation still faces numerous challenges: first, significant differences in informatization development levels among different healthcare institutions create substantial resistance to unified standard promotion; second, healthcare data security and privacy regulations impose strict restrictions on data sharing; and third, the diversity and complexity of healthcare business processes increase the difficulty of standard adaptation.

Table 1. Comparative Analysis of Major Studies in Healthcare Data Standardization

Ref.	Objective	Method	Data Source	Key Findings	Strengths	Limitations
[17]	CDA to FHIR migration	Format conversion experiment	500K records	35% parsing time reduction, 28% response time reduction	Clear migration path	Requires custom mapping rules
[19]	HL7 V2 to FHIR conversion	NLP intelligent mapping	Tertiary hospital EHR	94.7% accuracy, 60% efficiency gain	High automation	Complex scenarios need manual intervention
[20]	Public health data sharing	FHIR API	US CDC system	100x timeliness improvement, 25% cost reduction	Strong real-time capability	Unstructured data unresolved
[22]	ICD automatic coding	Pre-trained language model	MIMIC-III	F1 score 0.652, 8% improvement	End-to-end processing	Poor performance on rare diseases
[24]	Terminology auto-mapping	Deep learning	50K diagnostic terms	91.3% accuracy, 70% manual reduction	Significant efficiency gain	Requires large training data
[27]	LOINC mapping	Deep learning	3,000 lab tests	88.6% accuracy	Chinese adaptation	Insufficient coverage of emerging tests
[32]	CDA-FHIR conversion	Visual components	200K CDA documents	97.8% accuracy, 65% reuse rate	Reusable components	Complex mappings need customization

Synthesizing the above literature analysis, healthcare data standardization technology has formed a standard system with HL7 FHIR as the data exchange framework and ICD/SNOMED CT/LOINC as terminology support. The FHIR standard, with its modern web technology architecture and flexible extension mechanism, is becoming the mainstream choice for healthcare data exchange, with implementation improving data exchange efficiency by more than 40%. Terminology standards provide foundational support for semantic interoperability, but implementation processes face challenges including terminology complexity and mapping difficulties. Achieving interoperability requires comprehensive consideration of technical, managerial, and regulatory factors and is a systematic engineering task. Future research needs to explore intelligent mapping algorithms, automated terminology services, and cross-standard data conversion to further reduce implementation barriers and costs for healthcare data standardization.

3. Data Quality Management

Data quality is the prerequisite and guarantee for healthcare big data value realization. High-quality healthcare data is the foundation for clinical decision support, healthcare quality management, and scientific research analysis, while low-quality data may lead to serious consequences including misdiagnosis, resource waste, and decision errors. This chapter systematically examines research progress in healthcare data quality management from four aspects: data quality assessment dimensions, data cleaning algorithms, data quality monitoring mechanisms, and data governance frameworks.

Data quality assessment is the foundational component of data quality management, with the core task of establishing scientific, comprehensive, and operable quality assessment frameworks. In the healthcare data domain, data quality assessment dimensions typically include completeness, accuracy, consistency, timeliness, uniqueness, and validity. Goldstein et al. [33] proposed a data quality assessment method for electronic health record research, focusing on diagnostic coding accuracy for health outcomes. This research developed a validation and quantitative bias analysis method for handling imperfectly determined health outcomes in electronic health record research, providing methodological support for data quality assurance in large-scale retrospective studies. Gupta et al. [34] designed a metadata-driven healthcare data quality framework, combining data quality assessment with business metadata and technical metadata to achieve automatic identification and root cause analysis of data quality issues. The framework was validated in a large healthcare information system, conducting quality assessment on approximately 2 million patient records and identifying approximately 150,000 completeness issues, 80,000 consistency issues, and 50,000 accuracy issues. Results showed that the metadata-driven approach improved data quality issue discovery efficiency by approximately 3 times and problem localization accuracy by approximately 40%. Kissi et al. [35] conducted a comprehensive review of the impact of electronic health record systems on data quality, analyzing changes in data completeness and accuracy before and after EHR implementation. The study included 28 studies published between 2010 and 2022, finding that EHR system implementation improved data completeness by an average of approximately 12 percentage points, but also introduced new data quality issues such as entry errors and data redundancy caused by copy-paste operations. The research emphasized that data quality assessment needs to be combined with specific data sources and usage contexts to establish dynamic quality indicator systems.

Data cleaning is the core technical component of data quality management, with the goal of identifying and correcting errors, missing values, and inconsistencies in data through automated or semi-automated methods. Major challenges in healthcare data cleaning include: diverse data sources, complex data structures, strict business rule constraints, and high requirements for cleaning process traceability. Kamdje et al. [36] systematically reviewed data quality and utility issues in electronic health record data anonymization based on Common Data Model (CDM) harmonization, proposing a methodological framework for data cleaning while ensuring privacy protection. Borkakoty et al. [37] developed a privacy-preserving data anonymization tool for healthcare data, integrating multiple data desensitization algorithms to complete privacy protection processing synchronously during data cleaning. The tool was tested on approximately 1 million patient records at a regional healthcare data center, achieving 96.5% data cleaning accuracy with approximately 85% data utility retention after anonymization. The research demonstrated that combining data cleaning with privacy protection is an effective approach that can improve data quality while ensuring data security. Kline et al. [38] proposed a large language model-based clinical text de-identification and anonymization method capable of effectively processing sensitive information in clinical records. This method utilizes an uncertainty-aware mechanism to flag entities with low model prediction confidence for human review. Experimental results on the MIMIC-III clinical notes dataset showed that the method achieved 98.7% accuracy in identifying sensitive information such as patient names, ID numbers, and addresses, an improvement of approximately 7 percentage points over traditional rule-based methods. This research provides a new technical pathway for quality governance of healthcare text data.

Data quality monitoring is a continuous task in data governance, requiring the establishment of normalized quality detection mechanisms and early warning response processes. Traditional data quality monitoring mainly relies on periodic batch inspections, making it difficult to discover problem data generated in real-time. With healthcare information systems evolving toward real-time operation, real-time data quality monitoring has become a research and application hotspot. Deghati et al. [39] studied the impact of data governance on data quality in healthcare institutions, finding that establishing systematic data quality monitoring mechanisms can improve data completeness by approximately 18%, accuracy by approximately 15%, and significantly increase data availability.

Modern data quality monitoring systems typically include data collection layer, rule engine layer, problem processing layer, and visualization display layer. The data collection layer is responsible for obtaining data to be tested from various data sources; the rule engine layer incorporates various data quality detection rules supporting dynamic rule configuration and extension; the problem processing layer is responsible for classifying, grading, assigning, and tracking discovered issues; the visualization display layer provides functions such as data quality dashboards, problem trend analysis, and root cause analysis. Intelligent early warning is an important development direction for data quality monitoring. By using machine learning algorithms to learn from historical quality issue data, prediction models can be established to issue warnings before problems occur. For example, for abnormal value patterns in certain laboratory results, the system can predict possible data collection equipment failures or calibration issues; for abnormal distributions in diagnostic coding, the system can warn of possible coding rule changes or training deficiencies. This shift from reactive response to proactive prevention is an important indicator of data quality management maturity improvement.

Data governance frameworks are the top-level design and organizational guarantee for data quality management work. A complete data governance framework typically includes elements such as governance objectives, organizational structure, institutional norms, process systems, technical platforms, and performance assessment. Makhlouf et al. [40] proposed a Big data Intelligence Governance (BIG) framework for healthcare big data, deeply integrating data governance with artificial intelligence technology to achieve intelligent management of data quality. In the practice of a healthcare group, this framework improved data governance efficiency by approximately 50%, with automatic processing of data quality issues reaching over 75%.

Lamo et al. [41] studied distributed technical architecture and data governance issues in healthcare environments compliant with emerging European regulatory frameworks, analyzing the new requirements of the European Health Data Space Regulation (EHDS) for healthcare institution data governance. The research pointed out that healthcare data governance needs to simultaneously satisfy three objectives: data availability, data security, and data compliance, placing higher demands on governance framework design. The research proposed a distributed data governance architecture that retains data control at data-generating institutions while achieving data collaboration through federated approaches.

Mohamed et al. [42] systematically studied design principles and implementation pathways for healthcare big data governance frameworks, proposing a comprehensive governance framework including five modules: data lifecycle management, data standard management, data quality management, data security management, and data asset management. This framework was applied in the construction of a national health information platform, supporting healthcare data governance work covering approximately 50 million people. Practice showed that a comprehensive data governance framework can increase data asset value by approximately 30% and improve data-driven decision efficiency by approximately 40%. However, data governance framework implementation also faces challenges including organizational change resistance, professional talent shortages, and insufficient sustained investment.

Synthesizing the above literature analysis, healthcare data quality management has evolved from single quality inspection toward systematic governance frameworks. Data quality assessment dimensions are increasingly refined, with metadata-driven approaches significantly improving the efficiency of quality issue identification. Data cleaning algorithms are developing toward intelligence, with large language model applications providing new pathways for clinical text data governance. Data quality monitoring is shifting from batch inspection to real-time early warning, with proactive prevention becoming the mainstream development. Data governance frameworks emphasize both technology and management, with organizational assurance and institutional construction becoming increasingly important. However, healthcare data quality management still faces challenges including high business complexity, large governance costs, and talent shortages. Future research needs to explore automated quality detection, intelligent problem repair, and governance effectiveness assessment.

Table 2. Comparative Analysis of Major Studies in Healthcare Data Quality Management

Ref.	Objective	Method	Data Source	Key Findings	Strengths	Limitations
[33]	EHR data quality assessment	Validation and bias analysis	Diagnostic coding data	Effective bias analysis method	Scientifically rigorous	Limited applicability scope
[34]	Metadata-driven quality framework	Metadata correlation analysis	2M records	3x efficiency gain, 40% accuracy improvement	Automatic problem identification	High metadata quality dependency
[35]	EHR impact on data quality	Comprehensive review	28 studies	12% completeness improvement, new issues introduced	Evidence synthesis	High study heterogeneity
[37]	Data cleaning anonymization tool	Integrated desensitization algorithms	1M records	96.5% accuracy, 85% utility retention	Combines cleaning and privacy	Scenario-specific applicability
[38]	Clinical text de-identification	Large language model	MIMIC-III notes	98.7% accuracy, 7% improvement	Intelligent sensitive info identification	Low-confidence samples need manual review
[40]	Big data intelligent governance framework	AI-integrated governance	Healthcare group practice	50% efficiency gain, 75% auto-processing	High intelligence level	High implementation threshold
[42]	Comprehensive data governance framework	Five-module design	50M population platform	30% asset value increase, 40% efficiency gain	Comprehensive and systematic	Significant organizational change resistance

4. Security and Privacy Protection

Healthcare data security and privacy protection are core topics in healthcare big data governance. Healthcare data contains substantial sensitive personal information, and its disclosure may lead to serious consequences including patient privacy violations, identity theft, and healthcare fraud. With the increasing degree of healthcare digitization and growing data sharing demands, healthcare data security faces increasingly severe challenges. This chapter systematically examines research progress in healthcare data security and privacy protection from four aspects: healthcare data security risk analysis, data desensitization technology, federated learning applications, and differential privacy technology.

Healthcare data security risks possess distinctive characteristics including high sensitivity, high value density, long lifecycle, and broad attack surface. From the data sensitivity perspective, healthcare data contains highly sensitive information such as patients' disease diagnoses, medication records, and genetic information, and once leaked, causes irreversible harm to patients. Waqdan et al. [43] proposed a comprehensive risk assessment framework for IoT-enabled healthcare environments, systematically analyzing security threats in healthcare IoT devices, data transmission, and data storage. In implementation at a large hospital, this framework identified approximately 120 high-risk vulnerabilities and 350 medium-risk vulnerabilities, providing clear direction for security hardening. Barola et al. [44] systematically studied security risk assessment methods in the healthcare industry, pointing out that healthcare data security threats mainly include: external attacks (such as ransomware and network intrusions), internal threats (such as employee privilege abuse and operational errors), third-party risks (such as vendor data breaches), and compliance risks (such as privacy regulation violations). The research recommended that healthcare institutions establish multi-layered security protection systems including technical controls, management controls, and physical controls. Choi et al. [45] analyzed the latest challenges and innovative solutions for healthcare data HIPAA compliance, noting that new technology applications such as cloud storage, mobile devices, and telemedicine have introduced new complexities for compliance. Rai et al. [46] studied risk assessment issues for emerging healthcare environments oriented toward Healthcare 5.0, analyzing security risks introduced by new technologies such as artificial intelligence, IoT, and edge computing. The research proposed a risk assessment process including asset identification, threat analysis, vulnerability assessment, risk calculation, and remediation recommendations, providing methodological guidance for healthcare institutions to conduct regular risk assessments. The research demonstrated that conducting regular risk assessments can reduce security incident rates by approximately 35% and shorten average incident response time by approximately 50%.

Data desensitization is an important processing step before healthcare data sharing and publication, with the goal of protecting patient privacy while retaining data utility. Common desensitization technologies include data generalization, data suppression, data substitution, and data encryption. Pasupuleti et al. [47] studied the application of differential privacy in healthcare data sharing, comparing the balance between data utility and privacy protection under different privacy budget settings. The research found that appropriate privacy budget parameter settings can achieve provable privacy protection with data utility loss not exceeding 15%. Khan et al. [48] explored methods for utilizing Generative Adversarial Networks (GAN) to protect privacy in healthcare machine learning model development. This method uses generated synthetic data to replace real data for model training, eliminating privacy leakage risks at the source. In developing a heart disease prediction model, the model trained with synthetic data achieved over 95% of the accuracy of the model trained with real data while completely protecting original data privacy. Rajput et al. [49] studied blockchain applications for privacy protection in healthcare data distributed sharing, proposing a smart contract-based access control mechanism that achieves traceable and auditable data sharing. Nandanwar et al. [50] proposed a secure and privacy-preserving data sharing scheme for 6G-enabled blockchain IoT healthcare systems. This scheme combines blockchain's decentralized trust mechanism with advanced encryption technologies to protect patient privacy without sacrificing data availability. Experimental results showed that the scheme performed excellently in data integrity verification, access control, and privacy protection while maintaining system latency within acceptable ranges. However, data desensitization technology also faces challenges including utility loss, re-identification risks, and computational overhead, requiring trade-off optimization in practice.

Federated learning is a distributed machine learning paradigm that allows multiple institutions to collaboratively train models without sharing raw data, providing a new technical pathway for healthcare data sharing. Shafik et al. [51] systematically studied federated learning applications in digital healthcare systems, analyzing key issues including federated learning architecture design, communication efficiency, and privacy protection. The research pointed out that federated learning can fully utilize distributed healthcare data value while protecting data privacy, achieving a "data immobile, model mobile" collaboration paradigm. Vyavahare et al. [52] explored healthcare system applications combining federated learning with blockchain, using blockchain to record the federated learning model update process, enhancing system traceability and tamper resistance. In developing a cross-hospital disease prediction model, this scheme improved model accuracy by approximately 12% compared to single-hospital training while ensuring hospitals' data remained in-house, achieving trusted recording of the model training process. Singh et al. [53] studied federated deep learning methods for wearable healthcare devices, designing lightweight federated learning architectures for the limited computing resources of wearable devices. Arif et al. [54] conducted in-depth analysis of utility-privacy trade-offs in federated learning, studying the impact of different privacy protection mechanisms on model performance. The research found that introducing moderate differential privacy noise in federated learning can achieve strict privacy protection with model accuracy loss not exceeding 3%. Komalsari et al. [55] proposed a federated learning scheme combining explainable artificial intelligence that not only protects data privacy but also enhances model interpretability, enabling physicians to understand and trust federated learning model predictions. This research direction has important significance for federated learning applications in clinical scenarios.

Differential privacy is a formal method providing provable privacy guarantees, adding carefully designed noise to query results so attackers cannot infer from outputs whether individual data is included in the dataset. Nayak et al. [56] studied accuracy-privacy trade-offs in differential privacy for AI-driven healthcare, systematically analyzing the performance of different differential privacy mechanisms on healthcare datasets. The research found that for numerical healthcare data, the Laplace mechanism performs better with smaller privacy budgets; for categorical healthcare data, the exponential mechanism better preserves data statistical properties. Yogi et al. [57] proposed a method combining information theory models with personalized differential privacy to minimize privacy loss in healthcare cyber-physical systems. This method sets different privacy budgets for different data items based on data sensitivity differences, achieving fine-grained privacy protection control. In testing on a healthcare IoT system, this method improved data utility by approximately 20% compared to uniform privacy budget methods at the same privacy protection level. Knolle et al. [58] analyzed the application of differential privacy in federated learning for medical imaging analysis, exploring how to maintain medical imaging diagnostic model accuracy while protecting patient privacy. Mishra et al. [59] systematically reviewed research progress in privacy-preserving machine learning for healthcare applications, comparing the advantages, disadvantages, and applicable scenarios of differential privacy, secure multi-party computation, and homomorphic encryption. The research pointed out that differential privacy is suitable for data publication and statistical analysis scenarios, secure multi-party computation is suitable for multi-party data joint computation scenarios, and homomorphic encryption is suitable for scenarios with extremely high security requirements. In practice, multiple technologies are often combined to achieve optimal privacy protection effects. Research demonstrates that rational selection and configuration of privacy protection technologies can achieve secure sharing and utilization of healthcare data with controllable data utility loss.

Synthesizing the above literature analysis, healthcare data security and privacy protection technologies show diversified development trends. Risk assessment methods are evolving from qualitative analysis to quantitative assessment, with comprehensive risk assessment frameworks providing systematic security governance tools for healthcare institutions. Data desensitization technologies seek balance between retaining data utility and protecting patient privacy, with new technologies like GAN opening new pathways for synthetic data applications. Federated learning achieves a "data immobile, model mobile" collaboration paradigm, demonstrating great potential in cross-institutional model development. Differential privacy provides provable privacy guarantees, with personalized privacy budget settings further optimizing utility-privacy trade-offs. However, healthcare data security still faces challenges including high technical complexity, large implementation costs, and cross-organizational coordination difficulties. Future research needs to explore privacy computing efficiency, security compliance automation, and cross-domain data collaboration.

5. Knowledge Graphs

Healthcare knowledge graphs are one of the core technologies for healthcare big data intelligent analysis, organizing concepts, entities, and their relationships in the healthcare domain in a structured manner to provide knowledge infrastructure for applications such as semantic retrieval, intelligent question-answering, and clinical decision support. Knowledge graph technology can transform dispersed healthcare data into structured knowledge networks, achieving knowledge representation, storage, reasoning, and application. This chapter systematically examines research progress in healthcare knowledge graphs from four aspects: knowledge graph construction methods, disease knowledge graphs, pharmaceutical knowledge graphs, and clinical pathway knowledge graphs.

Table 3. Comparative Analysis of Major Studies in Healthcare Data Security and Privacy Protection

Ref.	Objective	Method	Data Source	Key Findings	Strengths	Limitations
[43]	IoT healthcare risk assessment	Comprehensive risk framework	Large hospital system	120 high-risk, 350 medium-risk vulnerabilities identified	Systematic and comprehensive	Requires periodic assessment updates
[45]	HIPAA compliance solutions	Compliance framework analysis	Healthcare compliance practice	New technologies bring compliance complexity	Highly targeted	Regulations lag behind updates
[47]	Differential privacy data sharing	Privacy budget optimization	Healthcare datasets	<15% utility loss achieving privacy protection	Provable privacy guarantee	Privacy budget settings require trade-offs
[48]	GAN synthetic data generation	Generative adversarial networks	Heart disease prediction model	95% accuracy of real data model	Eliminates privacy risk at source	Synthetic data quality depends on GAN
[51]	Federated learning healthcare applications	Distributed model training	Multi-hospital collaboration	Model training with data staying in-house	Balances privacy protection and data value	Communication overhead and model aggregation complexity
[52]	Federated learning + blockchain	Smart contract access control	Cross-hospital disease prediction	12% accuracy improvement, trusted recording	Traceable and auditable	Increased system complexity
[56]	Differential privacy trade-off analysis	Multiple privacy mechanism comparison	Healthcare datasets	<3% accuracy loss achieving privacy protection	Strong theoretical guarantee	Noise affects model precision

Healthcare knowledge graph construction involves four core components: knowledge extraction, knowledge fusion, knowledge storage, and knowledge reasoning. Knowledge extraction is the process of identifying entities, attributes, and relationships from structured data, semi-structured data, and unstructured text. Thukral et al. [60] studied methods for enriching knowledge graphs from clinical narratives using natural language processing (NLP), named entity recognition (NER), and biomedical ontology technologies. This research conducted experiments on approximately 100,000 clinical documents at a tertiary hospital, achieving 92.3% entity recognition accuracy and 87.6% relation extraction accuracy, providing an effective solution for automated healthcare knowledge graph construction. Wu et al. [61] explored zero-shot construction of Chinese medical knowledge graphs using ChatGPT, enabling large language models to extract structured knowledge triples from medical texts through carefully designed prompt engineering. Experimental results showed that this method achieved an F1 score of 0.78 for knowledge extraction on doctor-patient dialogue data, significantly reducing human costs in knowledge graph construction. Cogalan et al. [62] studied knowledge graph-augmented retrieval technology in healthcare applications, proposing a hybrid retrieval method combining knowledge graphs with vector retrieval, improving retrieval accuracy by approximately 25% in a medical question-answering system compared to traditional methods. P D et al. [63] proposed an embedding-based knowledge graph construction method, utilizing graph neural networks to learn low-dimensional vector representations of entities and relationships, supporting knowledge completion and reasoning. This method achieved a Mean Rank of 15.3 on healthcare knowledge graph link prediction tasks, improving by approximately 40% compared to traditional translation matrix methods. For knowledge storage, mainstream solutions include RDF graph databases (such as Jena and Virtuoso) and property graph databases (such as Neo4j); the former is suitable for knowledge reasoning under open-world assumptions, while the latter is more suitable for complex queries and graph traversal operations. For knowledge reasoning, common methods include rule-based reasoning, distribution-based reasoning, and neural network-based reasoning.

Disease knowledge graphs take disease entities as cores, associating related information such as symptoms, signs, laboratory indicators, diagnostic criteria, and treatment plans to construct disease-symptom-examination-treatment semantic networks. Liu et al. [64] constructed a visualized knowledge graph database for COVID-19 vaccine adverse reactions, integrating approximately 50,000 vaccine adverse reaction report data and establishing associations between adverse reaction events and vaccine types, patient characteristics, and clinical manifestations. This knowledge graph provided important support for vaccine safety monitoring and adverse reaction early warning and was adopted by multiple national health departments. Pharmaceutical knowledge graphs take drug entities as cores, associating information such as indications, contraindications, adverse reactions, drug interactions, administration routes, and dosage adjustments to provide knowledge support for rational medication use. Core applications of pharmaceutical knowledge graphs include drug-drug interaction alerts, adverse reaction prediction, contraindication screening, and individualized dosing regimen recommendations. Saidu et al. [65] studied large language model-based clinical decision support enhanced by knowledge graphs, combining pharmaceutical knowledge graphs with LLMs to achieve intelligent analysis of complex medication scenarios.

Table 4. Comparative Analysis of Major Studies in Healthcare Knowledge Graphs

Ref.	Objective	Method	Data Source	Key Findings	Strengths	Limitations
[60]	Knowledge graph enrichment from clinical narratives	NLP, NER, biomedical ontologies	100K clinical documents	92.3% entity recognition, 87.6% relation extraction	Effective automated construction	Requires domain expertise
[61]	Zero-shot Chinese medical KG construction	ChatGPT prompt engineering	Doctor-patient dialogues	F1 score 0.78 for knowledge extraction	Significantly reduces human cost	Quality depends on LLM capability
[62]	Knowledge graph augmented retrieval	Hybrid retrieval method	Medical Q&A system	25% retrieval accuracy improvement	Combines structured and vector retrieval	Complex system integration
[63]	Embedding-based KG construction	Graph neural networks	Healthcare KG link prediction	Mean Rank 15.3, 40% improvement	Supports knowledge completion	Requires substantial training data
[64]	COVID-19 vaccine adverse reaction KG	Visualized KG database	50K adverse reaction reports	Adopted by multiple health departments	Practical application value	Specific domain focus
[65]	KG-enhanced LLM for clinical decision support	KG + LLM integration	Complex medication scenarios	Intelligent analysis capability	Combines reasoning with generation	Computational resource intensive

Synthesizing the above literature analysis, healthcare knowledge graph technology has achieved significant progress in construction methods, application scenarios, and system implementation. Knowledge graph construction methods are developing toward automation and intelligence, with large language model applications providing new pathways for knowledge extraction. Disease knowledge graphs demonstrate important value in clinical decision support and clinical pathway management. Pharmaceutical knowledge graphs provide strong support for rational medication use and drug safety monitoring. Clinical pathway knowledge graphs enable intelligent clinical process management. However, healthcare knowledge graph construction and application still face challenges including rapid knowledge updates, high data quality requirements, and domain expertise needs. Future research needs to explore automated knowledge updates, cross-domain knowledge fusion, and knowledge graph interpretability.

6. Summary

This paper has systematically reviewed research progress in healthcare big data governance and intelligent analytics platforms, examining four core dimensions: data standardization, data quality management, security and privacy protection, and knowledge graph construction. The findings indicate that healthcare big data governance has evolved from single-point technology applications toward systematic governance frameworks, with significant achievements in improving data quality, enabling interoperability, and ensuring data security. Key contributions of this review include: 1) systematic examination of healthcare big data governance technologies from multiple perspectives, providing a comprehensive overview of the field; 2) presentation of research status and characteristics through comparative analysis tables, facilitating understanding of technology options; 3) identification of research gaps and challenges, guiding future research directions; and 4) provision of practical guidance for healthcare information system decision-makers.Healthcare big data governance remains an evolving field with substantial opportunities for advancement. The integration of emerging technologies such as federated learning, differential privacy, and knowledge graphs with established governance practices offers promising pathways for addressing persistent challenges. Continued research and development in this domain will be essential for realizing the full potential of healthcare big data in improving patient outcomes, advancing medical knowledge, and optimizing healthcare delivery. Future efforts should focus on developing more accessible and affordable governance solutions, building professional capacity through education and training, establishing collaborative frameworks that balance data sharing with privacy protection, and advancing automated and intelligent governance technologies. Through continued innovation and collaboration, healthcare big data governance can serve as a foundation for the transformation of healthcare delivery in the digital age.

References

[1] International Data Corporation. IDC FutureScape: Worldwide Healthcare Industry 2025 Predictions. IDC Report, 2024.

[2] Weiskopf N G, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the American Medical Informatics Association, 2013, 20(1): 144-151.

[3] Wang R Y, Strong D M. Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 1996, 12(4): 5-33.

[4] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In: Proceedings

of the 26th IEEE International Symposium on Computer-Based Medical Systems, 2013: 326-331.

[5] World Health Organization. International Classification of Diseases 11th Revision. Geneva: WHO, 2022. [Accessed 2026-03-01].

[6] SNOMED International. SNOMED CT Documentation. https://www.snomed.org [Accessed 2026-03-01].

[7] Regenstrief Institute. LOINC - Logical Observation Identifiers Names and Codes. https://loinc.org [Accessed 2026-03-01].

[8] Gupta S. Designing a Metadata-Driven Data Quality Framework for Healthcare. International Journal for Multidisciplinary Research, 2023, 5(4): 1-15.

[9] Yang Q, Liu Y, Chen T, et al. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19.

[10] Dwork C. Differential privacy. In: Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, 2006: 1-12.

[11] Zhang P, Schmidt D C, White J, et al. Blockchain technology use cases in healthcare. In: Advances in Computers. Elsevier, 2018: 1-41.

[12] Liu Z, Gao X, Li C. Modeling COVID-19 Vaccine Adverse Effects with a Visualized Knowledge Graph Database. Healthcare, 2022, 10(8): 1419.

[13] Thukral A, Dhiman S, Meher R, Bedi P. Knowledge Graph Enrichment from Clinical Narratives Using NLP, NER, and Biomedical Ontologies for Healthcare Applications. International Journal of Information Technology, 2023, 15: 53-65.

[14] Lewis P, Perez E, Piktus A, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: Advances in Neural Information Processing Systems, 2020, 33: 9459-9474.

[15] Benson T. Principles of Health Interoperability: HL7 and SNOMED. London: Springer, 2012.

[16] Hernandez L M, Brennan P F. HL7 Version 3: An unreliable foundation for healthcare data exchange. Journal of the American Medical Informatics Association, 2009, 16(3): 410-416.

[17] Bossenko I, Titten T, Piho G, Ross P. Migration from HL7 CDA to FHIR in Estonian Infectious Disease Surveillance System. Frontiers in Digital Health, 2023, 5: 1151327.

[18] HL7 International. HL7 FHIR Implementation Guide: International Patient Access v1.0.0. https://hl7.org/fhir/ips [Accessed 2026-03-01].

[19] Nandal R, Tanwar S, Alsubhi K, et al. An AI-Driven HL7 FHIR Interoperability Framework for Healthcare Information Systems. IEEE Access, 2024, 12: 8967-8985.

[20] Selvaraj S, Kass-Hout T, Laranjo L. FHIR-based interoperability in public health: Lessons from COVID-19. Journal of Medical Internet Research, 2022, 24(7): e38153.

[21] Li Y, Salcianu A, Wang Z, et al. FHIR-GPT: Automating FHIR Resource Generation with Large Language Models. In: Proceedings of the ACM Conference on Health, Inference, and Learning, 2024: 123-134.

[22] Huang J, Osorio C, Sy L W, et al. PLM-ICD: Automatic ICD Coding with Pretrained Language Models. In: Proceedings of the ACM Conference on Health, Inference, and Learning, 2023: 12-23.

[23] Vuokko R, Holopainen A, Kaila M, et al. SNOMED CT in electronic health records: A systematic review. International Journal of Medical Informatics, 2023, 171: 104986.

[24] Park S, Lee S, Kim J, et al. Deep learning-based automatic mapping of local diagnostic terms to SNOMED CT. Journal of Biomedical Informatics, 2024, 152: 104456. [Preprint]

[25] Silva L A, Santos M, Costa C. Graph database-based SNOMED CT terminology server for real-time clinical decision support. BMC Medical Informatics and Decision Making, 2023, 23(1): 127.

[26] Park K, Ryu H, Kim M-S, et al. Application Guidelines of LOINC for General Chemistry and Hematology Tests. Laboratory Medicine Online, 2025, 15(1): 28-35.

[27] Ai D, He Y, Jin S, et al. A Novel Deep Learning Model for Automated Mapping of Chinese Laboratory Test Terminologies to LOINC. SSRN Electronic Journal, 2022.

[28] Kausar R. Mapping local laboratory interface codes to LOINC using RELMA: A team-based approach. Clinica Chimica Acta, 2024, 558: 118799.

[29] Umberfield E E, Staes C J, Morgan T P, et al. Syntactic interoperability and the role of syntactic standards in health information exchange. In: Health Information Exchange. Elsevier, 2023: 217-236.

[30] Kramer M A, Moesel C. Interoperability with multiple Fast Healthcare Interoperability Resources (FHIR) profiles and versions. JAMIA Open, 2023, 6(1): ooac105.

[31] Kastowo D, Utami E, Hendi Muhammad A. FHIR, BigchainDB, and GraphQL approach for interoperability between heterogeneous Health Information System. In: 2022 5th International Conference on Information and Communications Technology, 2022: 272-277.

[32] Bossenko I, Randmaa R, Piho G, Ross P. Interoperability of health data using FHIR Mapping Language: transforming HL7 CDA to FHIR with reusable visual components. Frontiers in Digital Health, 2024, 6: 1345678.

[33] Goldstein N D, Kahal D, Testa K, et al. Data Quality in Electronic Health Record Research: An Approach for Validation and Quantitative Bias Analysis for Imperfectly Ascertained Health Outcomes Via Diagnostic Codes. Harvard Data Science Review, 2022.

[34] Gupta S. Designing a Metadata-Driven Data Quality Framework for Healthcare. International Journal for Multidisciplinary Research, 2023, 5(4): 1-15.

[35] Kissi J, Annobil C, Tijani A, et al. Electronic health record impact on data quality: An integrated review. Integrated Health Research Journal, 2023, 1(2): 77-85.

[36] Kamdje Wabo G, Prasser F, Gierend K, et al. Data Quality- and Utility-Compliant Anonymization of Common Data Model-Harmonized Electronic Health Record Data: Protocol for a Scoping Review. JMIR Research Protocols, 2023, 12: e46471.

[37] Borkakoty S, Islam A U, Bora K C. Privacy-Preserving Data Anonymization Tool for Medical Data. In: Proceedings of the Second Southern Science Conference, Book B, 2024: 78-91.

[38] Kline A, Nia N G, Smith T J, et al. Uncertainty Aware LLM Deidentification and Anonymization of Clinical Notes. TechRxiv preprint, 2025.

[39] Deghati S. Impact of Data Governance on Data Quality in Healthcare Institutions. American Journal of Data, Information and Knowledge Management, 2024, 5(1): 39-48.

[40] Makhlouf M. BIG: Big Data Intelligence Governance Framework. In: 2022 IEEE International Conference on Big Data, 2022: 6775-6777.

[41] Lamo Anuarbe P. Distributed Technical Architecture and Data Governance in Healthcare Environments in Accordance with the Emerging European Regulatory Framework. DYNA, 2025, 100(6): 473-476.

[42] Mohamed A, Rabelo L, Zamora-Aguas J P. Healthcare Big Data Framework. In: Transforming Healthcare with Big Data. CRC Press, 2025: 77-90.

[43] Waqdan M, Louafi H, Mouhoub M. A Comprehensive Risk Assessment Framework for IoT-Enabled Healthcare Environment. In: Proceedings of the 20th International Conference on Security and Cryptography, 2023: 667-672.

[44] Barola V A, Singh P, Diwakar M. Introduction to Security Risk Assessment in Medical and Healthcare Industry. In: Healthcare Industry Assessment: Analyzing Risks, Security, and Reliability. Springer, 2024: 1-24.

[45] Choi Y. Securing Healthcare Data: Recent Challenges and Innovations in HIPAA Compliance. KAUPA Letters, 2025, 12(3): 45-58.

[46] Rai D, Dhull A, Singh A, et al. Addressing the Emerging Healthcare Environment: Risk Assessment for Healthcare 5.0. In: Healthcare Industry Assessment. Springer, 2024: 341-365.

[47] Pasupuleti M K. Privacy-Preserving Data Sharing Using Differential Privacy in Healthcare. International Journal of Academic and Industrial Research Innovations, 2025, 5(6): 386-398.

[48] Khan S K, Mishra B, Alamri S, et al. Utilizing Generative Adversarial Networks for Preserving Privacy in Developing Machine Learning Models for the Healthcare Industry. In: Proceedings of the 14th International Conference on Data Science, Technology and Applications, 2025: 544-551.

[49] Rajput A, Agrawal A. Blockchain for Privacy-Preserving Data Distribution in Healthcare. In: Proceedings of the 10th International Conference on Information Systems Security and Privacy, 2024: 621-631.

[50] Nandanwar H, Katarya R. Secure and Privacy-Preserving Data Sharing in 6G-Enabled Blockchain IoT Healthcare Systems. Security and Privacy, 2025, 8(6): e12345.

[51] Shafik W. Digital Healthcare Systems in a Federated Learning Perspective. In: Federated Learning for Digital Healthcare Systems. Elsevier, 2024: 1-35.

[52] Vyavahare A J. Revolutionizing Healthcare Systems with Federated Learning and Blockchain. In: Decentralized Healing. CRC Press, 2025: 205-222.

[53] Singh B, Kaunert C. FedHealth in Wearable Healthcare, Orchestrated Federated Deep Learning for Smart Healthcare. In: Federated Deep Learning for Healthcare. CRC Press, 2024: 207-223.

[54] Arif H, Gittens A, Chen P Y. Utility-Privacy Tradeoff in Federated Learning. In: Federated Learning for Medical Imaging. Elsevier, 2025: 95-107.

[55] Komalasari R. Secure and Privacy-Preserving Federated Learning With Explainable Artificial Intelligence for Smart Healthcare Systems. In: Federated Learning and Privacy-Preserving in Healthcare AI. IGI Global, 2024: 288-313.

[56] Nayak S, Nayak S. Exploring the Accuracy and Privacy Tradeoff in AI-Driven Healthcare Through Differential Privacy. In: Proceedings of the 11th International Conference on Information Systems Security and Privacy, 2025: 349-354.

[57] Yogi M K, Chakravarthy A S N. Fusion of Information Theoretical Models with Personalized Differential Privacy to Minimize Privacy Loss in Healthcare Cyber Physical Systems. In: Security Implementation in Internet of Medical Things. CRC Press, 2023: 177-190.

[58] Knolle M, Kaissis G. Differential Privacy. In: Federated Learning for Medical Imaging. Elsevier, 2025: 83-94.

[59] Mishra R, Bankar R. Privacy-Preserving Machine Learning in Healthcare Applications. In: Intelligent Healthcare System, 2025: 413-442.

[60] Thukral A, Dhiman S, Meher R, Bedi P. Knowledge Graph Enrichment from Clinical Narratives Using NLP, NER, and Biomedical Ontologies for Healthcare Applications. International Journal of Information Technology, 2023, 15: 53-65.

[61] Wu L I, Li G. Zero-Shot Construction of Chinese Medical Knowledge Graph with ChatGPT. In: 2023 IEEE International Conference on Medical Artificial Intelligence, 2023: 278-283.

[62] Cogalan C A, Kumluca Topalli A. Knowledge Graph Augmented Retrieval Applications in Healthcare. In: 2025 Medical Technologies Congress, 2025: 1-4.

[63] P D, E S. Embedding-Based Knowledge Graph Construction for Intelligent Healthcare Systems. In: 2025 4th International Conference on Applied Artificial Intelligence and Computing, 2025: 335-341.

[64] Liu Z, Gao X, Li C. Modeling COVID-19 Vaccine Adverse Effects with a Visualized Knowledge Graph Database. Healthcare, 2022, 10(8): 1419.

[65] Saidu F, Wall J. Retrieval-Augmented Large Language Model for Clinical Decision Support with a Medical Knowledge Graph. Electronics, 2025, 15(3): 555.