Journal of Electronic Research and Application, 2026, Volume 10, Issue 8
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Current Status and Development Trends of Biomedical Big Data

Juan Zhang

North Health Medical Big Data Technology Co., Ltd., Jinan 250000, Shandong, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The deep integration of digital technologies has ushered biomedical science into a new era driven by data, making biomedical big data a cornerstone for precision medicine, medical research innovation, and public health prevention and control. Compared to traditional medical data, modern biomedical big data is characterized by multimodality, high dimensionality, and strong interconnectivity, fundamentally transforming conventional empirical and fragmented approaches to research and clinical practice. This paper systematically reviews the current applications of biomedical big data from key medical use cases, analyzes existing challenges in data governance, technological integration, and ethical compliance, moves beyond superficial generalizations, and identifies the main development directions of the field across three dimensions: technological evolution, clinical translation, and system building. It provides theoretical support for the standardized application, industrial implementation, and advancement of biomedical big data in the medical domain.

Keywords: Biomedical big data; Medical applications; Data governance; Precision medicine; Development trends

Online publication: Sept 8, 2026

1. Introduction

The deep integration of life sciences and information technology has liberated biomedical research from the constraints of traditional experimental observations and clinical experience, ushering it into a new paradigm of data-intensive science. Clinical records, omics data, medical imaging, and health monitoring information are continuously accumulating, forming a comprehensive biomedical big data system that spans the entire human lifespan and covers all stages of disease. These data are no longer mere byproducts of diagnosis and treatment but have become key production factors for decoding biological mechanisms, conquering complex diseases, and optimizing healthcare resource allocation. Biomedical big data is increasingly penetrating core areas such as basic medical research, precision clinical care, and public health management. However, significant challenges remain in data integration, technological adaptation, and secure governance. By systematically reviewing the current state of industry development through the lens of medical application needs, numerous existing issues can be identified. Addressing these challenges with foresight on future trends will facilitate the transformation of data value into tangible clinical outcomes, driving the modern medical system toward greater intelligence and precision.

2. Current development status of the biomedical big data industry

2.1. Diversified data systems covering all medical application scenarios

Currently, biomedical big data has established a multidimensional and comprehensive data architecture, breaking through the limitations of traditional medical data that were previously fragmented and limited in scope. It now covers nearly all medical scenarios, including basic research, clinical diagnosis and treatment, public health, and health management [1]. In terms of data types, these mainly include genomics data, clinical care data, medical imaging data, and dynamic health monitoring data. Genomics data represent the microscopic mechanisms of life, encompassing biological information at various levels such as genes, proteins, and metabolites, serving as fundamental data for elucidating disease mechanisms and enabling personalized medicine. Clinical care data consist of patient records, treatment plans, medication logs, follow-up information, and other materials generated from daily operations across healthcare institutions, forming the core resources for evidence-based clinical research. Medical imaging data include ultrasound, MRI, CT, and other imaging modalities, providing direct support for non-invasive disease diagnosis and disease progression assessment. Real-time health status data collected via intelligent medical devices also play a crucial role in daily health management and chronic disease prevention and control.

2.2. Deep implementation in application scenarios, empowering innovation in medical models

The application of biomedical big data has evolved from simple data storage and statistics to deeply empowering the entire medical chain, driving traditional empirical medicine toward precision medicine, evidence-based medicine, and preventive medicine. In clinical practice, integrating and analyzing multimodal data can help healthcare professionals uncover hidden disease characteristics, identify disease progression patterns, improve diagnostic and therapeutic strategies, and reduce misdiagnosis and missed diagnoses, making it particularly suitable for managing complex cases such as difficult-to-diagnose conditions and malignant tumors. In basic medical research, a data-driven approach is replacing the previous reliance on single experiments for validation, enabling researchers to discover connections within vast biological datasets, rapidly identify disease targets and mechanisms, and significantly shorten the time required for medical scientific breakthroughs. In public health, comprehensive integration and analysis of healthcare data allow for dynamic assessment of disease trends, supporting rapid response and precise prevention and control during public health emergencies, and optimizing the allocation of public health resources [2].

2.3. The technical system has gradually become more robust, enabling efficient data processing

With the continuous advancement of information technology, a processing framework tailored to the characteristics of biomedical data has gradually emerged, overcoming the limitations of traditional medical data processing, such as low efficiency and limited analytical dimensions. In terms of data collection and storage, high-throughput acquisition technologies and distributed storage systems have been widely adopted, effectively addressing the massive volume, multi-source nature, and heterogeneity of biomedical data, enabling unified aggregation and long-term preservation of diverse fragmented data [3]. In data processing and analysis, intelligent algorithms and machine learning techniques have been integrated into biomedical applications, facilitating multimodal data cleaning, matching, fusion, and in-depth mining to accurately identify underlying life cycle patterns or disease manifestations. On the application side, innovations such as medical knowledge graphs and intelligent diagnostic support systems have been developed, significantly improving the efficiency of translating big data analytics into clinical practice and scientific research. Through synergistic integration of these technologies, biomedical big data is being transformed into a valuable and practically usable medical resource.

3. Core challenges in the development of biomedical big data

3.1. Prominent data silo problem, making integration and sharing difficult

Data fragmentation and isolation represent the most critical challenges in today's biomedical big data field. Data collection standards, storage formats, and statistical methodologies vary significantly across medical institutions, research organizations, and health management agencies at different levels, preventing interoperability among healthcare data from different regions and tiers. Moreover, some institutions maintain a closed mindset regarding data, lacking proactive willingness to share information, which results in high-quality medical data being scattered and underutilized, failing to form large-scale, systematic data resources. Due to the heterogeneity of multi-source data, issues such as matching discrepancies and information redundancy arise during data integration, greatly reducing data usability. The existence of data silos not only limits the overall value of big data but also prevents medical research and clinical practice from conducting comprehensive analyses based on holistic datasets, severely hindering the large-scale implementation of precision medicine [4].

3.2. Lack of data governance framework, resulting in inconsistent data quality

Biomedical data is directly linked to life and health, and the authenticity, completeness, and standardization of data are crucial for ensuring effective applications. However, the industry has yet to establish a unified and comprehensive data governance system. During data collection, issues such as manual entry errors, equipment measurement biases, and inconsistent clinical documentation often lead to problems in raw data, including missing values, inaccuracies, and redundancy. In the data processing stage, standardized procedures for cleaning, filtering, and calibration are lacking, and different entities apply varying data processing standards, resulting in biased analytical outcomes that cannot be reliably applied across different scenarios. Due to the absence of routine quality control mechanisms and an inadequate system for data updating, maintenance, and traceability, some older data no longer meet current medical development needs, thereby diminishing their application value. The prevalence of low-quality data negatively impacts clinical decision support and scientific research, and may even lead to errors in medical decision-making.

3.3. Poor technical adaptability and low clinical translation efficiency

Currently, most big data technologies rely on general information technology frameworks, resulting in delayed development of specialized tools tailored to biomedical applications and insufficient compatibility between technology and clinical settings. General-purpose intelligent algorithms have limited capacity to analyze the complexity of biomedical data, failing to accurately capture nonlinear relationships within biological systems or uncover hidden patterns underlying complex diseases. As a result, many analytical outcomes suffer from poor interpretability and low relevance to clinical practice. Most big data research achievements remain at the experimental stage, lacking refinement and optimization for real-world clinical use, creating a clear gap between scientific discoveries and practical application, hindering the development of implementable diagnostic methods and interventions. Moreover, primary-level medical institutions often lack technical expertise and sufficient personnel, making it difficult for them to adopt intelligent big data solutions. Consequently, high-quality data and technological resources fail to achieve comprehensive coverage, leading to significant regional and hierarchical imbalances in industry development.

3.4. Inadequate safety ethics system increases compliance risks

Biomedical big data contains a significant amount of personal privacy information and core genetic data, making it highly sensitive. However, current industry security protection and ethical oversight systems still have notable shortcomings. In terms of data security, some institutions lack robust safeguards in data storage, transmission, and application, with insufficient encryption, data anonymization, and access control mechanisms, leading to risks such as data breaches, misuse, and tampering. Regarding ethical regulation, guidelines for data collection, sharing, and usage remain unclear, informed consent mechanisms for patients are inadequate, and issues such as excessive data collection and unauthorized use of private information persist in both research and commercial applications. Additionally, mechanisms for data traceability and accountability are underdeveloped, making it difficult to accurately assign responsibility when data security incidents occur. The absence of a comprehensive security and ethics framework not only violates individuals' rights to privacy but also hinders the development of the biomedical big data industry.

4. Future trends in biomedical big data

4.1. Deepening of data standardization and governance, with comprehensive data integration and sharing becoming the norm

The future of the biomedical big data industry will focus on standardization, fully addressing the issue of data silos and achieving seamless, cross-domain data interconnectivity. The industry will gradually establish unified standards for data collection, storage, classification, and transmission, standardizing data production processes across various medical and research institutions. It will also unify multimodal data formats and statistical methodologies to eliminate data heterogeneity at the source. Leveraging emerging technologies such as privacy-preserving computing and blockchain, secure and controllable data-sharing mechanisms will be built, breaking down data barriers between institutions and regions while safeguarding data privacy. This will enable collaborative data access across different entities and domains. A routine data quality governance framework will emerge, continuously improving data quality through intelligent validation, dynamic updates, and end-to-end traceability systems. Ultimately, this will create a high-quality, trustworthy biomedical big data resource repository, laying a solid foundation for applications across diverse industries.

4.2. Deep integration with technical scenarios, continuously enhancing clinical translation efficiency

The integration of information technology and biomedical sciences will deepen, leading to the gradual establishment of specialized, context-specific data technology systems. In the future, efforts will focus on addressing key challenges in medical settings by developing dedicated intelligent algorithms tailored for multi-omics biology, clinical diagnosis and treatment, and public health prevention and control. These algorithms will enhance the ability to analyze complex medical data and improve the interpretability of results, precisely meeting the needs of clinical research and patient care. A major priority will be breaking down barriers between research and clinical practice, transforming big data analysis outcomes into practical tools such as intelligent diagnostic support systems, personalized treatment plans, and disease risk prediction models, thereby closing the loop from data mining to real-world clinical application. Lightweight and user-friendly big data technologies for primary healthcare will continue to evolve and become widely adopted, helping to reduce disparities in healthcare across regions and levels, promoting nationwide sharing of health information, and making medical services increasingly intelligent.

4.3. Application boundaries continue to expand, forming a full-cycle health service system

The application of biomedical big data will extend from traditional disease diagnosis and treatment, as well as scientific research, to comprehensive health management throughout the entire life cycle, establishing an integrated intelligent healthcare system that covers prevention, diagnosis, rehabilitation, and care. In prevention, long-term collection and analysis of diverse health information enable prediction of disease susceptibility risks, allowing medical interventions to be implemented proactively before illness develops, shifting healthcare focus from disease treatment to health promotion. In diagnosis and treatment, multimodal data integration facilitates fully personalized precision medicine, achieving refined medical models such as “same disease, different treatments” and “different diseases, same treatment.” For rehabilitation and nursing, dynamic health monitoring data are used to develop individualized rehabilitation assessment and intervention systems, making chronic disease management and post-surgical recovery more scientific. Big data will also be deeply integrated into public health emergency response systems, improving rapid reaction and efficient handling of sudden outbreaks or epidemic events.

4.4. Standardization of the safety ethics system leads to orderly and controllable industry development

As the industry develops in the future, a compliance framework integrating technical safeguards, institutional constraints, and ethical oversight will emerge, achieving a synergistic effect between data value realization and security and privacy protection. Technologically, security measures such as data anonymization, encrypted transmission, access control, and privacy-preserving computing will be widely adopted, enabling data to be usable yet invisible, thereby mitigating risks of data breaches and misuse. Institutionally, more detailed regulations will be established for every stage of biomedical data collection, sharing, application, and trading, clearly defining the rights and responsibilities of all stakeholders and establishing mechanisms for accountability in data security. The ethical review system for medical data applications will be strengthened, with improved informed consent processes and privacy protection mechanisms to prevent data abuse and excessive data collection. Comprehensive standardization efforts will be implemented, helping the biomedical big data industry move away from unstructured growth and transition toward high-quality, sustainable development.

5. Peroration

Biomedical big data is the fundamental driving force behind the transformation and development of modern medicine. It has reshaped the approaches to medical research, clinical practice, and public health services, offering new solutions for diagnosing complex diseases, optimizing healthcare resources, and ensuring population-wide health. Although the industry has achieved initial scale, numerous practical challenges remain, including incomplete data integration, inconsistent data quality, weak technological translation, and inadequate attention to security and ethics. In the future, with strengthened standardized data governance, improved adaptability of technologies to specific applications, more comprehensive application systems, and increasingly robust compliance frameworks, biomedical big data will fully realize its core value. It will support the advancement of precision medicine, smart healthcare, and public health initiatives, while providing a strong underpinning for innovation in life sciences and safeguarding public health.

Disclosure statement

The author declares no conflict of interest.

References

[1] Li J, Yang N, 2025, Transformation of Biomedical Research Models Driven by Big Data. Medicine and Philosophy, 46(8): 1–6.

[2] Ke Q, Lu M, Chen F, et al., 2023, Current Status and Reflections on Curriculum Design for Biomedical Big Data Majors. Chinese Journal of Health Statistics, 40(6): 946–949.

[3] Zhang G, Zhao G, Li Y, 2023, Realizing the Value of Production Factors in Biomedical Big Data: Starting from Data Elements. Life Sciences, 35(12): 1545–1552.

[4] Yang S, You D, Shao F, 2022, Exploring the Cultivation of Biomedical Big Data Analysis Capabilities in Biostatistics Majors. Education Teaching Forum, 2022(41): 41–44.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.