

# Big Data-Driven Computer Science: Application Paradigms and Development Paths

Junfang Wang

Handan Preschool Teachers College, Handan, Hebei 056300

**Abstract:** *This article reviews recent advancements in the application of computer science and technology to big data. Through an analysis of key processes—including data collection, cleaning, preprocessing, storage, and management—it identifies core technologies for large-scale data processing. The discussion extends to algorithm optimization and machine learning techniques, underscoring their role in extracting value from data and improving decision-making. Additionally, the article addresses security and privacy challenges in big data, proposing relevant technical solutions. It aims to serve as a reference for further research in computer science within the big data domain.*

**Keywords:** Big data; Computer science; Technology application.

## 1. THE IMPORTANCE OF COMPUTER SCIENCE AND TECHNOLOGY IN BIG DATA PROCESSING

Computer science and technology play an indispensable and crucial role in big data processing. With the rapid development of information technology, data volume has experienced explosive growth, and big data has become a significant force driving social progress, corporate innovation, and scientific research.

Firstly, in terms of data collection, computer science technology has achieved automatic crawling and aggregation of massive, heterogeneous, and real-time data through the design of efficient crawler systems, sensor networks, and other technical means, laying a solid data foundation for big data processing. These technologies not only improve the efficiency and accuracy of data collection but also greatly expand the sources and scope of data.

Secondly, in terms of data storage, the development of computer science technology has driven the rise of emerging storage solutions such as distributed storage systems and cloud storage technologies. These technologies effectively address the limited storage capacity of a single node through horizontal scaling, achieving reliable storage and efficient access to petabytes or even exabytes of data. The application of technologies such as data compression and deduplication further improves the utilization and efficiency of storage space.

Thirdly, in terms of data processing, the contribution of computer science and technology is even more significant. The introduction of technologies such as parallel computing, distributed computing, and stream computing has enabled big data processing to complete complex computational tasks in a short period of time and respond to data changes in real time. The widespread application of technologies such as data mining and machine learning has enabled the hidden value in big data to be excavated and utilized, providing strong support for decision support, risk assessment, predictive analysis, and other fields.

Fourthly, in terms of big data applications, computer science and technology also play a significant role. By constructing data visualization platforms, intelligent recommendation systems, decision support systems, and other applications, big data technology can better serve various industries in society. These applications not only improve work efficiency and decision-making quality, but also promote industrial upgrading and transformation development.

Guo (2025) developed a method using Inertial Measurement Units (IMUs) and Long Short-Term Memory (LSTM) networks for real-time data completion and motion recognition [1]. Advancing system intelligence for governance, Zhang (2025) proposed a neuro-symbolic and blockchain-enhanced multi-agent framework designed for fair and consistent cross-regulatory audit intelligence [2]. Concurrently, significant effort is directed towards securing collaborative and distributed computing environments. Deng and Yang (2025) devised multi-layer defense strategies and privacy-preserving enhancements to protect federated learning frameworks from membership reasoning attacks [3]. The reliability of physical and engineered systems is another critical focus. Lin et al. (2025) established a Bayesian framework for modeling multivariate degradation data with dynamic covariates, providing

a method for assessing and predicting system reliability [4]. Specialized AI applications are rapidly evolving in legal, industrial, and commercial domains. Xie et al. (2024) advanced legal citation text classification using a Conv1D-based approach for multi-class categorization [5]. In industrial inspection, Wu, Luo, and Liao (2025) applied multi-level transfer learning to enable small-sample object detection of surface cracks in concrete structures [6]. For digital advertising, Yi (2025) created a system for real-time fair-exposure ad allocation using contextual bandits-with-knapsacks, targeting small businesses and underserved creators [7]. The performance of these advanced applications is underpinned by innovations in enabling hardware. Tang et al. (2020) contributed to this foundation with their work on the design and optimization of a shallow-angle grating coupler for vertical emission from indium phosphide devices [8]. In the realm of natural language processing and information extraction, Zheng and Jiang (2025) introduced a new methodology for Chinese term extraction from scientific publications [9]. Finally, research addresses the challenges of deploying generative AI responsibly. Chen, Chen, and Tian (2025) built a generative AI comment review system to ensure content compliance [10], while Deng (2025) proposed a homomorphic encryption-based mechanism for data integrity verification and anti-tampering in cloud storage environments, addressing critical security concerns for data at rest [11].

## **2. THE RESEARCH SIGNIFICANCE OF COMPUTER SCIENCE AND TECHNOLOGY APPLICATION IN THE CONTEXT OF BIG DATA**

In the context of big data, conducting in-depth research on the application of computer science and technology holds profound significance. Firstly, big data, as an important feature and valuable resource of the current era, presents significant challenges to traditional data processing and analysis methods due to its vast scale, diverse types, rapid generation, and low value density. The rapid development of computer science and technology, however, provides powerful tools and means to address these challenges. Studying the application of computer science and technology in the context of big data helps us better understand and grasp the essence and characteristics of big data, and explore more efficient, accurate, and intelligent data processing methods and technologies. This not only enhances the speed and efficiency of data processing but also uncovers hidden values and patterns in data, providing more scientific and precise support for decision-making, risk assessment, predictive analysis, and other fields [1]. With the widespread application of big data technology, various industries are actively exploring how to leverage big data to drive their innovation and development. In-depth research on computer science and technology can provide these industries with more advanced and reliable technical support and solutions, promoting the deep integration and collaborative development of big data technology with various industries. The research on the application of computer science and technology in the context of big data also carries important social significance. Through in-depth research on the application of big data technology, we can better protect personal privacy and data security, prevent data leakage and abuse. We can also utilize big data technology to optimize social resource allocation, improve public service levels, promote the modernization of social governance, and contribute more wisdom and strength to social development and progress.

## **3. THE INTEGRATION OF BIG DATA AND COMPUTER SCIENCE AND TECHNOLOGY**

### **3.1 The relationship and connection between big data and computer science and technology**

Big data and computer science technology maintain a close relationship characterized by mutual dependence and mutual promotion. As the core resource of the modern information society, big data poses severe challenges to traditional data processing methods and systems due to its immense scale and high complexity. On the other hand, computer science technology, as the cornerstone of information technology, provides robust technical support for the collection, processing, storage, analysis, and application of big data through continuous innovation and development. The close connection between the two not only manifests in the mutual integration at the technical level but also lies in their joint promotion of progress in the field of information technology and leading socio-economic transformation. Big data requires computer science technology to provide efficient, reliable, and intelligent technical means, while the development of computer science technology cannot be separated from the nourishment and promotion of big data. The two complement each other and jointly constitute an important feature of the information age.

### **3.2 The role of computer science technology in big data acquisition, storage, processing, and analysis**

In the lifecycle of big data, computer science and technology play a crucial role. In the data acquisition stage, computer science and technology enable the automated collection and integration of data from various sources and formats through the development of efficient crawler programs, API interfaces, and Internet of Things (IoT) technologies, ensuring the real-time, comprehensive, and accurate nature of big data. In terms of data storage, computer science and technology drive the development of new storage technologies such as distributed storage systems and cloud storage platforms. These technologies not only address the capacity issue of big data storage but also ensure data security and reliability through data redundancy and fault-tolerant mechanisms. In the data processing and analysis stage, computer science and technology shine even more brilliantly, achieving rapid processing, deep mining, and intelligent analysis of massive data through advanced technologies such as parallel computing, distributed computing, stream processing, and machine learning, providing powerful technical support for the exploration and utilization of data value [2].

### **3.3 The impact of computer science technology on big data application scenarios**

The development of computer science and technology has had a profound impact on the expansion and deepening of big data application scenarios. On the one hand, with the enhancement of computing power and optimization of algorithms, the application scenarios of big data are becoming increasingly diverse and enriched. In various fields such as finance, healthcare, education, transportation, and smart cities, big data technology, combined with computer science and technology, has driven the optimization of business processes, improved decision-making efficiency, and innovated service models. For example, in the financial sector, big data risk control systems can analyze customers' transaction behaviors in real time, identify potential risk points, and provide precise risk assessment and early warning services for financial institutions. In the medical field, the integration of big data and artificial intelligence has promoted the development of precision medicine, offering patients more personalized and efficient treatment plans. On the other hand, the development of computer science and technology has also facilitated the popularization and application of big data technology. Through the application of technologies such as cloud computing and big data platforms, the barriers and costs associated with using big data technology have been reduced, enabling more enterprises and organizations to leverage big data technology to drive their own development and innovation.

## **4. APPLICATION OF COMPUTER SCIENCE AND TECHNOLOGY IN BIG DATA PROCESSING**

### **4.1 Application of data mining and analysis technology in big data**

In big data processing, data mining and analysis technologies play a crucial role. These technologies utilize various algorithms such as statistics, machine learning, and pattern recognition to deeply explore hidden patterns and association rules within large datasets, thereby extracting valuable information and knowledge. Data mining technology can handle massive and complex data, discover potential relationships between data, provide scientific basis for corporate decision-making, while analysis technology further interprets and predicts data, helping enterprises grasp market trends, optimize product strategies, and enhance operational efficiency. In the context of big data, the application of data mining and analysis technologies not only improves the accuracy and efficiency of data processing but also promotes the maximization of data value exploration and utilization.

### **4.2 Application and development of artificial intelligence technology in big data**

As an important branch of computer science, artificial intelligence technology is playing an increasingly significant role in big data processing. By simulating human intelligent behaviors such as learning, reasoning, and decision-making, artificial intelligence technology conducts in-depth analysis and processing of big data. In the big data environment, artificial intelligence technology can automatically learn the characteristics of data, construct complex predictive models, and achieve precise data analysis and prediction. It can also be applied to intelligent processing of big data, such as automated classification, clustering, and anomaly detection, enhancing the level of intelligence in data processing. With the continuous development of artificial intelligence technologies such as deep learning, natural language processing, and reinforcement learning, the capability and effectiveness of big data processing will be further enhanced, bringing more innovation and development opportunities to various industries.

### **4.3 The contribution of cloud computing and edge computing to big data processing**

Cloud computing reduces the cost and threshold of big data processing by providing elastic and scalable computing resources and storage services. Enterprises can store massive amounts of data in the cloud and utilize the powerful computing capabilities of cloud platforms for efficient processing and analysis. Cloud computing also provides a rich set of data processing and analysis tools to help users quickly build big data applications. Edge computing, on the other hand, reduces data transmission latency and bandwidth consumption by sinking data processing and analysis capabilities to the edge of the network, thereby enhancing the real-time performance and efficiency of big data processing. In big data processing, cloud computing and edge computing complement each other, jointly providing strong support for the collection, storage, processing, and analysis of big data.

## **5. RESEARCH METHODS FOR THE APPLICATION OF COMPUTER SCIENCE TECHNOLOGY IN THE CONTEXT OF BIG DATA**

### **5.1 Technical methods for data collection, cleaning, and preprocessing**

In the context of big data, data collection, cleaning, and preprocessing are crucial prerequisites for ensuring data quality and enhancing the efficiency of subsequent analysis. In the data collection stage, computer science technology utilizes automated tools such as web crawlers, API integration, and real-time data stream capture techniques to efficiently and accurately retrieve data from various sources such as the Internet, IoT devices, and enterprise systems. These technologies not only broaden and deepen the scope of data collection but also ensure the real-time nature and diversity of the data [3]. Data cleaning is a key step in the data processing process, involving tasks such as removing duplicate data, correcting erroneous data, filling missing values, and handling outliers. Computer science technology applies data cleaning algorithms and tools, such as regular expression matching and data quality assessment models, to automatically identify and correct issues in the data, thereby enhancing the accuracy and reliability of the data. The preprocessing stage further processes the cleaned data to better meet subsequent analytical needs. This includes steps such as data conversion (e.g., format conversion, unit normalization), data standardization (scaling data to fall within a specific small interval), and data normalization (scaling data to fall within a specific small interval, typically [0,1]). These preprocessing techniques help reduce differences between data and enhance the accuracy and efficiency of data analysis.

### **5.2 Efficient big data storage and management technology**

Faced with the massive, high-speed, and diverse characteristics of big data, computer science and technology have provided a variety of efficient storage and management techniques. Distributed storage systems such as Hadoop HDFS achieve horizontal scaling and fault tolerance by distributing data blocks across multiple nodes in the network. This distributed architecture not only enhances the capacity and reliability of the storage system but also accelerates data read and write speeds through parallel processing mechanisms. To further improve the utilization rate and access efficiency of storage space, data compression and deduplication techniques are widely used in big data storage. Data compression technology reduces storage requirements by eliminating redundant information in data, while deduplication technology reduces storage space waste by identifying and deleting duplicate data blocks. In terms of data management, new data management solutions such as NoSQL databases and data warehouses provide flexible data models and efficient query performance. NoSQL databases support non-relational data structures, such as key-value pairs, documents, column families, and graphs, which better accommodate the diversity of big data. Data warehouses, on the other hand, integrate data from different sources into a unified storage environment through the data integration, transformation, and loading (ETL) process, providing convenient data access interfaces for data analysis.

### **5.3 Big data analysis and algorithm optimization methods**

Big data analysis is a crucial step in unlocking the value of big data and discovering patterns hidden within it. Computer science technology offers a variety of big data analysis methods and algorithm optimization strategies. Machine learning algorithms, such as decision trees, random forests, and neural networks, are widely applied in tasks like big data classification, clustering, and prediction. These algorithms achieve intelligent analysis and mining of big data by automatically learning data features and constructing predictive models. To expedite the big data processing process and enhance analysis efficiency, parallel computing and distributed computing technologies are extensively utilized in big data analysis. These technologies accelerate the entire analysis process by dividing large datasets into multiple smaller chunks and processing these chunks in parallel across multiple

computing nodes. Algorithm optimization is also a significant method to improve the efficiency of big data analysis [4]. Researchers are continually exploring new algorithmic models and optimization strategies, such as variants of gradient descent algorithms and parallel implementations of random forests, to further enhance the accuracy and efficiency of big data analysis.

#### 5.4 Application of computer science technology in big data security and privacy protection

With the widespread application of big data and the continuous accumulation of data, issues related to data security and privacy protection have garnered increasing attention. Computer science and technology have provided various solutions and technical means in this regard. Encryption technology is an important means to protect data confidentiality. By applying encryption algorithms such as AES and RSA to encrypt sensitive data, the security of data during transmission and storage can be ensured. Even if the data is illegally obtained, it cannot be decrypted and accessed by unauthorized users. Access control and identity authentication technologies are used to ensure that only authorized users can access sensitive data. These technologies prevent unauthorized access and data leakage by implementing strict access permission management and identity authentication mechanisms. In terms of privacy protection, new technologies such as differential privacy and federated learning are applied in big data analysis. Differential privacy technology protects user privacy information from being leaked by adding random noise during data analysis; while federated learning allows multiple participants to jointly train machine learning models without sharing raw data, thus achieving a win-win situation of effective data utilization and privacy protection. Data desensitization and anonymization technologies are also widely used in the process of big data processing to reduce the risk of data leakage. These technologies reduce the risk of data leakage and protect user privacy by desensitizing sensitive data or converting it into an anonymous form.

## 6. CONCLUSION

The research on the application of computer science and technology in the context of big data has not only driven the innovation of data processing and analysis technologies but also promoted the development of data security and privacy protection. With the continuous advancement of technology and the continuous expansion of application scenarios, it is reasonable to believe that computer science and technology will play a more important role in the era of big data, bringing more intelligent, efficient, and secure data processing solutions to various industries. In the future, we should continue to deepen research in related fields, promote the innovation and application of big data technology, and jointly embrace a brighter digital future.

## REFERENCES

- [1] Guo, Y. (2025, May). IMUs Based Real-Time Data Completion for Motion Recognition With LSTM. In Forum on Research and Innovation Management (Vol. 3, No. 6).
- [2] Zhang, T. (2025). A Neuro-Symbolic and Blockchain-Enhanced Multi-Agent Framework for Fair and Consistent Cross-Regulatory Audit Intelligence.
- [3] Deng, X., & Yang, J. (2025, August). Multi-Layer Defense Strategies and Privacy Preserving Enhancements for Membership Reasoning Attacks in a Federated Learning Framework. In 2025 5th International Conference on Computer Science and Blockchain (CCSB) (pp. 278-282). IEEE.
- [4] Lin, Z., Liu, X., Xiang, Y., & Hong, Y. (2025). Modeling multivariate degradation data with dynamic covariates under a Bayesian framework. *Reliability Engineering & System Safety*, 111115.
- [5] Xie, Y., Li, Z., Yin, Y., Wei, Z., Xu, G., & Luo, Y. (2024). Advancing legal citation text classification A Conv1D-based approach for multi-class classification. *Journal of Theory and Practice of Engineering Science*, 4(02), 15-22.
- [6] Wu, J., Luo, L., & Liao, N. (2025). Small-Sample Object Detection of Surface Cracks in Concrete Structures of High-Rise Buildings via Multi-Level Transfer Learning. *Innovation & Technology Advances*, 3(2), 57–72. <https://doi.org/10.61187/ita.v3i2.262>
- [7] Yi, X. (2025, October). Real-Time Fair-Exposure Ad Allocation for SMBs and Underserved Creators via Contextual Bandits-with-Knapsacks. In Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (pp. 1602-1607).
- [8] Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020). Design and Optimization of Shallow-Angle Grating Coupler for Vertical Emission from Indium Phosphide Devices.

- [9] Zheng, H., & Jiang, T. (2025). A New Methodology for Chinese Term Extraction from Scientific Publications. *Innovation & Technology Advances*, 3(2), 19–45. <https://doi.org/10.61187/ita.v3i2.222>
- [10] Chen, Rensi, Ziming Chen, and Yueqi Tian. "Building a Generative AI Comment Review System for Content Compliance." *Proceedings of the 2nd International Symposium on Integrated Circuit Design and Integrated Systems*. 2025.
- [11] Deng, X. (2025). Homomorphic Encryption-Based Data Integrity Verification and Anti-Tampering Mechanism in Cloud Storage Environment.