# Research on Big Data Analysis Methods Based on Artificial Intelligence Technology

**Jun Wang**

*Nanjing 30000IoT Technology Co., Ltd, Nanjing, Jiangsu, China*
*wangjun@30000iot.com*

*Abstract:* This paper aims to explore big data analysis methods based on artificial intelligence technology and analyze their applications in addressing challenges and issues in the real world. Firstly, an overview of big data analysis and artificial intelligence technology is provided, including their definitions, importance, and application areas. Secondly, a detailed discussion on big data analysis methods based on artificial intelligence technology is presented, covering aspects such as data preprocessing, feature extraction and selection, as well as model establishment and evaluation. Based on this foundation, the paper showcases the practical applications of these methods in domains such as healthcare, finance, and e-commerce through case studies. Lastly, the strengths and limitations of current methods are discussed, along with suggestions for future development directions and methodological improvements. This study aims to provide theoretical guidance and practical experience for research and practice in the field of big data analysis.

## 1. Introduction

With the rapid development of information technology and the prevalence of the Internet, the era of big data has arrived. Big data not only refers to the increase in data volume but more importantly, the extraction of valuable information and knowledge from these vast amounts of data. Big data analysis, as a powerful tool and method, has become one of the essential means to address various complex problems in the real world. However, with the continuous expansion of data scale and the increasing variety of data types, traditional data analysis methods and technologies face numerous challenges, such as the cost of data storage and processing, data quality and accuracy, privacy protection, among others. Meanwhile, the rapid development of artificial intelligence technology has provided new insights and solutions for big data analysis. Machine learning, deep learning, natural language processing, computer vision, and other artificial intelligence technologies are playing increasingly important roles in big data analysis, providing strong support for data processing, pattern recognition, predictive analysis, and more. Big data analysis methods based on artificial intelligence technology can not only handle massive amounts of data more effectively but also discover hidden patterns and information in the data, providing scientific basis for decision-making and innovation. This paper aims to delve into big data analysis methods based on artificial intelligence technology and explore their applications in addressing various challenges and

problems in the real world. We will first introduce the basic concepts and current development status of big data analysis and artificial intelligence technology. Then, we will discuss in detail the big data analysis methods based on artificial intelligence technology, including data preprocessing, feature extraction and selection, as well as model establishment and evaluation. Through case studies, we will demonstrate the practical applications of these methods in domains such as healthcare, finance, and e-commerce. Finally, we will analyze the strengths and limitations of current methods and propose future development directions and methodological improvements, aiming to provide theoretical guidance and practical experience for research and practice in the field of big data analysis.

## 2. Big Data Analysis and Artificial Intelligence

### 2.1. Overview of Big Data Analysis

With the rapid development of information technology and the widespread use of the Internet, big data has become an essential resource in today's society. Big data not only refers to the vast quantity of data but also encompasses its diversity, velocity, and value. Big data analysis involves the collection, storage, processing, and analysis of large-scale data using various techniques and tools to discover patterns, trends, and value hidden within the data. Big data analysis can help businesses and organizations better understand customer needs, optimize operational processes, improve product quality, forecast future trends, and ultimately make better business decisions and foster innovation. The process of big data analysis typically includes four main stages: data collection, data storage, data processing, and data analysis. Firstly, data needs to be collected from various sources, including structured data (such as databases, spreadsheets) and unstructured data (such as text, images, videos). Then, this data is stored in corresponding data storage systems, such as data warehouses, data lakes, etc. Subsequently, various data processing techniques are applied to clean, transform, and integrate the data to ensure its quality and consistency. Finally, data mining and analysis tools and algorithms are utilized to explore and analyze the data, uncovering patterns, associations, and trends to support decision-making. In conclusion, big data analysis has become an important means for enterprises and organizations to enhance competitiveness and create value. With the continuous increase in data volume and the enrichment of data types, big data analysis will play an increasingly important role in the future as a crucial engine driving innovation and development [1].

### 2.2. Overview of Artificial Intelligence Technology

Artificial Intelligence (AI) is a computer system that mimics human intelligence behaviors. With the continuous development of computer science and information technology, artificial intelligence technology has made significant progress and has been widely applied in various fields. The core of artificial intelligence technology is to enable computer systems to simulate human cognitive abilities, including learning, reasoning, planning, language understanding, visual perception, etc. In the process of developing artificial intelligence technology, techniques such as machine learning (ML) and deep learning (DL) have played a crucial role. Machine learning is a technique that improves the performance of computer systems by training data, including supervised learning, unsupervised learning, and reinforcement learning methods. Deep learning is a special form of machine learning that simulates the working principles of the human brain's neurons through multi-layer neural networks, enabling the extraction and representation of advanced features from complex data, thus achieving more accurate and effective data analysis and prediction. In addition to machine learning and deep learning, artificial intelligence technology also involves fields such as

natural language processing (NLP), computer vision, and intelligent recommendation systems. Natural language processing technology enables computers to understand and process natural language, including speech recognition, text analysis, machine translation, etc. Computer vision technology enables computers to understand and process image and video data, including object detection, image recognition, face recognition, etc. Intelligent recommendation systems utilize artificial intelligence technology to analyze and predict user preferences and behaviors, providing personalized recommendation services to users. In summary, artificial intelligence technology has become an important component of the information technology field, exerting profound impacts on socio-economic development and human life. With the continuous advancement of technology and the expansion of application scenarios, artificial intelligence technology will play an increasingly important role in the future, creating smarter, more convenient, and safer living environments for humanity[2].

## 3. Big Data Analysis Methods Based on Artificial Intelligence Technology

### 3.1. Data Preprocessing

Data preprocessing is a crucial step in the process of big data analysis, involving operations such as data cleaning, transformation, integration, and normalization. Its aim is to improve data quality, reduce noise and redundancy, and provide a high-quality data foundation for subsequent data analysis and mining. In big data analysis based on artificial intelligence technology, various intelligent methods and techniques are commonly employed during the data preprocessing stage to enhance the efficiency and accuracy of data processing.Data cleaning is the first step in data preprocessing, which includes operations such as handling missing values, detecting and handling outliers, and removing duplicate data. Artificial intelligence technology can assist in identifying and processing anomalies and noise in the data. For example, machine learning algorithms can automatically identify outliers or missing values and take appropriate actions to handle or fill them. Moreover, artificial intelligence technology can predict potential anomalies in the future by learning patterns and rules from historical data, thus enabling proactive prevention and handling.Data transformation is the second step in data preprocessing, involving operations such as format conversion, attribute transformation, and feature extraction[3]. In big data analysis, data often comes in various types and formats, which need to be transformed into a unified format and representation for subsequent analysis and mining. Artificial intelligence technology, utilizing techniques such as natural language processing and computer vision, automates the processing and transformation of unstructured data into structured data or feature vectors, thereby providing a data foundation for the establishment and application of machine learning and deep learning models.Data integration is the third step in data preprocessing, which involves integrating data from multiple sources into a unified data warehouse or data lake for comprehensive analysis and mining. Artificial intelligence technology can leverage techniques such as natural language processing and knowledge graphs to semantically understand and correlate data from different sources, achieving automatic data integration and fusion to enhance data comprehensiveness and accuracy.Data normalization is the final step in data preprocessing, including operations such as standardization, normalization, and discretization, to ensure data consistency and comparability. Artificial intelligence technology, utilizing machine learning and deep learning techniques, models and learns the distribution and features of the data to achieve automated normalization and processing, reducing manual intervention and errors.In summary, big data preprocessing methods based on artificial intelligence technology can enhance the efficiency and accuracy of data processing, providing a high-quality data foundation for subsequent data analysis and mining, thus enabling more precise and effective big data analysis and application.

## 3.2. Feature Extraction and Selection

Feature extraction and selection are essential aspects of big data analysis based on artificial intelligence technology, involving the extraction of valuable features from raw data and the selection of the most relevant features for model building and prediction. In big data analysis, where data often exhibits high dimensionality and complexity, intelligent methods and techniques are required for feature extraction and selection to reduce feature dimensionality, improve model efficiency, and enhance prediction accuracy. Feature extraction involves extracting features relevant to the target task from raw data. In big data analysis based on artificial intelligence technology, feature extraction typically employs machine learning and deep learning methods to automatically extract the most representative and discriminative features by learning patterns and rules from the data. For example, in image data, convolutional neural networks (CNNs) can be used to extract features such as edges, textures, and colors. In text data, techniques such as bag-of-words models and word embeddings can be used to extract features such as word frequency and word vectors. Feature selection involves selecting the most relevant and valuable features from the extracted features to reduce feature dimensionality, improve model efficiency, and enhance prediction accuracy. In big data analysis based on artificial intelligence technology, feature selection commonly employs various strategies and algorithms, such as filter, wrapper, and embedded methods. Filter methods rank and select features based on statistical metrics or correlation coefficients. Wrapper methods evaluate and select feature subsets through techniques such as cross-validation or heuristic search. Embedded methods embed feature selection into the model training process, achieving feature selection and model establishment simultaneously through techniques such as regularization. In summary, feature extraction and selection are crucial aspects of big data analysis based on artificial intelligence technology, aiding in improving the efficiency and accuracy of data analysis, providing robust support for the establishment of accurate and effective models, and enabling more refined and intelligent big data analysis and application.

## 3.3. Model Establishment and Evaluation

In big data analysis based on artificial intelligence technology, model establishment and evaluation are critical steps in achieving data analysis goals. This stage involves selecting suitable machine learning or deep learning models, training, optimizing, and evaluating them to obtain accurate descriptions or predictions of the data.Model establishment: Firstly, it is necessary to select the appropriate model based on the features of the data and the analysis goals. Commonly used models in big data analysis include linear regression, decision trees, support vector machines, neural networks, etc. Depending on the situation, a single model or a combination of multiple models can be employed for modeling [4]. During the model establishment process, data needs to be partitioned, including dividing the dataset into training, validation, and testing sets for model training, parameter tuning, and evaluation. Model training and optimization: Once the model is selected, it needs to be trained using the training set and optimized through various optimization algorithms to improve its fitting and generalization abilities. Various optimization algorithms such as stochastic gradient descent and genetic algorithms can be utilized during the training process to reduce the model's loss function and enhance its performance. Model evaluation: After training, the model needs to be evaluated and fine-tuned using the validation set to assess its performance and generalization ability. Common evaluation metrics include accuracy, precision, recall, F1 score, etc. Additionally, techniques such as cross-validation can be used for robustness assessment to ensure the reliability and stability of the model. Finally, the model is validated using the testing set to evaluate its prediction capability on unseen data. Model interpretation and application: For complex models, model interpretation and explainability analysis can be conducted to aid in understanding the

model's decision-making process and prediction results. Furthermore, the trained model is applied to real-world scenarios to achieve predictions and decision support for future data.In summary, model establishment and evaluation are core aspects of big data analysis based on artificial intelligence technology, enabling accurate descriptions and predictions of data and providing scientific basis for decision-making. Through steps such as selecting models, optimizing model parameters, and evaluating model performance, high-quality and reliable data analysis results can be obtained, providing crucial support for intelligent big data applications[5].

## 4. Data Privacy and Security

Against the backdrop of big data analysis, concerns regarding data privacy and security are increasingly capturing people's attention. With the development and application of big data technologies, the issue of personal data leaks and misuse has become more prominent, posing challenges to individual rights and societal stability. Therefore, data privacy protection and security have become crucial issues that cannot be overlooked in the process of big data analysis.Firstly, big data often contains a vast amount of sensitive information, such as personal identities, health records, financial transactions, etc. The leakage of such information may lead to infringements of personal privacy rights and even pose risks to personal safety and property. Hence, safeguarding user data privacy in the process of big data analysis has become an urgent matter. Traditional data anonymization and encryption techniques are no longer sufficient to meet the demands for protecting user data privacy. Further research and exploration of effective data privacy protection methods, such as differential privacy and homomorphic encryption, are necessary to ensure the security and privacy of user data. Secondly, data sharing and exchange play a crucial role in big data analysis, but they also entail risks of data leaks and misuse. Particularly in cross-institutional and cross-border data sharing processes, uncertainties about data ownership and management authority may easily lead to improper use and abuse of data. Therefore, it is essential to establish effective data sharing mechanisms and privacy protection measures to ensure that data is shared and used reasonably and securely, protecting user privacy rights and data security. Additionally, data security issues are also receiving significant attention. Big data analysis involves the storage and processing of massive amounts of data, and data security issues encompass aspects such as data storage, transmission, and processing. Data is vulnerable to hacking attacks and malicious tampering during storage and transmission, leading to data leaks and damage. Hence, effective data security measures, such as encrypted transmission, access control, and identity verification, are needed to safeguard the security and integrity of data during storage and transmission. In conclusion, data privacy protection and security are crucial issues in the process of big data analysis. By adopting effective privacy protection and security measures, user privacy rights and data security can be safeguarded, ensuring the legitimacy and credibility of the big data analysis process.

## 5. Application Cases

In the field of healthcare, big data analysis and artificial intelligence technology have achieved significant results. On one hand, by utilizing big data analysis techniques to deeply mine medical data, doctors can obtain more information about disease trends and treatment outcomes, thereby assisting in the formulation of diagnoses and treatment plans. For example, using machine learning algorithms to analyze medical imaging data can achieve tumor detection and assist in diagnosis, thereby improving the accuracy and efficiency of early tumor detection. On the other hand, artificial intelligence technology also plays an important role in clinical practice. By employing natural language processing techniques to analyze clinical documents, doctors can quickly access medical history information and medical knowledge, thereby assisting in clinical decision-making and

treatment plan formulation. Additionally, in the financial sector, big data analysis and artificial intelligence technology also play a crucial role. For instance, analyzing customer behavior data using machine learning models can achieve fraud detection and risk warning, helping banks and financial institutions reduce risks and improve service quality. Furthermore, mining market data using big data analysis techniques can assist investors in formulating more scientific investment strategies and asset allocation plans, thereby enhancing investment returns and risk management levels. These application cases fully demonstrate the potential and value of big data analysis and artificial intelligence technology in the healthcare and financial sectors, providing important support and driving force for the development and progress of relevant industries[6].

## 6. Challenges in Research on Artificial Intelligence Technology in Big Data Analysis Methods

In the research of applying artificial intelligence technology to big data analysis methods, despite significant progress, several challenges are faced. Firstly, data quality and accuracy are pressing issues. Big data often suffers from problems such as incompleteness, inconsistency, and noise, posing challenges to the training and application of artificial intelligence models. Addressing these issues requires advanced data cleaning and preprocessing techniques to improve data quality and accuracy. Secondly, data privacy and security concerns are also significant. Big data contains vast amounts of sensitive information, such as personal health records and financial transaction information, and ensuring data sharing and analysis while protecting data security remains a challenge. Effective data privacy protection and security mechanisms need to be designed to safeguard the privacy and security of user data. Additionally, the issue of model interpretation and explainability is a hotspot. Complex artificial intelligence models often struggle to explain their decision-making processes and prediction results, affecting the credibility and acceptability of the models. Enhancing model interpretability to make their decision-making processes more transparent and understandable is an important challenge. Moreover, innovation in algorithms and technologies is essential. With the increasing scale and diversity of data, continuous innovation and improvement of artificial intelligence algorithms and technologies are needed to address more complex data analysis tasks and application scenarios. For example, innovation in areas such as processing and analyzing unstructured data and integrating multimodal data is currently a research hotspot and challenge. Finally, computational resources and efficiency issues need to be addressed. Big data analysis often requires significant computational and storage resources, especially when training deep learning models. Improving the efficiency and speed of data analysis using distributed computing and parallel computing technologies is an important problem that needs to be addressed. In summary, research on artificial intelligence technology in big data analysis methods still faces numerous challenges, requiring interdisciplinary research and collaboration, continuous innovation and breakthroughs to better address the complex problems and challenges in the real world.

## 7. Conclusion

In this paper, we have delved into the methods of big data analysis based on artificial intelligence technology and demonstrated their applications in the fields of healthcare and finance through specific cases. By outlining big data analysis and artificial intelligence technology, we have gained an understanding of their fundamental concepts, significance, and application areas. Building upon this foundation, we have elaborated on methods such as data preprocessing, feature extraction and selection, and model establishment and evaluation, elucidating their roles and significance in big data analysis. Through case studies, we have showcased the practical applications of these methods in healthcare and finance, illustrating their crucial role in improving medical diagnosis and enhancing financial risk management. Furthermore, we have discussed the strengths and limitations

of current methods, and proposed directions for future development and methodological improvements. Summing up the above, we can conclude that big data analysis methods based on artificial intelligence technology hold immense potential and application prospects, playing a crucial role in promoting socio-economic development and addressing real-world problems. However, challenges such as data quality, privacy protection, and model interpretability still exist and require further research and exploration. We believe that with continued technological advancement and methodological refinement, big data analysis methods based on artificial intelligence technology will increasingly play a vital role in the future, providing robust support for achieving intelligence and sustainable development.

## References

*[1] Zhang Z. Big data analysis with artificial intelligence technology based on machine learning algorithm [J]. Journal of Intelligent & Fuzzy Systems, 2020, 39(5): 6733-6740.*
*[2] Kibria M G, Nguyen K, Villardi G P, et al. Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks [J]. IEEE access, 2018, 6: 32328-32338.*
*[3] Zhu H. Big data and artificial intelligence modeling for drug discovery [J]. Annual review of pharmacology and toxicology, 2020, 60: 573-589.*
*[4] Duan Y, Edwards J S, Dwivedi Y K. Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda[J]. International journal of information management, 2019, 48: 63-71.*
*[5] Kumar P, Singh A, Rajput V D, et al. Role of artificial intelligence, sensor technology, big data in agriculture: next-generation farming[M]//Bioinformatics in Agriculture. Academic Press, 2022: 625-639.*
*[6] Samara D, Magnisalis I, Peristeras V. Artificial intelligence and big data in tourism: a systematic literature review [J]. Journal of Hospitality and Tourism Technology, 2020, 11(2): 343-367.*