iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76747

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

Recent Advances in Big Data Analytics and Artificial Intelligence: A Comprehensive Review

Bharat Shelke[1], Prajkta More[2]
[1]*Associate Professor, Department of Computer science, SCS College, Omerga, Dharashiv*
[2]*Research Scholar, Department of Management Science, Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhajinagar*

***Abstract**: The rapid growth of data generated from digital platforms, sensors, and interconnected systems has led to the emergence of Big Data as a critical resource for intelligent decision-making. At the same time, Artificial Intelligence (AI) has evolved significantly, offering powerful techniques to extract meaningful insights from large-scale and complex datasets. This paper presents a comprehensive review of recent advances in Big Data analytics and Artificial Intelligence, focusing on their integration, methodologies, and practical applications. Key Big Data frameworks and analytics paradigms are discussed alongside state-of-the-art AI techniques, including machine learning, deep learning, and natural language processing. The review highlights recent trends such as real-time analytics, scalable AI models, and generative intelligence, while also examining major challenges related to data quality, privacy, scalability, and model interpretability. Furthermore, emerging research directions, including explainable AI, edge–cloud intelligence, and sustainable AI systems, are explored. This review aims to provide researchers and practitioners with a consolidated understanding of current developments and future opportunities at the intersection of Big Data analytics and Artificial Intelligence.*

***Keywords:** Big Data Analytics; Artificial Intelligence; Machine Learning; Deep Learning; Real-Time Data Processing; Distributed Computing; Intelligent Systems; Emerging Trends.*

I. INTRODUCTION

In recent years, the rapid expansion of digital technologies has resulted in an unprecedented volume of data generated from diverse sources such as social media platforms, Internet of Things (IoT) devices, cloud services, enterprise systems, and scientific applications. This massive growth of data, commonly referred to as Big Data, has transformed the way information is collected, stored, and analyzed. Traditional data processing techniques are no longer sufficient to handle the scale, complexity, and velocity of such data, thereby necessitating advanced analytical approaches. Big Data analytics focuses on extracting meaningful patterns, knowledge, and insights from large and heterogeneous datasets. However, the true potential of Big Data can only be realized when it is combined with Artificial Intelligence (AI). AI provides intelligent mechanisms that enable machines to learn from data, adapt to new patterns, and make accurate predictions or decisions with minimal human intervention. Techniques such as machine learning, deep learning, and natural language processing have demonstrated remarkable success in handling complex Big Data analytics tasks. The integration of Big Data analytics and AI has led to significant advancements across various domains, including healthcare, finance, smart cities, manufacturing, and cybersecurity. AI-driven analytics enables real-time decision-making, predictive analysis, automation, and enhanced operational efficiency. Moreover, recent developments in scalable computing frameworks, cloud platforms, and distributed processing have further strengthened the synergy between Big Data and AI.

Despite these advancements, several challenges remain, such as data quality issues, privacy and security concerns, high computational costs, and the lack of model transparency. Addressing these challenges is essential for the development of reliable and trustworthy intelligent systems. Consequently, a systematic review of recent progress, emerging trends, and open research issues is required. This paper presents a comprehensive review of recent advances in Big Data analytics and Artificial Intelligence. It examines key technologies, methodologies, applications, challenges, and future research directions, aiming to provide a consolidated reference for researchers, academicians, and practitioners working in this rapidly evolving field.[1][2][3]

II. BIG DATA ANALYTICS: AN OVERVIEW

Big Data analytics refers to the process of examining large, complex, and heterogeneous datasets to uncover hidden patterns, correlations, and useful insights that support data-driven decision-making. Unlike traditional data analytics, Big Data analytics deals with massive data volumes generated at high speed and in multiple formats, requiring scalable storage and processing techniques.[4][5][6]

©IJRASET: All Rights are Reserved

A. Characteristics of Big Data

Big Data is commonly described using the concept of the 5Vs:

1) Volume: Refers to the enormous amount of data generated from sources such as social media, sensors, transaction systems, and multimedia content.
2) Velocity: Indicates the speed at which data is produced, collected, and processed, often in real time or near real time.
3) Variety: Represents the diversity of data formats, including structured, semi-structured, and unstructured data such as text, images, audio, and video.
4) Veracity: Addresses the reliability, accuracy, and quality of data, which may be affected by noise, inconsistency, or incompleteness.
5) Value: Emphasizes the importance of extracting meaningful and actionable insights from raw data.

Understanding these characteristics is essential for designing effective analytics solutions and selecting appropriate processing tools.

B. Big Data Analytics Frameworks

To manage and analyze Big Data efficiently, several distributed computing frameworks have been developed. Apache Hadoop is one of the earliest frameworks, providing distributed storage through the Hadoop Distributed File System (HDFS) and batch processing using the MapReduce programming model. While Hadoop is suitable for large-scale batch analytics, it has limitations in handling real-time data.

To overcome these limitations, Apache Spark emerged as a powerful in-memory processing framework that supports faster data analytics, iterative machine learning tasks, and real-time stream processing. Other widely used frameworks include Apache Flink, Apache Storm, and Kafka, which are designed for high-throughput, low-latency data streaming applications.

C. Role of Analytics in Decision-Making

Big Data analytics plays a crucial role in transforming raw data into valuable knowledge. It supports descriptive, diagnostic, predictive, and prescriptive analytics, enabling organizations to understand past trends, identify root causes, forecast future outcomes, and recommend optimal actions. When combined with AI techniques, Big Data analytics becomes more intelligent, adaptive, and capable of handling complex decision-making scenarios.

III. ARTIFICIAL INTELLIGENCE TECHNIQUES FOR BIG DATA

Artificial Intelligence (AI) provides advanced computational techniques that enable intelligent analysis of large-scale and complex datasets. When applied to Big Data, AI techniques enhance the ability to discover hidden patterns, perform accurate predictions, and automate decision-making processes. This section reviews the major AI techniques widely used in Big Data analytics.[7][8][9][10]

A. Machine Learning Techniques

Machine Learning (ML) is a core component of AI that allows systems to learn from data without explicit programming. In the context of Big Data, ML algorithms are used for tasks such as classification, clustering, regression, and anomaly detection. Supervised learning techniques, including decision trees, support vector machines, and ensemble methods, are commonly applied in predictive analytics.

Unsupervised learning approaches, such as k-means clustering and hierarchical clustering, are useful for discovering data structures and patterns in unlabeled datasets. The scalability of ML algorithms has been significantly improved through distributed computing frameworks and parallel processing.

B. Deep Learning Models

Deep Learning (DL), a subset of machine learning, has gained significant attention due to its ability to automatically extract high-level features from raw data. Deep neural networks, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, have demonstrated superior performance in handling unstructured Big Data such as images, videos, and textual information.

These models benefit from large datasets and high computational power, making them well-suited for Big Data environments. Recent advancements in hardware accelerators and cloud-based platforms have further enabled the large-scale deployment of deep learning models.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Natural Language Processing and Computer Vision

Natural Language Processing (NLP) focuses on enabling machines to understand, analyze, and generate human language. In Big Data analytics, NLP techniques are widely used for text mining, sentiment analysis, document classification, and information extraction from massive textual datasets. Similarly, Computer Vision techniques analyze large-scale visual data, enabling applications such as object detection, facial recognition, and video analytics. The integration of NLP and Computer Vision with Big Data analytics has expanded the scope of intelligent applications across multiple domains.

IV. RECENT ADVANCES AND EMERGING TRENDS

Recent years have witnessed rapid advancements in the integration of Big Data analytics and Artificial Intelligence, driven by the need for scalable, efficient, and intelligent data processing systems. These advances have significantly improved the capability of organizations to analyze complex data in real time and derive actionable insights.[11][12]

A. Scalable and Distributed AI Models

One of the major advancements in this field is the development of scalable AI models capable of processing massive datasets across distributed computing environments. Techniques such as distributed machine learning, federated learning, and parallel deep learning enable model training over large-scale data without centralized storage. These approaches improve performance, reduce latency, and enhance data privacy, making them suitable for enterprise-level Big Data applications.

B. Real-Time and Streaming Analytics

The growing demand for real-time decision-making has led to significant progress in streaming analytics. Modern Big Data platforms support continuous data ingestion and real-time processing using tools such as Spark Streaming, Apache Flink, and Kafka-based pipelines. AI models integrated with streaming frameworks enable instant predictions, anomaly detection, and event-driven responses in applications such as fraud detection, network monitoring, and smart infrastructure.

C. Generative AI and Foundation Models

Generative AI and foundation models represent a recent breakthrough in AI-driven Big Data analytics. Large-scale pre-trained models, including transformer-based architectures, are capable of learning generalized representations from vast datasets and adapting to multiple downstream tasks. These models have shown strong performance in text generation, data summarization, recommendation systems, and knowledge discovery, highlighting their potential to transform data analytics workflows.

V. APPLICATIONS OF BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

The integration of Big Data analytics and Artificial Intelligence has enabled intelligent solutions across a wide range of application domains. By leveraging large-scale data and advanced AI models, organizations can improve efficiency, accuracy, and decision-making capabilities.[13][14][15]

A. Healthcare

In healthcare, Big Data and AI are used for disease diagnosis, medical image analysis, patient monitoring, and personalized treatment planning. Machine learning and deep learning models analyze electronic health records, medical images, and genomic data to support early detection of diseases and improve clinical outcomes. Real-time analytics also enables continuous patient monitoring through wearable devices and IoT-based healthcare systems.

B. Finance and Banking

The finance sector extensively uses Big Data analytics and AI for fraud detection, credit risk assessment, algorithmic trading, and customer behavior analysis. AI-driven models analyze large volumes of transactional and market data to identify suspicious activities and predict financial risks. Automated decision-making systems enhance operational efficiency while reducing human error.

C. Smart Cities and Industry

In smart city applications, Big Data and AI support traffic management, energy optimization, public safety, and environmental monitoring.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Data collected from sensors, cameras, and connected devices is analyzed in real time to improve urban services and resource utilization. Similarly, in industrial environments, AI-enabled Big Data analytics facilitates predictive maintenance, quality control, and supply chain optimization, contributing to the development of intelligent and sustainable industrial systems.

VI. CHALLENGES AND OPEN ISSUES

Despite significant advancements in Big Data analytics and Artificial Intelligence, several challenges and open research issues continue to limit their widespread adoption and effectiveness. Addressing these issues is essential for building reliable, scalable, and trustworthy intelligent systems.[16][17][18]

A. Data Quality and Availability

Big Data often originates from heterogeneous and unstructured sources, leading to issues such as missing values, noise, inconsistency, and redundancy. Poor data quality can negatively impact the performance and reliability of AI models. Additionally, limited access to high-quality, labeled datasets remains a major obstacle, particularly in sensitive domains such as healthcare and finance.

B. Scalability and Computational Cost

Processing and analyzing massive datasets require substantial computational resources, storage infrastructure, and energy consumption. Training deep learning models on Big Data is computationally expensive and time-consuming, making scalability a critical challenge. Efficient resource management and cost-effective computing solutions are still active areas of research.

C. Privacy and Security Concerns

The use of large-scale personal and sensitive data raises serious privacy and security concerns. Data breaches, unauthorized access, and misuse of information pose significant risks. Ensuring data confidentiality while maintaining analytical performance remains challenging, especially in distributed and cloud-based environments.

D. Model Interpretability and Trust

Many advanced AI models, particularly deep learning architectures, function as black boxes, making their decision-making processes difficult to interpret. The lack of transparency reduces trust and limits adoption in critical applications. Developing interpretable and explainable AI models is an important open research issue.

VII. FUTURE RESEARCH DIRECTIONS

Future research in Big Data analytics and Artificial Intelligence is expected to focus on developing more efficient, transparent, and sustainable intelligent systems. One promising direction is the design of explainable and trustworthy AI, which aims to improve model interpretability and enhance user confidence, particularly in critical domains.[19][20][21]

Another important area is edge–cloud intelligence, where data processing and AI inference are performed closer to data sources to reduce latency, bandwidth usage, and privacy risks. Additionally, hybrid AI models that combine symbolic reasoning with data-driven learning are gaining attention for their ability to improve robustness and generalization.

Research efforts are also moving toward energy-efficient and green AI, addressing the environmental impact of large-scale model training. Furthermore, the integration of AI with emerging technologies such as blockchain and digital twins offers new opportunities for secure, transparent, and intelligent Big Data applications.

VIII. CONCLUSION

This paper presented a comprehensive review of recent advances in Big Data analytics and Artificial Intelligence, highlighting key techniques, emerging trends, applications, challenges, and future research directions. The synergy between Big Data and AI has enabled intelligent data-driven solutions across diverse domains, including healthcare, finance, and smart cities.

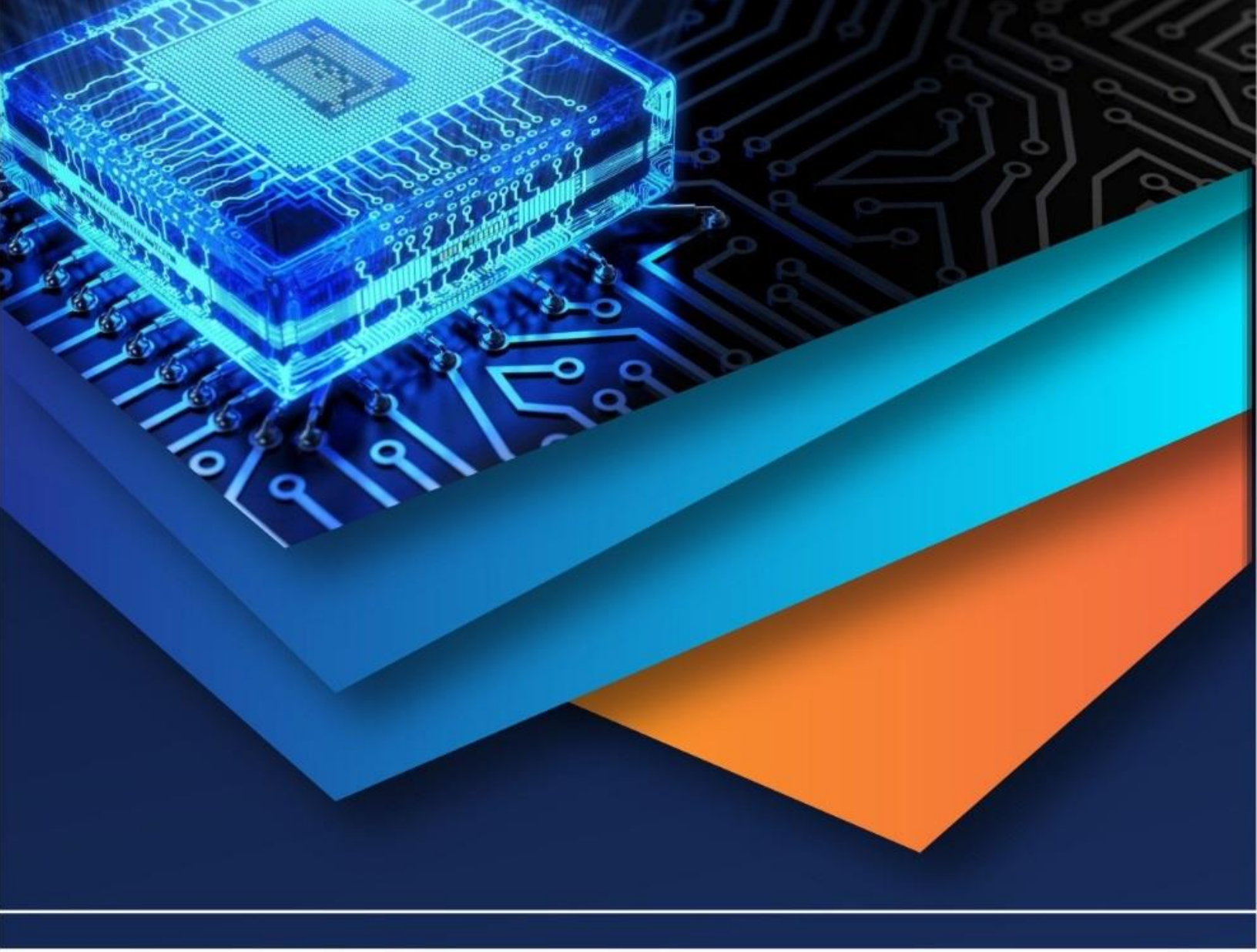
Despite notable progress, challenges related to data quality, scalability, privacy, and model interpretability remain open. Addressing these challenges is essential for realizing the full potential of AI-driven Big Data analytics. Overall, this review provides valuable insights for researchers and practitioners and serves as a foundation for future advancements in this rapidly evolving field.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” Mobile Networks and Applications, vol. 19, no. 2, pp. 171–209, 2014.
[2] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” Science, vol. 349, no. 6245, pp. 255–260, 2015.
[3] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” International Journal of Information Management, vol. 35, no. 2, pp. 137–144, 2015.
[4] P. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. New York, NY, USA: McGraw-Hill, 2012.
[5] J. Dean and S. Ghemawat, “MapReduce: Simplified data processing on large clusters,” Communications of the ACM, vol. 51, no. 1, pp. 107–113, 2008.
[6] M. Zaharia et al., “Apache Spark: A unified engine for big data processing,” Communications of the ACM, vol. 59, no. 11, pp. 56–65, 2016.
[7] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[8] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.
[9] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[10] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Pearson, 2023.
[11] P. Kairouz et al., “Advances and open problems in federated learning,” Foundations and Trends® in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.
[12] R. Bommasani et al., “On the opportunities and risks of foundation models,” arXiv preprint arXiv:2108.07258, 2021.
[13] A. Esteva et al., “A guide to deep learning in healthcare,” Nature Medicine, vol. 25, pp. 24–29, 2019.
[14] A. Bahrammirzaee, “A comparative survey of artificial intelligence applications in finance,” Journal of Financial Innovation, vol. 6, no. 3, 2020.
[15] M. Batty et al., “Smart cities of the future,” The European Physical Journal Special Topics, vol. 214, pp. 481–518, 2012.
[16] R. Kitchin, “Big data, new epistemologies and paradigm shifts,” Big Data & Society, vol. 1, no. 1, 2014.
[17] B. D. Mittelstadt et al., “The ethics of algorithms: Mapping the debate,” Big Data & Society, vol. 3, no. 2, 2016.
[18] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2016.
[19] F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” arXiv preprint arXiv:1702.08608, 2017.
[20] W. Shi et al., “Edge computing: Vision and challenges,” IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, 2016.
[21] T. H. Davenport and R. Ronanki, “Artificial intelligence for the real world,” Harvard Business Review, vol. 96, no. 1, pp. 108–116, 2018.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)