Review Article

ACCENTS Transactions on Information Security, Vol 7(27)
ISSN (Online): 2455-7196
http://dx.doi.org/10.19101/TIS.2022.725015

Unlocking the power and unveiling the challenges of big data management and analysis: insights from diverse domains

Om Prakash* **and Sandhya Gawade**
School of Computer Science and Engineering, Madhyanchal Professional University, Bhopal, India

Received: 20-June-2022; Revised: 22-July-2022; Accepted: 24-July-2022

Abstract
The multifaceted realm of big data was investigated in this paper, encompassing its management and analysis across diverse domains. The pivotal role of big data in sectors like healthcare, education, competitive intelligence, and the Internet of Things (IoT) was explored. The integration of advanced analytics, cloud-based solutions, and innovative techniques such as deep learning and artificial intelligence was examined. Moreover, potential challenges, including scalability concerns, integration complexities, publication bias, attribute relationships, and dimensionality reduction evaluations, were identified in the study. By shedding light on the advantages and limitations of big data analytics, this research contributed to a comprehensive understanding of its implications in contemporary society.

Keywords
Big data management, Data analytics, Healthcare, Education, IoT.

1.Introduction
Massive data is being generated daily from various sources, including social data, machine data, and transactional data, and is characterized by the three Vs: velocity, volume, and variety [1–3]. Understanding data quality is crucial in data-intensive domains, as data and processing techniques impact the reliability of data sources and analytics [4, 5]. Technologies such as the Internet of Things, geolocation systems, cameras, smart devices, and social media contribute to the generation of diverse data types used for decision-making, often employing computational intelligence methods like deep learning and artificial intelligence [6–8]. The management of big data has become more complex with the introduction of new strategies like data lakes, affecting industries across the board [9, 10].

The motivation behind this research stems from the increasing importance of managing and analyzing big data, which is essential in various domains, including healthcare, where patient data and medical information are abundant [11, 12]. Deep learning and artificial intelligence have found applications in the medical field, but scalability and privacy remain key concerns.

Additionally, as big data clustering and analytics become more prevalent, it's crucial to consider tools and environments for efficient processing. Cloud computing is a practical solution for big data processing, storage, and complex analytics [13, 14]. Moreover, the integration of big data and deep learning technologies is explored in the context of Internet of Things (IoT) security [15, 16]. Data envelopment analysis is used to assess decision-making units' effectiveness and can benefit from improved scalability [17–20].

This paper aims to provide insights into the management and analysis of big data, particularly in the context of healthcare and IoT security. It explores the application of deep learning, artificial intelligence, and data envelopment analysis in handling large-scale data. The research also investigates the role of cloud computing in supporting big data processing and analytics. Additionally, it discusses the potential of advanced machine learning techniques for forecasting demand and evaluating market potential based on consumer reviews.

The contribution of this research lies in its exploration of various aspects of big data management, including its application in healthcare, IoT security, and consumer goods industries. It offers

*Author for correspondence

insights into the use of deep learning and artificial intelligence, scalability considerations, and the role of cloud computing. Additionally, it highlights the potential for advanced machine learning techniques in forecasting demand and assessing market potential. This paper contributes to the understanding of the challenges and opportunities presented by big data in diverse domains. *Figure 1* shows the areas and applicability of big data. It includes healthcare, e-commerce, energy Safety, education, and agriculture. The structure of this paper is organized as follows: Section 2 discusses related work in the field, providing context for the research. Section 3 provides a detailed discussion of the findings, drawing conclusions and implications. Finally, Section 4 provides a comprehensive conclusion to the work.

Figure 1 Areas and applicability of big data

2.Literature review

The section provides overview and analyis for the tasks performed by big data and its analysis from multiple perspectives. Both IoT and Big Data are centered on data that requires innovative processing methods owing to its high volume, velocity, and diversity.

In 2019, Fu et al. [16] introduced an optimized system for scheduling large-scale data in social internet communities. Their scheduling model for big data considered data communication volume for each task during goal data transmission. They efficiently addressed data collision challenges and improved the scheduling of social network data.

In 2019, Olivera et al. [17] explored the significant potential applications of big data in inflammatory bowel disease, encompassing cost-effectiveness of care, disease heterogeneity characterization, drug safety and development, and illness progression and therapy response prediction models. They reviewed various data sources, including clinical trial data, electronic health records, e-health apps, and diverse omics data, for studying inflammatory bowel disease. They noted variations in data quality among different sources, with some sources more prone to poor quality.

In 2020, Sestino et al. [18] investigated how companies manage digital transformation with a focus on the roles of Big Data and the Internet of Things (IoT). They conducted a literature survey spanning the last ten years (2008–2019) to identify 204 relevant articles. Their findings indicated that IoT and Big Data primarily drive the reengineering of corporate processes, products, and services. However, they acknowledged the possibility of missing relevant documents in other languages, as their study focused on English-language articles.
In 2020, Qi et al. [19] aimed to elucidate the fundamentals of big data management, emphasizing its importance in the mining industry. They provided an overview of big data and management, highlighting challenges faced by the mining sector. They also discussed data sources in the mining industry and how big data can benefit the sector.

Their research projected a future where big data is integrated with technologies like automation, the Internet of Things, and augmented reality to create a global database project, facilitating its broader application.

In 2020, Bragazzi et al. [20] presented a review outlining potential applications of Big Data and artificial intelligence in global pandemic management efforts.

In 2020, Hasan et al. [21] underscored the impact of big data on financial services, financial markets, and online banking. They provided a theoretical framework to comprehend big data's effects on finance, drawing insights from 180 related articles. They observed that big data has transformed the financial sector, enabling real-time stock market insights, enhancing fraud detection and prevention, and enabling precise risk analysis through machine learning. Future research directions may focus on facilitating small businesses' access to managing large datasets.

In 2020, Baig et al. [22] conducted a systematic review of 40 primary papers published between 2014 and 2019 to examine the growing interest in big data in education. They identified four key research areas: integrating big data into the curriculum, studying learner behavior and performance, developing models and educational data warehouses, and enhancing the educational system. Most studies in big data education focused on student actions and outcomes. They proposed a framework for further research and highlighted new perspectives for effective big data applications in education.

In 2020, Agrawal and Prabakaran [23] explored big data's future role in digital healthcare. They balanced its potential benefits for patient outcomes with growing physician concerns, examining perspectives from the US, UK, and other nations. They emphasized the need for healthcare data regulations, using oncology as an illustrative example. Proposed guidelines included establishing a unique global patient ID.

In 2021, Leon et al. [24] recommended a Big Data architecture designed to address scalability and adaptability requirements. They highlighted the effectiveness of this architecture in improving genetic identification, particularly in challenging conditions like epilepsy, and its ongoing evaluation in pediatric oncology and sudden cardiac death cases.

In 2021, Nilashi et al. [25] proposed a two-stage approach framework to study the correlation between customer satisfaction and eco-friendly hotel performance. They collected data from TripAdvisor, considering reviews and complaints about local cuisine quality. After data preprocessing, they conducted segmentation using learning vector quantization and employed Partial Least Squares Structural Equation Modeling. Their approach achieved high AUC scores, ranging from 0.94 to 0.98.

In 2021, Nikolov et al. [26] presented a Big Data workflow method based on message-oriented middleware, software container technologies, and a domain-specific language to enable scalable process execution and workflow specification. They compared the scalability of their method to Argo Workflows and emphasized the need to address centralized message-oriented communication limitations.

In 2021, Arfanuzzaman et al. [27] assessed the potential for expanding big data and artificial intelligence initiatives in South Asian cities to tackle urban development challenges. They examined recent advances in these technologies for sustainable development objectives in South Asia, emphasizing the potential for transformative change. Their recommendations included infrastructure development, regional research collaboration, technological readiness enhancement, and the removal of inhibiting conditions to fully leverage these technologies in South Asia.

In 2022, Rohini et al. [28] explored the wireless communication, a dynamic field that has witnessed explosive growth due to commercial network operators and the advent of big data. They introduced a high-level overview of big data applications in next-gen wireless networks, emphasizing machine learning's role in deciphering vast data. The study aimed to enhance community wireless channels by predicting mobile users' demands through advanced analytics, ultimately improving wireless network efficiency and performance.

In 2023, Lutfi et al. [29] investigated Big Data Analytics (BDA) adoption in Jordan's retail industry. Their integrated model, based on technology-organization-environment and resource-based view theories, revealed significant influences of factors like relative advantage, readiness, top management and government support, data variety, and velocity on

BDA adoption. Moreover, the study established a significant link between BDA adoption and firm performance, offering insights for improving BDA adoption in developing nations.

In 2023, Safa et al. [30] introduced the Health Care Big Data Analytics Model (HCBDA) to efficiently manage massive patient data in healthcare. HCBDA monitors patients' health conditions, uses sensor nodes and IoT devices for data collection, and employs advanced algorithms for disease prediction. The approach significantly enhances disease prediction accuracy, reaching up to 96%.

This section provides a comprehensive overview of various studies encompassing big data and its applications, particularly in diverse domains like healthcare, finance, education, and wireless communication. These studies highlight the importance of data quality, innovative processing techniques, and the integration of advanced analytics methods. They shed light on both the potential advantages, such as improved disease prediction and enhanced decision-making, and the challenges, including scalability and data quality variations, associated with big data utilization across different sectors.

3.Discussion and analysis

Based on an extensive review of the literature and a thorough analysis of the findings, several notable benefits and limitations pertaining to big data and analytics have been identified. The major advantages are discussed below.

The integration of big data and analytics can yield a robust competitive intelligence analytics system. By harnessing advanced analytics techniques, this system can significantly improve the accuracy and effectiveness of traditional competitive intelligence methods. Companies can gain a competitive edge by leveraging insights derived from extensive data analysis, allowing them to make informed strategic decisions.

Investing in cloud-based analytics solutions can yield substantial returns for businesses. Cloud platforms offer scalability, flexibility, and cost-efficiency, enabling organizations to process and analyze vast amounts of data without the need for significant on-premises infrastructure. This approach not only reduces operational costs but also accelerates data-driven decision-making.

Big data management plays a pivotal role in the education sector, particularly in the context of data mining. These solutions are increasingly essential for the effective management of educational processes and resources. Leveraging big data analytics in education can lead to improved learning outcomes, personalized education experiences, and data-driven decision support for educators and administrators.

Big data proves invaluable in providing visualizations of pandemic-related information. This includes risk assessments, breakdowns of risk factors, tracking of confirmed cases, monitoring preventive measures, and more. Utilizing big data analytics in the context of pandemics enhances situational awareness and supports evidence-based decision-making for healthcare authorities and the public.

The major limitations are as follows:
As new components and technologies emerge in the field of big data analytics, the process of integrating them into existing frameworks must be streamlined. The rapid pace of innovation can make it challenging for organizations to keep their analytics ecosystems up-to-date and fully compatible with novel tools and techniques.

Studies included in the literature review may not be representative of all research conducted in the field, and this bias could affect the generalizability of certain findings. Researchers and analysts should remain vigilant in addressing this potential source of bias.

Establishing clear and meaningful relationships between different attributes or characteristics within big datasets can be a complex task. Understanding how various data elements interact and influence outcomes is critical for accurate analysis. Failing to discern these relationships can lead to erroneous conclusions and ineffective decision-making.

High-dimensional data, such as images and text, present unique challenges in terms of dimensionality reduction. While various techniques exist for reducing dimensionality and enhancing the interpretability of data, evaluating their effectiveness with different types of high-dimensional data remains an ongoing research area. Identifying the most suitable dimensionality reduction methods for specific applications is crucial for meaningful analysis.

The amalgamation of big data and analytics offers substantial benefits across various domains, from competitive intelligence and education to pandemic response. However, it is imperative to recognize and address the associated limitations, such as scalability constraints, the need for streamlined integration, potential publication bias, attribute relationships, and the evaluation of dimensionality reduction techniques. These considerations are essential for maximizing the potential of big data analytics while mitigating challenges and risks.

4.Conclusion

Big data continues to shape and revolutionize various domains, offering transformative benefits across industries. It has the potential to enhance decision-making, improve educational outcomes, bolster competitive intelligence, and enable data-driven pandemic responses. However, harnessing these benefits requires overcoming challenges such as scalability constraints, integration complexities, and issues related to data quality and analysis. To fully leverage the potential of big data, organizations and researchers must remain vigilant in addressing these obstacles and stay attuned to evolving technologies and methodologies. Big data's journey is far from over, promising continued innovation and exploration in the realms of data management and analytics.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Li W, Chai Y, Khan F, Jan SR, Verma S, Menon VG, et al. A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. Mobile Networks and Applications. 2021; 26:234-52.

[2] Zhan K. Sports and health big data system based on 5G network and internet of things system. Microprocessors and Microsystems. 2021; 80:103363.

[3] Aceto G, Persico V, Pescapé A. Industry 4.0 and health: internet of things, big data, and cloud computing for healthcare 4.0. Journal of Industrial Information Integration. 2020; 18:100129.

[4] Karthik SA, Hemalatha R, Aruna R, Deivakani M, Reddy RV, Boopathi S. Study on healthcare security system-integrated internet of things (IoT). In perspectives and considerations on the evolution of smart systems 2023 (pp. 342-62). IGI Global.

[5] Sarosh P, Parah SA, Bhat GM, Muhammad K. A security management framework for big data in smart healthcare. Big Data Research. 2021; 25:100225.

[6] Lv Z, Qiao L. Analysis of healthcare big data. Future Generation Computer Systems. 2020; 109:103-10.

[7] Selvaraj S, Sundaravaradhan S. Challenges and opportunities in IoT healthcare systems: a systematic review. SN Applied Sciences. 2020; 2(1):139.

[8] Hariri RH, Fredericks EM, Bowers KM. Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data. 2019; 6(1):1-6.

[9] Bag S, Wood LC, Xu L, Dhamija P, Kayikci Y. Big data analytics as an operational excellence approach to enhance sustainable supply chain performance. Resources, Conservation and Recycling. 2020; 153:104559.

[10] Dubey AK, Kushwaha GR, Shrivastava N. Heterogeneous data mining environment based on dam for mobile computing environments. In international conference on advances in information technology and mobile communication 2011 (pp. 144-9). Berlin, Heidelberg: Springer Berlin Heidelberg.

[11] Duan Y, Edwards JS, Dwivedi YK. Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. International Journal of Information Management. 2019; 48:63-71.

[12] Kumari S, Patidar K, Kushwah R, Saxena G. A survey and analysis of big data management based on computational methodologies. ACCENTS Transactions on Image Processing and Computer Vision. 2020; 6(20):48-53.

[13] Didas M. The barriers and prospects related to big data analytics implementation in public institutions: a systematic review analysis. International Journal of Advanced Computer Research. 2023; 13(64):29-54.

[14] Hussin SK, Omar YM, Abdelmageid SM, Marie MI. Traditional machine learning and big data analytics in virtual screening: a comparative study. International Journal of Advanced Computer Research. 2020; 10(47):72-88.

[15] Dubey AK, Kumar A, Agrawal R. An efficient ACO-PSO-based framework for data classification and preprocessing in big data. Evolutionary Intelligence. 2021; 14:909-22.

[16] Fu W, Liu S, Srivastava G. Optimization of big data scheduling in social networks. Entropy. 2019; 21(9):902.

[17] Olivera P, Danese S, Jay N, Natoli G, Peyrin-Biroulet L. Big data in IBD: a look into the future. Nature Reviews Gastroenterology & Hepatology. 2019; 16(5):312-21.

[18] Sestino A, Prete MI, Piper L, Guido G. Internet of things and big data as enablers for business digitalization strategies. Technovation. 2020; 98:102173.

[19] Qi CC. Big data management in the mining industry. International Journal of Minerals, Metallurgy and Materials. 2020; 27:131-9.

[20] Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How big data and artificial intelligence can help better manage the COVID-19

pandemic. International Journal of Environmental Research and Public Health. 2020; 17(9):3176.
[21] Hasan MM, Popp J, Oláh J. Current landscape and influence of big data on finance. Journal of Big Data. 2020; 7(1):1-7.
[22] Baig MI, Shuib L, Yadegaridehkordi E. Big data in education: a state of the art, limitations, and future research directions. International Journal of Educational Technology in Higher Education. 2020; 17(1):1-23.
[23] Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. Heredity. 2020; 124(4):525-34.
[24] León A, Pastor O. Enhancing precision medicine: a big data-driven approach for the management of genomic data. Big Data Research. 2021; 26:100253.
[25] Nilashi M, Abumalloh RA, Almulihi A, Alrizq M, Alghamdi A, Ismail MY, et al. Big social data analysis for impact of food quality on travelers' satisfaction in eco-friendly hotels. ICT Express. 2021.
[26] Nikolov N, Dessalk YD, Khan AQ, Soylu A, Matskin M, Payberah AH, et al. Conceptualization and scalable execution of big data workflows using domain-specific languages and software containers. Internet of Things. 2021; 16:100440.
[27] Arfanuzzaman M. Harnessing artificial intelligence and big data for SDGs and prosperous urban future in South Asia. Environmental and Sustainability Indicators. 2021; 11:100127.
[28] Rohini P, Tripathi S, Preeti CM, Renuka A, Gonzales JL, Gangodkar D. A study on the adoption of wireless communication in big data analytics using neural networks and deep learning. In 2nd international conference on advance computing and innovative technologies in engineering 2022 (pp. 1071-6). IEEE.
[29] Lutfi A, Alrawad M, Alsyouf A, Almaiah MA, Al-Khasawneh A, Al-Khasawneh AL, et al. Drivers and impact of big data analytic adoption in the retail industry: a quantitative investigation applying structural equation modeling. Journal of Retailing and Consumer Services. 2023; 70:103129.
[30] Safa M, Pandian A, Gururaj HL, Ravi V, Krichen M. Real time health care big data analytics model for improved QoS in cardiac disease prediction with IoT devices. Health and Technology. 2023; 13(3):473-83.

Om Prakash is currently pursuing an M.Tech in Computer Science & Engineering at Madhyanchal Professional University in Bhopal, Madhya Pradesh, India. He holds a B.Tech degree in Computer Science & Engineering from Pune University in Pune, Maharashtra. His primary area of interest is Network Security.
Email: omsinghssm1@gmail.com

Sandhya Gawade is currently serving as an Assistant Professor in the Department of Computer Science and Engineering at Madhyanchal Professional University, situated in Bhopal, India. She possesses both a Bachelor of Engineering (B.E.) and a Master of Engineering degree in Computer Science and Engineering from Rajiv Gandhi Technical University in Bhopal, Madhya Pradesh. Her research interests encompass the fields of Data Mining, Optimization, and Machine Learning.
Email: sandhya.gawhade@gmail.com