RESEARCH

Big data applications: overview, challenges and future

Afzal Badshah[1] · Ali Daud[2] · Riad Alharbey[3] · Ameen Banjar[3] · Amal Bukhari[3] ·
Bader Alshemaimri[4]

Abstract

Big Data (i.e., social big data, vehicular big data, healthcare big data etc) points to massive and complex data, that require special technologies and approaches for storage, processing, and analysis. Similarly, big data applications are software and systems utilizing large and complex datasets to extract insights, support decision-making, and address diverse business and societal challenges. Recently, the significance of big data applications has grown immensely for organizations across diverse sectors as they increasingly rely on insights derived from data. The increasing reliance on data insights has rendered traditional technologies and platforms inefficient due to scalability limitations and performance issues. This study contributes by identifying key domains impacted by big data, examining its effect on decision-making, addressing inherent complexities and opportunities, exploring core technologies, and offering solutions for potential concerns. Additionally, it conducts a comparative analysis to demonstrate the superiority of this research. These contributions provide valuable insights into the evolving landscape shaped by big data applications.

Keywords Big data · Big data applications · Data analytics

1 Introduction

In the present digital era, big data has emerged as a transformative force, transforming how organizations collect, store, and analyze extensive datasets. The significant increase in data produced from various origins such as social media, autonomous vehicles, and sensors has made Big Data Analytics (BDA) essential for businesses and industries globally. Big data, as defined by the 3 Vs, encompasses large volumes of diverse and rapidly arriving data with potential uncertainties about its quality and availability (Laney 2001). The three Vs

✉ Ali Daud
  alimsdb@gmail.com

1   Department of Software Engineering, University of Sargodha, Sargodha, Pakistan

2   Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates

3   Department of Information Systems and Technology, College of Computer Science
    and Engineering, University of Jeddah, Jeddah, Saudi Arabia

4   Software Engineering Department, College of Computing and Information Sciences, King Saud
    University, Riyadh, Saudi Arabia

comprise volume (referring to large datasets), variety (involving diverse data formats), and velocity (indicating the rapid generation of data) (Badshah et al. 2024).

The Big Data Ecosystem comprises six key tools essential for efficient large-scale data management (shown in Fig. 1). *Data Technologies*, including Apache Hadoop and Apache Spark, analyze and process Big Data beyond traditional capabilities. *Analytics and Visualization tools*, such as Tableau and SAS, uncover patterns, while *Business Intelligence tools* like Cognos transform raw data for business analysis. *Cloud Service Providers*, like AWS and GCP, offer fundamental infrastructure. *NoSQL Databases*, including MongoDB and Cassandra, handle Big Data processing, and *Programming Tools* like R and Python perform analytical tasks and operationalize Big Data, completing this vital ecosystem (Coursera 2023).

The applications of big data are diverse and far-reaching, spanning healthcare, supply chain and logistics, marketing and advertising, smart cities, media and entertainment, cybersecurity, climate & earth science, industry, and education. The primary objective of big data lies in its analysis for diverse purposes. Harnessing the capabilities of BDA enables organizations to discover important insights, recognize patterns, and make informed, data-driven decisions. These decisions, in turn, enhance operational efficiency, drive innovation, and improve customer experiences. From personalized healthcare treatments to predictive maintenance in manufacturing, big data is transforming industries and shaping the future of how we live and work (Himeur et al. 2023; Talaoui et al. 2023).

In the current technological research landscape, big data plays a pivotal role, focusing on the analysis, processing, and extraction of valuable information from extensive and intricate datasets. The foundation of BDA is intricately linked with advanced technologies, specifically Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and the Internet of Things (IoT). Figure 2 illustrates the processing stages of big data. For a better understanding of this article, Table 1 outlines the terminologies used in the study.

The big data market is expected to have remarkable growth globally, with revenue projections ranging to USD 473.6 Billion by 2030, reflecting a growth rate of 12.7% from 2022 to 2030 (Research and Consulting 2023). This substantial growth underscores the increasing recognition of big data's critical role across industries and sectors. Simultaneously, current estimates indicate a massive increase in data generation, with the world expected to produce 175 zettabytes of data by 2025 (Statista 2023), as shown in Fig. 3. This exponential increase highlights the expanding scope and importance of big data as a critical tool for managing, analyzing, and deriving insights from this colossal volume of information.
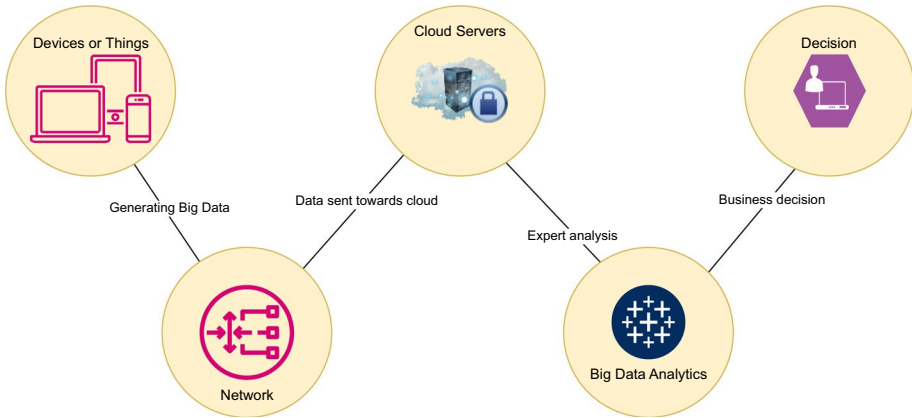
Fig. 1 Big data ecosystem
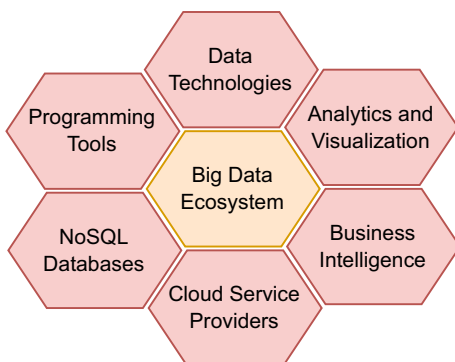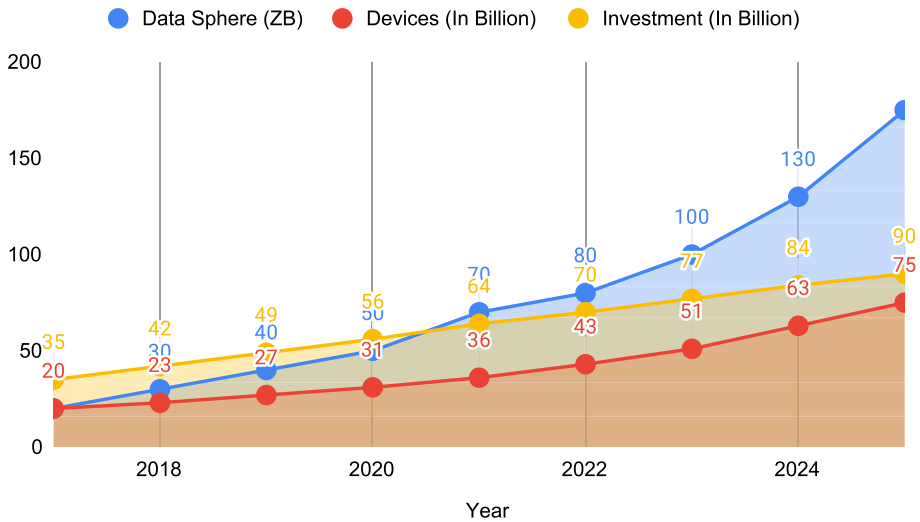
Fig. 2 Structure of the big data cycle

Fig. 3 Devices, data and revenue forecast from 2017 to 2025

Table 1 Terminologies employed in the study

Terminology	Definition	Terminology	Definition
SBD	Social big data	VBD	Vehicular big data
EBD	Earth big data	EdBD	Educational big data
HBD	Healthcare big data	IBD	Industrial big data
UBD	Urban big data	BDA	Big data analytics
AI	Artificial intelligence	ML	Machine learning
DL	Deep learning	IoT	Internet of things
IoE	Internet of everything	ICT	Information and communication technologies
GIS	Geographic information system	ITS	Intelligent transportation system

The massive utilization of big data is propelled by the exponential surge in data volume, the extensive utilization of cloud computing, global digital transformation, increasing internet and smartphone usage, and accelerated adoption due to the impact of the COVID-19 pandemic. Leading companies, such as Google, Amazon, and other tech giants play a crucial role in the big data ecosystem, significantly contributing to the development and advancement of big data technologies, influencing trends, and shaping the future trajectory of this dynamic field.

Numerous comprehensive literature survey papers have extensively explored big data applications. Focusing on healthcare, (Hong et al. 2018; Abouelmehdi et al. 2018; Rajabion et al. 2019; Galetsi et al. 2019) conducted a thorough review of big data's impact in the healthcare sector, often termed Healthcare Big Data (HBD) (HBD). Vehicular Big Data (VBD), referring to big data in vehicles, has received significant attention with comprehensive reviews by researchers in Nguyen et al. (2018), Torre-Bastida et al. (2018), Ghofrani et al. (2018), Mishra et al. (2018). Concurrently, Urban Big Data (UBD), associated with smart cities, has been deeply explored by authors in Allam and Dhunny (2019), Karimi et al. (2021), Mohammadi and Al (2018), Huang et al. (2021). Exploring the intersection of big data and cybersecurity, Alani (2021), Ullah and Babar (2019), Srivastava and Jaiswal (2019) provide a comprehensive review. The industrial sector, often referred to as Industrial Big Data (IBD), underwent scrutiny in Qi (2020), Misra et al. (2020), Mosavi et al. (2018), while the education sector, explored under the umbrella of big data, is thoroughly reviewed in Luan et al. (2020), Baig et al. (2020), Li and Jiang (2021). Notably, authors in Akter and Wamba (2019), Amani et al. (2020), Huang et al. (2018), and Akter and Wamba (2019) have extensively explored the utilization of big data in earth sciences and disaster management, usually referred to as Earth Big Data (EBD). This collective exploration paints a comprehensive picture of the diverse applications and impacts of big data across various domains.

After an in-depth analysis of the available literature, it becomes apparent that individual literature reviews have been conducted across various domains, such as big data in healthcare, vehicles, finance, agriculture, education, etc. However, a wide gap exists in the collective analysis of big data applications. In bridging this gap, it is crucial to undertake a comprehensive assessment of how big data substantially contributes to diverse fields, discern the challenges it presents, delve into ethical concerns, and illuminate emerging applications. Therefore, this article aims to make the following contributions:

- This research systematically identifies and analyzes key domains profoundly influenced by big data applications, providing a comprehensive understanding through the exploration of prominent use cases.
- The study examines the transformation of decision-making processes in these domains due to big data, emphasizing how data-driven insights contribute to informed decision-making and enrich the existing knowledge on the subject.
- This research addresses limitations and potentials in diverse fields' big data applications, emphasizing inherent complexities and opportunities.
- This research delves into the core technologies employed for storing, processing, and analyzing large datasets, elucidating their significance in big data applications.
- This research systematically identifies and addresses potential concerns within big data, offering viable solutions and mitigation strategies.
- The study conducts a comparative analysis of this research survey with related surveys to demonstrate the unique contributions and superiority of this study.

The subsequent sections encompass the following: In Sect. 2, the research methodology employed for conducting this study is explained, covering details on the research string, as well as the inclusion and exclusion criteria. Section 3 undertakes the classification of literature concerning big data applications and conducts a thorough analysis of this body of work. Section 10 explored the technologies used to process and store big data. Exploring concerns associated with the utilization of big data in various domains, Sect. 5 delves into potential concerns with big data and introduces potential solutions that can be applied to address the aforementioned issues discussed earlier. Section 6 carried out a detailed comparison of this study with the related literature to show the uniqueness of this paper. Finally, section 7 concludes the study.

2 Research methodology

For the big data applications, we exclusively considered articles published from 2018 to 2023. Our search encompassed Google Scholar, Scopus, IEEE Xplore, and Science Direct to identify pertinent papers. Google Scholar offers access to papers published in any journal, whereas research libraries provide access to a more limited but high-quality selection of papers published in affiliated journals and by specific publishers.

In exploring the electronic world, a pivotal element is the search string, defining the search's quality. The search string incorporates keywords that encapsulate the population, methodology, and outcomes. The methodology of this research paper is organized into three stages: (i) the Planning phase, (ii) the Conducting phase, and (iii) the Reporting phase. The following section of this part examines these phases.

2.1 Planning the review

In the initial stage, we precisely designed the review's framework, which encompassed the formulation of the study protocol, identification of relevant journals and research papers, the establishment of including and excluding criteria, and defining our reporting strategy. The planning phase serves two fundamental purposes: (i) emphasizing the significance and necessity of this study, setting it apart from similar research endeavours; and (ii) formulating a robust protocol for conducting a comprehensive search of relevant studies while establishing clear criteria for their inclusion and exclusion.

Creating a well-defined review protocol is of utmost importance. A suitable protocol guides us towards a comprehensive review, while an invalid one may divert authors from the main focus. Therefore, this stage encompasses the examination and determination of developing research queries, exploration approaches, and criteria for selection.

2.2 Conducting the review

During this stage, the research is executed following the protocol delineated in Phase 1. The primary emphasis lies in identifying pertinent research studies, and subjecting them to scrutiny based on three pivotal criteria.

2.2.1 Population

To start, we define the population of the study. For instance, this research encompasses various domains impacted by big data applications, so the population of this review includes diverse fields where big data plays a significant role. The primary keyword is 'Big Data'. This overarching term encompasses various facets, including 'Earth Big Data (EBD)', 'Vehicular Big Data (VBD)', 'Healthcare Big Data (HBD)', 'Urban Big Data (UBD)', 'Industrial Big Data (IBD)', and 'Education Big Data (EdBD)'. The inclusion of these synonyms ensures a comprehensive exploration of diverse fields impacted by big data applications.

2.2.2 Methodology or technique

The second criterion involves the methodology or technique employed to achieve the intended outcomes. In this context, big data applications serve as the fundamental technique for obtaining desired results across different domains. The methodology or technique employed in the research is centred around the domain of Data Processing Techniques. The keywords utilized for this area include 'Data Analytics', 'Machine Learning', and 'Artificial Intelligence'. These keywords are chosen to encapsulate the fundamental techniques employed to achieve intended outcomes across different domains.

2.2.3 Outcome

The third and final check revolves around the outcomes achieved in each research study. In the case of this review, the outcomes pertain to the applications and impacts of big data within the specified domains. The keywords employed in this context encompass 'Healthcare', 'Supply-chain', 'Logistics', 'Marketing', 'Advertisement', 'Smart Cities', 'Media', 'Cybersecurity', 'Climate', 'Industry', and 'Education'. These keywords are strategically chosen to align with specific domains, ensuring a focused investigation into the applications and impacts of big data within each specified area.

Therefore, the following research statement is used to search the paper on different platforms. Table 2 shows the details of the research string.

(Big Data) AND (Data Analytics OR Machine Learning OR Artificial Intelligence) AND (Healthcare OR Supply Chain OR Transport OR Marketing OR Advertisement OR Smart Cities OR Social Media OR Climate OR Earth Science OR Industry)

Table 2 The search string employed to explore the relevant literature

Area	Keywords	Synonyms in literature
Population	Big data	Big Data OR Earth Big Data OR Vehicular Big Data OR Healthcare Big Data OR Urban Big Data OR Industrial Big Data OR Education Big Data
Methodology	Data processing techniques	Data Analytic OR Machine Learning OR Artificial Intelligence
Outcomes	Big data applications	Healthcare OR Supply-chain OR Logistics OR Marketing OR Advertisement OR Smart Cities OR Media OR Cybersecurity OR Climate OR Industry OR Education

Pursuing these criteria, the subsequent critical step involves formulating research questions. In the context of this review, the investigation questions are as follows:

(1) Which key domains are profoundly influenced by big data applications, and what are the prominent use cases within these domains?
(2) How has big data transformed decision-making processes in these key domains, and how do data-driven insights contribute to informed decision-making?
(3) What are the limitations and potentials associated with big data applications in diverse fields, and what inherent complexities and opportunities do they present?
(4) What core technologies are employed for storing, processing, and analyzing large datasets in big data applications, and what is their significance?
(5) What potential concerns exist within big data applications, and what viable solutions and mitigation strategies can address these concerns?

These research questions guide the review process, providing a structured framework for analyzing the selected research studies and synthesizing their findings. The questions address various aspects of big data applications, from their impact on different domains to the challenges, ethical considerations, and emerging trends associated with their use.

2.3 Quality assessment

The evaluation of this study's quality depends on various essential parameters. To ensure the robustness and relevance of the papers included in this review, the following parameters have been established.
    Inclusion criteria:

(1) Relevant to big data applications: Selected papers must address topics related to big data applications, ensuring the content aligns with the central theme of this review.
(2) methodology and results presentation: Included research articles should present their methodology and results in a clear and organized manner, enhancing the comprehensibility of their findings.
(3) citation threshold: Research articles under consideration must have a minimum of 10 citations, reflecting their impact and recognition within the academic community. However, to include the recent articles, this count is reduced to 5 for 2023 publications.
(4) Publication year: Selected research articles should have been published since 2018 to ensure the incorporation of recent developments in the field.

Exclusion criteria:

(1) Irrelevance to big data role: Research papers that do not discuss the position of big data in any domain will be excluded from the review, as they fall outside the scope of this study.
(2) Inadequate presentation of results and methodology: Papers that do not adequately present the results and methodology used to achieve desired outcomes will be excluded to maintain the quality and rigour of the review.
(3) Insufficient citations: Research papers that have failed to garner at least 10 citations may be excluded to ensure the inclusion of well-recognized and influential works.

(4) Publication year: *Research articles not published between 2018 and 2023 will be excluded to focus on more recent and relevant literature.*

2.4 Reporting the review

In the final stage, this study extracts and presents papers that are relevant to the keywords and research questions. The impact of the review depends on how the final assessment is presented in the paper.

We conducted a Google search that yielded a total of 22,080 results across various categories, including; Conferences (17,582), Journals (3485), Books (443), Magazines (328), Early Access Articles (228), Standards (12) and Courses (2).

We used the above-mentioned inclusion and exclusion criteria to filter the content. After careful consideration, we selected only 125 papers which were fulfilling the inclusion criteria.

Table 3 shows the number of publications yearwise and categorywise from 2018 to 2023, while Table 13 displays the category-wise publications since 2018.

3 Big data application classification

Big data applications have become ubiquitous, establishing as a fundamental technology with a pervasive role across various fields, just like to computers. This section systematically categorizes these applications, delving into their impact, challenges, and prospects. The primary domains of big data applications include Healthcare, Supply Chain and Transport, Market and Transport, Smart Cities, Media, Cyber Security, Earth Science, Industry, Education, and others. Figure 4 visually presents the classification of these diverse big data applications.

3.1 Big data in healthcare

Big data plays a central role in Health 4.0, reshaping healthcare through data-driven research. Analyzing biomedical omics and clinical data offers both challenges and opportunities for healthcare improvement (Ahmed et al. 2023). The healthcare industry generates vast data, including hospital records, medical exams, and research, necessitating proper management for meaningful insights (Philip et al. 2022). Healthcare's BDA can enable personalized medicine, clinical risk management, and forecasting, alongside standardizing medical terminology and patient registration (Tohka and Van Gils 2021; Masood et al. 2018b). Table 4 summarizes the big data applications and Fig. 5 shows the HBD categorization.

Integrating biomedical and healthcare data empowers modern organizations to revolutionize medical therapies and personalize treatment (Mehta and Pandit 2018). Big data and e-business complement modern hospital management, transforming fragmented systems into comprehensive, omnidirectional healthcare management (Dash et al. 2019). Health analytics using big data aids in developing effective medical policies, improving healthcare services, and enhancing disease prediction, drug recommendations, and treatment outcomes (Zhang et al. 2023).

A robust BDA platform at Xiangya Hospital's Gastroenterology Department, China, facilitates comprehensive digestive medicine analysis. The platform combines electronic

Table 3 A comprehensive summary of publications organized by year and category since 2018

| Category	Subcategory	2018	2019	2020	2021	2022	2023	Total
Health care	Hospitals, telehealth, pharmaceutical research	4	5	3	2	3	2	19
Supply chain and logistic	Transportation management, electric vehicles, autonomous drones	4	5	1	2	2	1	15
Marketing and advertising	Digital marketing, customer analytics, ad personalization	3	1	1	1	3	1	10
Smart cities	Urban planning, IoT infrastructure, sustainable technologies	6	2	1	2	1	1	13
Media and entertainment	Streaming services, content recommendation, user engagement	2	4	1	2	2	1	12
Cybersecurity	Network security, threat detection, encryption technologies	2	5	2	2	1	1	13
Climate & earth science	Climate modeling, environmental monitoring, geospatial analysis	5	3	4	1	0	1	14
Industry	Manufacturing processes, automation, quality control	4	1	4	2	2	1	14
Education	E-learning platforms, learning analytics, student performance tracking	4	1	5	2	2	1	15
Total		34	27	22	16	16	10	125
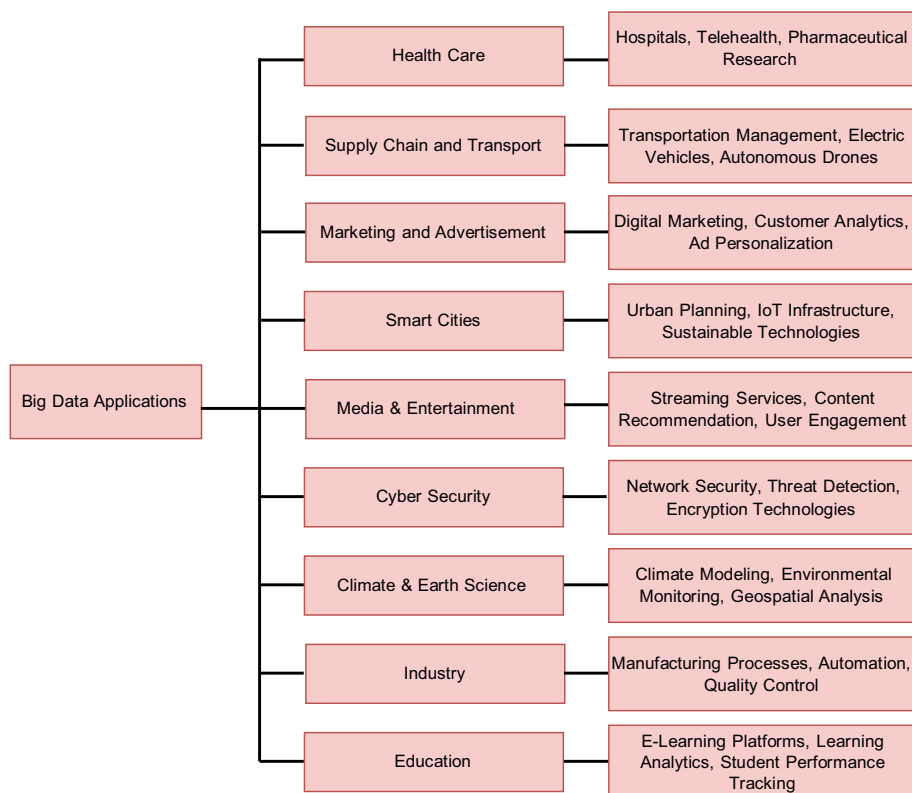
**Fig. 4** Big data applications classification

medical records and colonoscopy data, offering insights for optimal colorectal cancer screening ages and improving healthcare management (Yan et al. 2019). Leveraging prescription big data can enhance dosage prediction in pediatric medication. Traditional clinical decision support systems often lack accurate pediatric data. In Wu et al. (2019), the authors propose a data-driven approach for precise pediatric medication dosage predictions. Authors in Zhou et al. (2021) introduce a track-able patient health data search system for smart city hospital management, ensuring data privacy and efficient analysis. In Makkie et al. (2018), authors discuss challenges in analyzing MRI big data and introduce a distributed computing platform using Hadoop and Spark for fMRI data processing.

In telemedicine, an innovative big-data visualization methodology is proposed (Galletta et al. 2018). This graphical tool allows remote monitoring of patient health using coloured circles to represent various health data, adhering to the geoJSON standard for data classification. Additionally, authors in Hong et al. (2019) suggest a medical-history-based algorithm for predicting potential diseases accurately. This algorithm utilizes HBD and DL technology, providing references for targeted medical examinations and reducing delays in treatment due to unclear symptoms or limited professional knowledge. Similarly, authors in Yadav and Jadhav (2019) employ medical big data in disease recognition.

While health big data is vital for disease detection, migrating to the cloud faces challenges like data standards and sensitivity. Authors endorse a cloud-native healthcare data

Table 4 Detail summary of big data applications in healthcare

Paper & year	Main contributions	Limitation
Yan et al. (2019)	Comprehensive digestive medicine analysis at Xiangya Hospital's Gastroenterology Department, China	Limited insights into optimal colorectal cancer screening ages from the combined electronic medical records and colonoscopy data
Wu et al. (2019)	Proposing a data-driven approach for precise pediatric medication dosage predictions	The challenge of accurate pediatric data in traditional clinical decision support systems
Zhou et al. (2021)	Introduction of a traceable system for searching patient health data in smart city hospital management	Challenges related to data privacy and the need for efficient analysis
Makkie et al. (2018)	Discussion on challenges in analyzing MRI big data and introduction of a distributed computing platform using Hadoop and Spark	The challenges involved in processing MRI big data and the potential of distributed computing
Galletta et al. (2018)	Proposing an innovative big-data visualization methodology in telemedicine	Limited insights into remote patient health monitoring using coloured circles for data representation
Hong et al. (2019)	Introduction of a medical-history-based algorithm for predicting potential diseases	Challenges in reducing delays in treatment due to unclear symptoms
Yadav and Jadhav (2019)	Employing medical big data in disease recognition	Limitations in accurately predicting potential diseases
Wu et al. (2019)	Endorsing a cloud-native healthcare data ingestion service to address challenges in migrating to the cloud	Challenges and best practices in migrating HBD to the cloud
Zhou et al. (2020)	Presenting a scalable system for storing and analyzing healthcare data from IoT devices	Limited insights into secure storage and analysis of healthcare data from IoT devices using big data systems and blockchain architecture

Fig. 5 Big data in healthcare
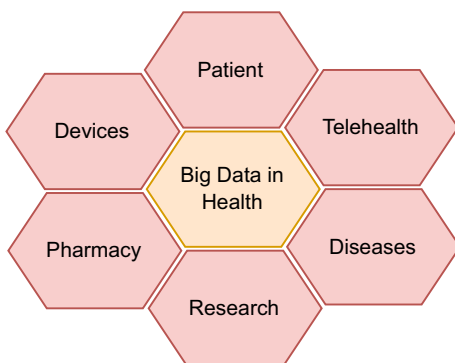
ingestion service in Wu et al. (2019) to address these challenges and establish best practices. Similarly, authors in Zhou et al. (2020) present a scalable system that securely stores and analyzes healthcare data from IoT devices using big data systems and blockchain architecture.

In the future, healthcare organizations will increasingly embrace big data for success. The use of HBD will enhance marketing strategies, especially with the growing popularity of wearable technology and the IoT. The integration of constant patient monitoring data from these sources will provide valuable insights, enabling healthcare marketers to identify and engage patients more effectively.

However, several concerns are associated with the utilization of HBD. One of the primary challenges involves network congestion and delays. The massive data generation, particularly during peak hours, congests the network. Real-time healthcare applications running during these times are significantly affected (Adeghe et al. 2024). This is the main reason why healthcare real-time applications do not trust the network. Furthermore, HBD is directly linked to lives. Therefore, it will take time and maturity to build trust in this technology. The health care data is also used for several tasks, such as research and treatment, however, no consent is taken in this regard (Al Teneiji et al. 2024).

## 3.2 Big data in logistics and transport

The integration of big data in logistics and transport has gained significant attention (Yadav and Jadhav 2019). Researchers have delved into BDA within SCM, identifying its potential to rectify deficiencies, enhance efficiency, and reduce costs (Lwin et al. 2019; Jahani et al. 2023). Particularly, in the context of the COVID-19 pandemic, logistics firms harnessed big data and supply chain integration (SCI) to optimize supply chain performance (Ved and B 2019; Fosso Wamba et al. 2018). Table 5 summarizes the big data applications and Fig. 6 shows the logistic and transport big data categorization.

Moreover, the synergy between big data analytics technology capability (BDATC) and SCI has been observed to bolster supply chain performance by fostering proactive and reactive capabilities, as well as resource reconfiguration (Leng et al. 2020). Blockchain technology has also made inroads in logistics and supply chain systems, lending technical support and mitigating risks (Chen et al. 2022b). In tandem, AI and big data analysis are utilized to scrutinize logistics service supply chain models, augmenting customer satisfaction and optimizing logistics operations (Farchi et al. 2023).

Table 5 Detail summary of big data applications in logistics and transport

Paper & Year	Main contributions	Limitation
Zhu and Du (2022)	Evaluation of maritime logistics enterprises' service capability using a vast big data pool from the IoT supply chain system. Proposal of an approach for evaluating service capabilities	The scalability and practical effectiveness of the proposed evaluation approach lack a comprehensive understanding
Yang et al. (2022)	Introduction of a quality management platform in supply chain logistics, leveraging edge-cloud blockchain and IoT for elevated quality management	Limited understanding of the scalability and real-world effectiveness of the quality management platform
Jiang (2019)	Advocating for an intelligent supply chain information collaboration model based on IoT and big data to improve information collaboration efficiency and mitigate the risks of supply chain disruption	Lack of comprehensive understanding regarding the scalability and practical effectiveness of the proposed collaboration model
Lyu and Zhao (2019)	Investigation of compressed sensing as a valuable method for conducting risk assessments in big data for Internet supply chain finance	Limited insights into the scalability and real-world effectiveness of compressed sensing in risk assessments
Zhili et al. (2021)	Use of blockchain technology as a robust solution to address security challenges in ITS and big data integration	Limited insights into the scalability and practical effectiveness of blockchain in ITS and big data integration

Fig. 6 Big data in supply chain
and logistics
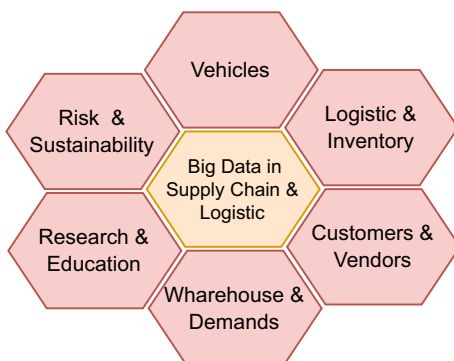
Significantly, the assessment of service capability in maritime logistics enterprises relies heavily on the extensive big data resources derived from the IoT supply chain system. This evaluation is crucial due to the numerous factors influencing maritime logistics, including overseas transportation routes. In Zhu and Du (2022), the authors suggest an approach for evaluating the service capabilities of maritime logistics enterprises by leveraging big data from the IoT supply chain system.

Moreover, an advanced cloud blockchain and Internet of Everything (IoE) enabled quality control platform seeks to improve quality management and bolster consumer confidence in perishable supply chain logistics, as discussed in Yang et al. (2022). This platform enables swift sensor data acquisition, ensuring authentication and transparency within cold supply chain logistics. In Jiang (2019), the authors endorse an intelligent supply chain model based on the IoT and big data. The objective of this model is to enhance information collaboration efficiency while mitigating the risks of supply chain disruption.

In the context of internet supply chain finance, compressed sensing proves to be a valuable method for conducting risk assessments within big data. Authors in Lyu and Zhao (2019) investigated the development of a risk assessment system for Internet supply chain finance, harnessing the power of compressed sensing and big data analysis. Furthermore, blockchain technology emerges as a robust solution to address security challenges in ITS and big data integration (Zhili et al. 2021). By using blockchain, data trustworthiness, transparency, and integrity are assured, surpassing the security standards of centralized databases.

The future holds significant promise for integrating big data in logistics and transportation. As highlighted by a research study (Insider 2023), last-mile delivery, a substantial portion of total shipping expenses, faces challenges such as carrier collaboration, manual processes, driver retention, fuel costs, WISMO ("Where is my order?") calls, and return costs. These challenges provide opportunities for optimization through the effective application of big data solutions.

While the integration of big data in logistics and transport is essential, there are associated concerns that need consideration. The primary concern is the privacy of drivers' locations, which may be misused. Similarly, the utilization of big data in logistics may also jeopardize customer privacy. Therefore, it is necessary to address all these concerns when planning the future of big data in logistics and transport (Albqowr et al. 2024).

3.3 Big data in marketing and advertising

Big Data has a substantial influence on marketing and advertising, enabling organizations to collect and scrutinize vast data reservoirs for informed decision-making (Craig and Ludloff 2011). It empowers precise targeting and customization of advertising messages, guided by consumer behaviours and preferences (Chen 2022; Cockcroft and Russell 2018). Real behaviours data marketing entails the collection of internet-driven behavioural data for in-depth analysis of advertising content, timing, and format. This, in turn, fosters more effective customer relationship management and enhances customer retention (Del Vecchio et al. 2022; Beauvisage et al. 2023). Table 6 summarizes the big data applications and Fig. 7 shows the marketing and advertising big data categorization.

In the financial sector, Big Data has ascended to prominence, with companies leveraging its capabilities for market analysis, customer insights, and informed decision-making. Authors in Hassani et al. (2018) explore into the pertinence of Big Data approaches in the financial realm, particularly within corporate banking, highlighting opportunities for technological advancements.

Furthermore, in the context of telecom big data, authors in Jia et al. (2019) propose a meticulous user classification scheme based on decision trees, aimed at amplifying marketing efficiency and effectiveness. The advent of Big Data technology has ushered in a paradigm shift in online advertising delivery, seamlessly integrating data, users, platforms, and businesses. Authors in Jieyu (2020) investigated the development of a precise online delivery system hinged on Big Data technology.

Cloud computing and Big Data technology have found extensive applications in the world of e-commerce advertising promotion, elevating the core competitiveness of enterprises within this industry. The authors in Zhang (2022) investigate the utilization of Big Data and cloud computing technology to enhance e-commerce advertising. They propose a distributed system built on Hadoop for this purpose. Similarly, authors in Ducange et al. (2018) furnish an in-depth analysis of SBD and its application in shaping marketing strategies, encompassing a comprehensive methodology and classification of contemporary use cases.

E-commerce and advertising cannot survive without big data. Nowadays, the action plans of e-commerce and advertising agencies rely heavily on big data analysis. This technology enhances targeted advertising, enabling businesses to reach potential customers more effectively. Therefore, it can be stated that the e-commerce and advertising domains represent significant applications of big data.

However, the deployment of Big Data in marketing and advertising gives rise to substantial concerns regarding privacy and the potential for government surveillance, as discussed in Tang et al. (2022). Despite its advantages, Big Data in marketing and advertising presents challenges such as the crucial need for consent and the complexities surrounding transparency, identity, power dynamics, and inclusivity (Yin et al. 2021). Therefore, it is necessary to prioritize customer data privacy when planning the integration of big data in commerce and advertising.

3.4 Big data in smart cities

The combination of the IoT and BDA technologies holds the potential to be a game-changer in the construction of smart cities (Bibri 2019). These technologies provide opportunities

Table 6 Detail summary of big data applications in marketing and advertising

Paper & year	Main contributions	Limitation
Hassani et al. (2018)	Exploration of Big Data approaches in corporate banking for market analysis and customer insights	Limited insights into technological advancements opportunities
Jia et al. (2019)	Proposal of a user classification scheme based on decision trees in tel-ecom big data for amplifying marketing efficiency	Limited insights into the effectiveness of the classification scheme
Jieyu (2020)	Investigation of a precise online advertising delivery system hinged on Big Data technology	Challenges in seamlessly integrating data, users, platforms, and businesses
Zhang (2022)	Examining the integration of big data and cloud computing technology within the realm of e-commerce advertising promotion	Limited insights into the specific applications of the distributed system on Hadoop
Ducange et al. (2018)	In-depth analysis of SBD and its application in shaping marketing strate-gies	Comprehensive methodology and classification of contemporary use cases

Fig. 7 Big data in marketing and advertising
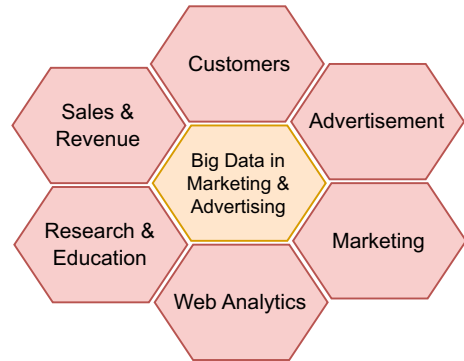
for efficient disaster management activities, analysis, and the acquisition of valuable information for decision-making (Shah et al. 2019; Ding et al. 2023). Table 7 summarizes the big data applications and Fig. 8 shows the UBD categorization.

A plethora of devices connected to the internet in smart cities continuously generates vast amounts of data. Addressing this data deluge, researchers in Wang et al. (2018) propose enhanced multi-order distributed algorithms to efficiently process this big data in the realm of smart city services. Similarly, authors in Alahakoon et al. (2020) advocate for a comprehensive framework designed to handle the substantial data inflow from sources such as sensors, IoT devices, and social networks within smart cities. This framework encompasses data processing workflows, ML algorithms, and statistical techniques aimed at extracting meaningful insights from the data.

The evaluation of smart cities has resulted in the generation of massive quantities of data. Unfortunately, a significant portion of this data often goes to waste due to the absence of established mechanisms and standards for extracting valuable information. Authors in Chang (2021) discuss the issues and approaches linked with leveraging big data and ML to enable cognitive smart cities, thereby enhancing the utilization of this data.

In alignment with this, authors in Wu et al. (2018) present a framework designed to efficiently process the large amounts of data generated by sensors in smart cities. This architectural model comprises various layers and components for data processing and analysis. ML techniques are integral to this framework, ensuring the acquisition of accurate data and the delivery of precise information to end-users, ultimately resulting in an elevated Quality of Experience (QoE) performance.

In anticipation of the growing prevalence of cameras in smart cities, video surveillance is becoming a key component of data collection. This evolution necessitates the development of efficient techniques for processing substantial volumes of video data. Several papers in the field look into this topic. Tian et al. (2018) propose a block-level background modelling (BBM) algorithm for efficient video coding, complemented by a rate-distortion optimization algorithm designed to enhance compression performance.

The part of big data in the implementation of smart cities is crucial, as it enables the analysis of extensive data volumes to extract valuable insights. In He et al. (2018), the authors utilize special technologies for municipal governance and planning in smart cities. Similarly, in Kandt and Batty (2021), authors delve into the value of big data in shaping long-term urban planning. They emphasize how urban analytics can inform these long-term urban policies within smart cities.

The perspective of big data in smart cities promises transformative advancements. The integration of big data and the IoT is set to revolutionize urban living. Expect more

Table 7 Detail summary of big data applications in smart cities

Paper & year	Main contributions	Limitation
Wang et al. (2018)	Proposal of enhanced multi-order distributed algorithms for processing big data in smart city services	Limited insights into the efficiency of the proposed algorithms
Alahakoon et al. (2020)	Advocacy for a comprehensive framework for handling substantial data inflow in smart cities	Challenges in implementing data processing workflows and ML algorithms
Chang (2021)	Discussion on challenges and approaches for utilizing big data and ML in cognitive smart cities	Limited insights into the specific challenges associated with cognitive smart cities
Wu et al. (2018)	Presentation of a framework for the efficient processing of data generated by sensors and devices in smart cities	Challenges in integrating ML techniques for improved Quality of Experience (QoE) performance
Tian et al. (2018)	Proposal of a block-level background modelling (BBM) algorithm for efficient surveillance video coding	Limited insights into the rate-distortion optimization algorithm's impact on compression performance
He et al. (2018)	Utilization of spatial information technologies, such as GIS, for municipal governance and planning in smart cities	Limited insights into the specific applications of GIS in municipal governance
Kandt and Batty (2021)	Investigation into the potential and constraints of big data in shaping long-term urban policy and planning	Insights into how urban analytics can inform long-term urban policies in smart cities
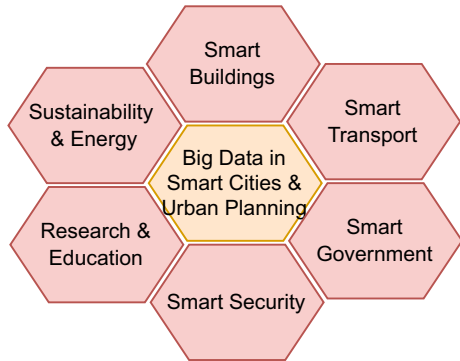
sophisticated data analytics, real-time insights for resource management, improved infrastructure planning, and AI-driven solutions to address urban challenges. This evolution aims to create proactive, sustainable, and resilient smart cities.

However, the widespread use of big data in smart cities brings critical concerns. Security and privacy issues surrounding the vast data generated by IoT devices and sensors need careful attention. Protecting data from unauthorized access and ensuring citizen privacy requires robust security measures and regulatory frameworks. Ethical considerations in data collection, storage, and usage demand scrutiny to prevent misuse. Striking a moderation between reaping the benefits of big data in urban development and safeguarding individual privacy is crucial for fostering trust and ensuring sustainable and inclusive smart city growth (Thilagavathi et al. 2019; Elhoseny et al. 2018).

3.5 Big data in media

The intersection of big data and entertainment is a dynamic field with vast potential for insights, innovation, and, at the same time, several challenges to navigate Abbasi et al. (2018); Daud et al. (2013). Table 8 summarizes the SBD applications and Fig. 9 shows the media big data categorization.

Social media platforms are prolific producers of what's referred to as SBD (Badshah et al. 2022b). This treasure trove of data is a window into user behaviour, trends, and interactions, offering valuable insights (Esfahani et al. 2019). Companies recognize the power of this data and utilize it to personalize marketing strategies, pinpoint specific demographics, and boost sales (Ghani et al. 2019; Rahman and Reza 2022). Social media also serves as a powerful platform for businesses to engage with their customer base, foster loyalty, and even function as online retail spaces (Liu et al. 2021; Hayat et al. 2019).

However, the employing of SBD raises significant concerns related to privacy and the potential misuse of personal information, as highlighted in Bansal et al. (2018). Thus, the combination of big data and social media presents a dual landscape, offering opportunities for innovation, effective marketing, and improved decision-making. However, it is laden with challenges and ethical considerations. Similarly, authors in Mani and Chouk (2022) and Vargo et al. (2018) discussed privacy and security issues in media big data.

To investigate the role of social media big data, the authors in Jimenez-Marquez et al. (2019) propose a comprehensive two-stage framework tailored for the big data era. The first stage emphasizes data preparation and the selection of a ML model, while the second stage utilizes established layers of big data architectures to extract insights from the data.

Table 8 Detail summary of big data applications in media

Paper & year	Main contributions	Limitation
Jimenez-Marquez et al. (2019)	Proposal of a two-stage framework for analyzing SBD, versatile for large and small datasets	Limited insights into the scalability of the framework in large-scale events
Zhang et al. (2022)	Introduction of the Big Data-assisted Social Media Analytics for Business (BD-SMAB) model for decision-making in marketing strategies	Challenges related to misinformation and limited data affecting decision-making
Sivarajah et al. (2020)	Exploration of the role of social websites in B2B marketing for sustaining and expanding businesses	Limited insights into challenges faced by B2B organizations in leveraging social media
Ilieva and McPhearson (2018)	Investigation of opportunities and challenges in harnessing SBD for urban sustainability research	Limited insights into the scalability and real-world impact of using social media data for urban sustainability

Fig. 9 Big data in media
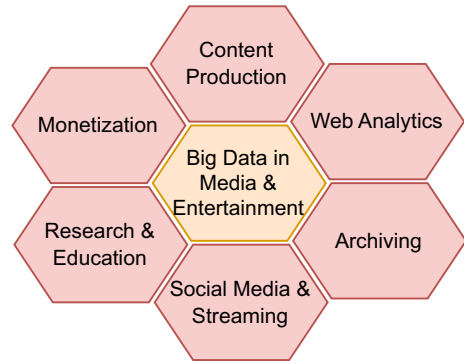
This versatile framework accommodates both large and small datasets and is illustrated through a case study focused on analyzing reviews of hotel-related businesses. Similarly, in the study (Zhang et al. 2022), the authors introduce the Big Data-assisted Social Media Analytics for Business (BD-SMAB) model to enhance decision-making in marketing strategies and competitive analysis.

Social media is a focal point for marketing, especially for business-to-business (B2B) organizations aiming to sustain and expand through strategic operations and marketing activities, as explained by authors in Sivarajah et al. (2020).

The potential of SBD is also recognized in the realm of urban sustainability research and practice. Its unique advantages, including vast scale and near-real-time observation, offer insights into human behaviour within urban environments. Authors in Ilieva and McPhearson (2018) delve into the potential and issues associated with harnessing social media data for urban sustainability research and practice, shedding light on a promising avenue for urban development.

The integration of big data in entertainment and social media is currently revolutionizing user experiences, content creation, and industry dynamics. With ongoing technological advancements, big data is driving personalized content recommendations, offering predictive insights, enhancing user engagement, enabling targeted advertising, optimizing content distribution, and facilitating real-time trend analysis (Hariri et al. 2019). Emphasizing data security and privacy measures, these developments are transforming the industry, providing tailored and immersive experiences, improving content relevance, and ensuring efficiency in advertising and content distribution. To remain competitive in these evolving sectors, a seamless integration of BDA is essential to meet the dynamic expectations of users in today's rapidly advancing technological landscape (Amalina et al. 2019).

Despite its benefits, social media big data faces challenges such as misinformation and limited data, making it difficult to distinguish the truth. Current solutions struggle with scalability in large-scale events (Zhang et al. 2018). Furthermore, this big data is wrongly used by companies, as they share it with commercial entities. These companies enforce their narratives through advertising. Therefore, it is necessary to address these concerns while working on SBD, especially in entertainment, particularly on social media.

3.6 Big data in cyber security

Playing a crucial role in cybersecurity, big data is especially significant in domains such as intrusion detection, anomaly detection, spamming and spoofing detection, malware and

ransomware detection, code security, and cloud security (Walters and Novak 2021). The integration of BDA with ML can effectively address unknown risks and insider threats, providing advanced threat analytics (Saravanan and Prakash 2021). It enables the discovery of irregularities and suspicious activities, leading to the deployment of effective intrusion detection systems (França et al. 2021). Additionally, BDA can enhance data security and privacy, mitigating cybersecurity breaches and supporting secure information sharing (Rassam et al. 2017). The application of BDA in cybersecurity is an emerging trend, presenting potential future directions for research and development (Wang and Jones 2021). Table 9 summarizes the applications and Fig. 10 shows the cybersecurity big data categorization.

By leveraging big data and advanced analytics techniques, organizations can improve their operational intelligence and security capabilities, staying ahead of evolving cyber threats. Authors in Kantarcioglu and Xi (2016) discussed security issues faced in the big data environment, particularly in the context of cloud computing.

Surveilling the security of the IoT through multidimensional streaming big data encounters various challenges, including substantial data volumes, redundancy, and scalability issues. To tackle these obstacles, the authors in Ullah et al. (2022) present an algorithm called ODIS. This algorithm extracts vital information from data across distributed sensor nodes, considering the spatial and temporal dependence structure of the data. ODIS establishes a precise data structure model to understand IoT system behaviours and employs testing methods to quantify the uncertainty linked with monitoring tasks. Adversarial data mining is an emerging field that combines BDA with cybersecurity. Authors in Li et al. (2019) used adversarial data mining techniques to handle malicious adversaries in cyber security applications.

In Tao et al. (2019), the authors introduced a parameter-wise adaptation that autonomously initiates the tuning process. This system adjusts the configuration parameters of the framework for various security datasets and subsequently executes the BDCA system with the adapted configuration. Similarly, Rawat et al. (2019) explores the economic aspects of safeguarding big data security and privacy, encompassing investment decisions and cyber insurance.

To tackle the challenges posed by cyber threats in the cloud, the authors in Subroto and Apriyana (2019) have devised a cloud computing-based system for cybersecurity management. This system aims to streamline the analysis process of extensive network data. The constructed system is built on the MapReduce framework and encompasses end-user devices, cloud infrastructure, and a monitoring center.

Big data is advancing cybersecurity, making it more intelligent for the future. This increased intelligence will enable systems to promptly counter cyber attacks. Consequently, cybersecurity experts are acquiring additional skills in both big data and cybersecurity, driven by the recognition of the crucial role played by these combined capabilities (Zhang and Ghorbani 2021).

Big data in cybersecurity offers potent advantages but introduces challenges, including privacy concerns, security issues, data accuracy, scalability, and cost management. Successfully navigating these hurdles requires a comprehensive strategy addressing legal compliance, robust security measures, data quality assurance, and cost-effective implementation (Rao and Lakshmanan 2024).

3.7 Big data in earth science

An extensive array of data about our planet, which is usually also referred to as Earth Big Data (EBD) is generated from Earth observation systems on diverse platforms, such as satellites, aeroplanes, and ground-based setups. This includes geoscience, statistical, and

Table 9 Detail summary of big data applications in cyber security

Paper & year	Main contributions	Limitation
Ullah et al. (2022)	Proposal of an online distributed IoT security monitoring algorithm (ODIS) for handling large volumes of multidimensional streaming big data	There is a scarcity of comprehensive understanding regarding the scalability and practical deployment of the ODIS algorithm
Li et al. (2019)	Exploration of adversarial data mining techniques for handling malicious adversaries in cybersecurity	Limited insights into the real-world effectiveness of adversarial data mining in cybersecurity applications
Tao et al. (2019)	Proposal of a parameter-driven adaptation system for Big Data Cyber Security Analytics (BDCSA) systems	Lack of comprehensive understanding regarding the scalability and adaptability of the proposed parameter-driven adaptation system
Rawat et al. (2019)	Investigation of the economic perspective of protecting big data security and privacy, including investment decisions and cyber insurance	There is a shortage of comprehensive insights into the real-world economic impacts and practical implementation of the proposed economic perspective
Subroto and Apriyana (2019)	Development of a cyber security management system based on cloud computing and constructed using the MapReduce framework	Limited insights into the scalability and efficiency of the developed cloud computing-based cyber security management system

Fig. 10 Big data in cyber security
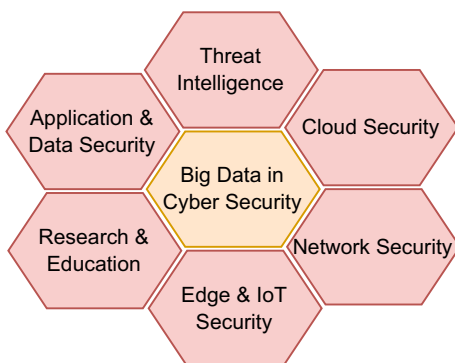
social data (Yang et al. 2019). Integrating Earth observation data with other forms within a geographic context offers the potential to model Earth systems more accurately, linking human activities with their impacts on Earth processes (EOS 2023). Table 10 summarizes the applications and Fig. 11 shows the earth's big data categorization.

Big data applications in climate and earth studies have gained increasing importance in recent years. These applications involve the utilization of large volumes of data generated from climate and weather modelling (Huang et al. 2018). The analysis of this big climate data has led to advancements in understanding climate change, assessing environmental conditions, and predicting future climate trends. Leveraging BDA, including data mining techniques and the integration of heterogeneous data sources, has empowered researchers to study climate change in a more comprehensive and interdisciplinary manner. Open data resources, like Google Earth Engine, have been used to evaluate environmental conditions and assess vulnerability to climate change in specific regions (Amani et al. 2020). Overall, big data tools and techniques have provided valuable insights into climate-related issues and have the potential to contribute to sustainability and resilience-building efforts.

Big data on climate and earth is used for several purposes. The foremost use is the monitoring. Authors in Hassani et al. (2019) designed BDA to enhance seasonal change monitoring and understanding of climate change. The second big use of big data is to predict the climate and conditions. Authors in Knüsel et al. (2019) used Big data techniques in rainfall prediction, helping farmers make wise decisions on crop yield and studying the timing of floods or droughts. Similar concepts are discussed and proposed by authors in Sebestyén et al. (2021) and use the big data collected by different sensors for climate monitoring and prediction. Authors in Silva et al. (2018) discussed in detail the studies, which investigated big data climate monitoring and prediction.

Along with climate monitoring and prediction, big data is used for Sustainable Urban Planning and Infrastructure. Authors in Leung et al. (2019); Ameer and Shah (2018) used big data and its analytics tools in urban planning and smart city decision management. Similarly, authors in Sarker et al. (2020) used BDA for smart cities' air pollution prediction. They introduced a spark-based architecture for smart urban planning that utilizes BDA to classify air quality. This architecture is implemented on a dataset of vehicle pollution in Aarhus City, Denmark.

Disaster management has become a significant concern, and Big Data is being utilized for natural disaster management. Authors in Yu et al. (2018), utilized big data for disaster management derived from remote sensing imagery, social media data, crowdsourced data, GIS, and mobile metadata. Similarly, in Sarker et al. (2020b), the authors investigated several studies exploring the use of big data in disaster management.

Table 10 Detail summary of big data applications in earth science

Paper & year	Main contributions	Limitation
Huang et al. (2018)	Advancements in climate and earth studies through the analysis of big climate data	Challenges such as continuous growth of Earth's big data and limitations of current tools in thoroughly analyzing it
Hassani et al. (2019)	Design of BDA for enhanced seasonal change monitoring and understanding of climate change	Lack of comprehensive understanding regarding the scalability and real-world implementation of the designed BDA
Knüsel et al. (2019)	Use of Big Data techniques in rainfall prediction for informed decision-making in agriculture	Lack of comprehensive understanding regarding the scalability and real-world effectiveness of Big Data techniques in rainfall prediction
Leung et al. (2019)	Utilization of big data and analytics tools in urban planning and smart city decision management	Limited exploration of the scalability and practical impacts of big data in urban planning
Yu et al. (2018)	Application of big data produced from various sources for disaster management	Limited exploration into the scalability and practical effectiveness of big data in disaster management

Fig. 11 Big data in earth science
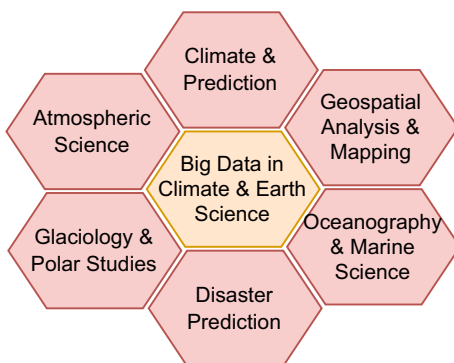
The main challenge associated with the Earth's big data is its continuous growth. Every country deploys satellites, balloons, aeroplanes, and other tools that consistently gather data. However, reaping benefits from this data is contingent upon having appropriate tools. Regarding the sheer volume of Earth's big data, our current tools are not advanced enough to thoroughly analyze it Sudmanns et al. (2019).

The foremost concern regarding Earth Big Data is individual privacy. This data is constantly generated without regard for the privacy of specific locations, making it accessible to anyone for various purposes. The data finds application in numerous fields, including science, weather prediction, and defence. The issue of precisely identifying the responsible party or owner of this data remains unresolved (Farley et al. 2018). Therefore, there is a need to explore whether it is feasible to collect this data with individual consent and whether regulations can be established to govern this vast dataset.

3.8 Big data in industry

Big data is being applied in various industries, including construction, sports, tourism, and the legal field. In the construction industry, big data is utilized to enhance construction efficiency, reduce material waste and expenses, improve planning and decision-making processes, and enhance construction site safety (Nguyen et al. 2020). In the sports industry, big data analysis and AI are used to analyze player performance, broadcast events, and improve sports marketing strategies (Patel et al. 2020). In tourism, big data is used for revenue management, marketing strategies, customer experience, and market research, aiding in the development and recovery of the industry (Li et al. 2022). In the legal industry, BDA tools are used for tasks such as billing, marketing, and identifying trends in cases (Bhure and Desai 2023). Table 11 summarizes the applications and Fig. 12 shows the industry big data categorization.

Authors in Lies (2019) covered big data's transformative role in automotive marketing, emphasizing precision marketing and data-driven consumer insights. A similar theme is explored in Liu and You (2021), where big data correlates with a 2.895% increase in new energy vehicle technology innovation, advocating its integration with the industry for national benefits.

Classification benefits from big data too, as seen in Li et al. (2019), where cellular company customer records are categorized to enhance marketing efficiency. In Chen et al. (2022a), big data analysis is used to create tailored data packages. The chemical industry harnesses big data for intelligent manufacturing, evaluating strengths, weaknesses, and

Table 11 Detail summary of big data applications in the industry

Paper & year	Main contributions	Limitation
Lies (2019)	Transformative role of big data in automotive marketing with a focus on precision marketing and data-driven consumer insights	Authors failed to discuss the scalability and real-world implementation of big data in automotive marketing
Li et al. (2019)	Utilization of big data for cellular company customer records classification to enhance marketing efficiency	Lack of comprehensive understanding regarding the scalability and real-world effectiveness of big data in customer records classification
Jiyang et al. (2020)	Harnessing big data for intelligent manufacturing in the chemical industry, evaluating strengths, weaknesses, and future trends	Limited exploration into the scalability and real-world impacts of big data in intelligent manufacturing
Mohammadpoor and Torabi (2020)	Adoption of big data in Huabei Oilfield for oil production engineering through a "seven-step method" system and a data mining platform	Limited investigation into the scalability and practical effectiveness of big data in oil production engineering

Fig. 12 Big data in industry
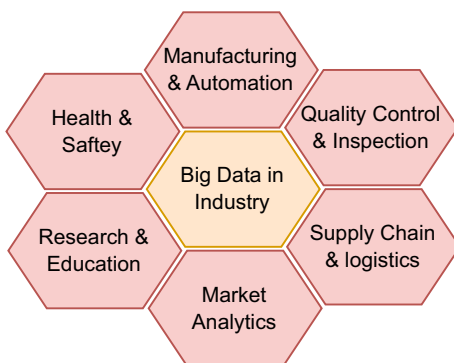
future trends (Jiyang et al. 2020). Similarly, Huabei Oilfield adopts big data with a "seven-step method" system and a data mining for oil production engineering, enhancing data-driven processes (Mohammadpoor and Torabi 2020).

Big data is reshaping industries, particularly production, in alignment with market analysis. The expanding realm of big data is certain to amplify its influence on the industry. Utilizing big data analysis will enhance customer-centric production strategies, ultimately leading to improved revenue outcomes (Vassakis et al. 2018).

Big data utilized for market analysis is collected from various sources, raising significant concerns about the privacy and security of this data. Therefore, it is imperative to ensure that the data collection and analysis do not compromise someone's privacy and security (Del Vecchio et al. 2018).

3.9 Big data in education

Big data has the potential to enhance teaching and learning, improve educational research, and advance education governance (Fischer et al. 2020). Although the utilization of big data in education is not a new concept, recent technological advancements have spurred increased research in this area (Ray and Saeed 2018; Amjad et al. 2018). There is an interest in leveraging big data to analyze student behavior and performance, enhance the educational system, and integrate big data into the curriculum (Baig et al. 2020). Popular tools and techniques for working with big data in the education industry include educational data mining and learning analytics (Qian et al. 2022). The convergence of the ability to collect, store, manage, and process data, along with data from online educational platforms, presents unprecedented opportunities for educational institutions, learners, educators, and researchers. Table 12 provides a summary of the applications, and Fig. 13 illustrates the categorization of big data in education.

In educational technology, the most investigated these days is personalized learning. With the help of personalized learning, the personalized content or subjects are recommended to the learners and they can learn in their own space (Munshi and Alhindi 2021). Authors in Yuwen et al. (2018) carried out some experiences to appropriately suggest the courses to the learners using BDA. Their results show that their accuracy for the course recommendation is much better than the already working algorithms. Similarly, authors in Kanth et al. (2018) highlighted the challenges of identifying student misconceptions, predicting dropouts, and improving educational quality, with a focus on leveraging data

Table 12 Detail summary of big data applications in education

Paper & year	Main contributions	Limitation
Yuwen et al. (2018)	Investigation of personalized learning using BDA, showcasing improved course recommendation accuracy	Lack of scalability and real-world effectiveness of big data in personalized learning
Kanth et al. (2018)	Highlighting challenges in identifying student misconceptions, predicting dropouts, and improving educational quality, proposing solutions with a focus on data and advanced technologies	Lack of comprehensive insights into the scalability and real-world impacts of big data in addressing educational challenges
Zhang et al. (2021)	Utilization of big data for student management and discipline by analyzing daily routines, learning styles, and behavior	Limited insights into the scalability and real-world effectiveness of big data in student management
Hu et al. (2022)	Proposal of a hybrid teaching method using big data, demonstrating increased student engagement in comparison to traditional classes	There is a scarcity of comprehensive insights into the scalability and real-world effectiveness of big data in hybrid teaching

Fig. 13 Big data in education
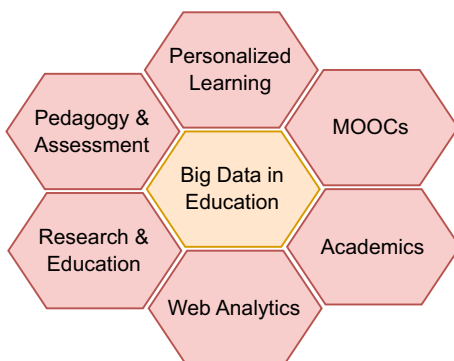
and advanced technologies. The authors aim to enhance personalized learning and propose various supervised learning methods as solutions.

Student management and discipline represent significant challenges in educational institutions. The authors in Zhang et al. (2021) addressed this issue by leveraging big data. Through the analysis of students' daily routines, learning styles, and behavior, they obtained insights to aid in student management. In Liang (2020), the authors present an education management model utilizing big data, demonstrating improved information levels and a broader application of big data in educational management. Similarly, authors in Badshah (2023a, 2023b) utilize similar concepts for student management and enhancing their productive engagement.

The big data is also changing the way of teaching. Flipped classrooms (Hao 2021) and homeschooling (Inayatulloh et al. 2022) are the leading examples. Authors in Hu et al. (2022) explored the same by proposing the hybrid teaching method. Their investigation shows that students were more actively engaged in the learning concerning the normal classes.

Education is intricately linked with big data as both a producer and consumer. Millions of individuals, whether learners, teachers, or administrators, are actively engaged in this dynamic field. The demand for virtual classes has surged during the COVID-19 pandemic, further emphasizing the role of big data in meeting these evolving educational needs. Concepts like personalized learning and home-based schooling are gaining prominence, relying entirely on the insights and capabilities provided by big data. In this interconnected landscape, the symbiotic relationship between education and big data continues to shape the future of learning.

While there is a considerable list of advantages, the use of big data in education also raises several concerns. Foremost among them is the risk of misuse, as the data of thousands of learners, including institutional geography and learner locations, may be mishandled. Additionally, concerns about data bias and algorithmic bias pose potential challenges that need careful consideration to ensure fair and equitable outcomes (Lin et al. 2024).

4 Key technologies

In big data, several key enabling technologies play pivotal roles in facilitating the storage, processing, and analysis of extensive and intricate datasets. These technologies serve as the backbone for the vast potential of big data applications. Here are some of the key enabling technologies discussed.

Table 13 Category-wise publications since 2018

Category	Subcategories	Key studies
Health care	Hospitals, telehealth, pharmaceutical research	Philip et al. (2022), Mehta and Pandit (2018), Dash et al. (2019), Yan et al. (2019), Wu et al. (2019), Zhou et al. (2021), Makkie et al. (2018), Galletta et al. (2018), Hong et al. (2019), Demirbaga and Aujla (2022), Zhou et al. (2020)
Supply chain and logistic	Transportation management, electric vehicles, autonomous drones	Yadav and Jadhav (2019), Lwin et al. (2019), Ved and B (2019), Leng et al. (2020), Chen et al. (2022b), Zhu and Du (2022), Yang et al. (2022), Jiang (2019), Lyu and Zhao (2019), Zhili et al. (2021)
Marketing and advertising	Digital Marketing, customer analytics, ad personalization	Craig and Ludloff (2011), Chen (2022), Del Vecchio et al. (2022), Tang et al. (2022), Yin et al. (2021), Hassani et al. (2018), Jia et al. (2019), Jieyu (2020), Zhang (2022), Ducange et al. (2018)
Smart cities	Urban planning, IoT infrastructure, sustainable technologies	Shah et al. (2019), Wang et al. (2018), Alahakoon et al. (2020), Chang (2021), Elhoseny et al. (2018), Tian et al. (2018), Wu et al. (2018), He et al. (2018), Thilagavathi et al. (2019), Kandt and Batty (2021)
Media and entertainment	Streaming services, content recommendation, user engagement	Badshah et al. (2022b), Esfahani et al. (2019), Ghani et al. (2019), Liu et al. (2021), Bansal et al. (2018), Jimenez-Marquez et al. (2019), Zhang et al. (2022), Sivarajah et al. (2020), Zhang et al. (2018), Ilieva and McPhearson (2018)
Cybersecurity	Network security, threat detection, encryption technologies	Walters and Novak (2021), Saravanan and Prakash (2021), França et al. (2021), Rassam et al. (2017), Wang and Jones (2021), Kantarcioglu and Xi (2016), Li et al. (2019), Ullah et al. (2022), Tao et al. (2019), Rawat et al. (2019), Subroto and Apriyana (2019)
Climate & earth science	Climate modeling, environmental monitoring, geospatial analysis	Hassani et al. (2019), Knüsel et al. (2019), Sebestyén et al. (2021), Silva et al. (2018), Leung et al. (2019), Ameer and Shah (2018), Sarker et al. (2020), Yu et al. (2018), Sarker et al. (2020b), Amani et al. (2020), Huang et al. (2018)
Industry	Manufacturing processes, automation, quality control	Nguyen et al. (2020), Patel et al. (2020), Li et al. (2022), Bhure and Desai (2023), Lies (2019), Liu and You (2021), Li et al. (2019), Chen et al. (2022a), Jiyang et al. (2020), Mohammadpoor and Torabi (2020)

Table 13 (continued)

Category	Subcategories	Key studies
Education	E-Learning platforms, learning analytics, student performance tracking	Fischer et al. (2020), Ray and Saeed (2018), Baig et al. (2020), Qian et al. (2022), Yuwen et al. (2018), Kanth et al. (2018), Zhang et al. (2021), Badshah et al. (2023a), Liang (2020), Munshi and Alhindi (2021), Hu et al. (2022), Hao (2021), Inayatulloh et al. (2022)

4.1 Hadoop

At the forefront, Hadoop stands as a distributed storage and processing framework that enables parallelized handling of large datasets. Its architecture allows for efficient and scalable data processing, making it a cornerstone in the big data ecosystem (Apache 2023a).

4.2 Apache spark

Complementing Hadoop, Spark emerges as an in-memory data processing engine that significantly enhances the speed and efficiency of BDA. It excels in iterative computations and ML algorithms, contributing to improved data processing capabilities, as discussed in Apache (2023c).

4.3 NoSQL databases

In the era of diverse data types, NoSQL databases like MongoDB (Apache 2023d) and Cassandra (Cassandra 2023) play a vital role. These non-relational databases accommodate unstructured and varied data, providing flexibility and scalability crucial for managing the complexities of modern data.

4.4 Data warehousing

Technologies such as Amazon Redshift (Amazone 2023) and Google BigQuery (Google 2023) exemplify the capacity to store and retrieve large volumes of structured data. These solutions for data warehousing enable organizations to effectively handle and retrieve their data for analytical purposes.

4.5 Machine learning

The integration of ML algorithms and frameworks, including TensorFlow (Tensor 2023) and scikit-learn Learning (2023), empowers data scientists to derive actionable insights and predictions from vast datasets. ML becomes an invaluable tool in uncovering patterns and trends within the data.

4.6 Data integration tools

Apache NiFi Apache (2023b) and Talend (2023) exemplify the significance of data integration tools. These platforms facilitate the seamless integration of diverse data sources, ensuring a unified and coherent dataset ready for comprehensive analysis.

4.7 Data visualization tools

Platforms like Tableau (2023) and Power BI Microsoft (2023) add a layer of accessibility to big data insights. These visualization tools transform complex datasets into

digestible visualizations, enabling stakeholders to interpret and understand data-driven narratives.

4.8 Blockchain technology

Highlighting security and transparency, blockchain technology contributes to safeguarding the integrity of transactions and data sharing in the realm of big data. The decentralized nature of blockchain enhances both trust and data immutability (Badshah 2023c).

4.9 Edge computing

To fulfil the demand for real-time analytics, edge computing facilitates data handling near the data source. This minimizes latency and improves the efficiency of analytics for applications such as the IoT (Amjad et al. 2018).

4.10 Cloud computing

Services offered by AWS, Azure, and Google provide a scalable and flexible infrastructure for big data storage and processing. Cloud computing has become a cornerstone, furnishing organizations with the resources required to manage the continuously expanding volumes of data (Badshah 2023a).

5 Potential concerns and solutions

As we have explored the expansive landscape of employing big data across various applications, it becomes imperative to acknowledge and address potential challenges and concerns associated with its widespread utilization (Ajah and Nweke 2019). These concerns, spanning privacy, security, biases, and misuse, highlight the need for understanding the implications and risks inherent in BDA (Ikegwu et al. 2024). In this section, we delve into these concerns, acknowledging the multifaceted nature of navigating complexities when harnessing vast datasets. Additionally, we present solutions to tackle these challenges, offering a roadmap for a more secure and responsible digital environment. This section provides a detailed examination of proactive measures, carefully crafted to address distinct aspects of concern.

5.1 Privacy

A substantial concern associated with the utilization of big data across various applications is the issue of privacy. Almost every field grapples with this concern due to the underdeveloped nature of regulations on data security and privacy. The existing rules lack maturity, posing a challenge in adequately protecting user data (Amaithi Rajan and V 2023; Masood et al. 2018a).

To address privacy issues linked with big data, it is crucial to advocate for the development and implementation of robust regulations governing data security and privacy. Collaborating with regulatory bodies and policymakers to create comprehensive and mature frameworks will enhance the protection of user data (Price and Cohen 2019).

5.2 Security

The lack of data privacy raises security concerns, not just for the data itself but also for individual security. Organizational data exposure or the public availability of individual locations can lead to notable security problems. The connection between data vulnerability and individual security issues exacerbates the overall concern (Khan and Ahmad 2023).

Mitigating security risks involves reinforcing data privacy measures. Implementing encryption, access controls, and regular security audits can fortify the protection of organizational and individual data. Additionally, fostering awareness about cybersecurity practices among users is essential for minimizing vulnerabilities (Ikegwu et al. 2022).

5.3 Biases

Algorithmic bias, especially in IoT devices, is a common problem nowadays. Similarly, BDA may also exhibit biases in their calculations, disrupting decision-making processes (Rehman et al. 2022).

Addressing algorithmic biases in BDA requires continuous monitoring and evaluation of algorithms. Implementing diversity in datasets and adopting ethical guidelines for algorithm development can help mitigate biases, ensuring fair and unbiased decision-making processes (Favaretto et al. 2019; Amjad et al. 2012).

5.4 Misuse

Misuse of big data is a major concern, with companies often utilizing this data without considering the welfare of customers. Many individuals are unaware of how their data is being used for the benefit of companies. Mitigating potential misuse requires increased transparency and ethical considerations (Stegenga et al. 2023).

Preventing the misuse of big data involves enhancing transparency in data usage and fostering ethical considerations. Implementing clear data usage policies, obtaining explicit consent from users, and educating individuals about how their data is utilized contribute to responsible and ethical data practices (Bag et al. 2023).

5.5 Different cyber laws

The internet has transformed the world into a global village, however, the issue is that cyber rules and regulations vary widely. Every country has different rules, leading to conflicts on the internet. An action may be a crime in some countries and not in others, highlighting the need for harmonizing international cyber regulations (Rawat et al. 2023).

When the big data concerns are collectively looked at, it is noticed that all these concerns are linked with international cyber laws. Due to this gap, the digital world has these

issues. It is, therefore, important and need of the day to go ahead toward international cyber rules, which will equally work in all countries (Favaretto et al. 2019).

5.6 Doubted accuracy

Despite its advantages, social media big data faces challenges such as misinformation and limited data, making it difficult to distinguish the truth. Therefore, it is not always guaranteed that the big data used for decision-making is correct (Badshah et al. 2022b).

Ensuring the accuracy of big data used for decision-making requires implementing rigorous data validation processes. Incorporating fact-checking mechanisms, promoting data transparency, and investing in data quality assurance measures contribute to the reliability of information derived from big data (Khan et al. 2016).

5.7 Reason for network congestion

The rapid growth in data generation leads to network congestion, slowing internet speed and impeding real-time communication. This poses challenges, particularly in critical applications like hospitals, where trust in the network's reliability is compromised. Addressing congestion is crucial for ensuring seamless real-time interactions and maintaining the dependability of data-driven systems (Anitha et al. 2023).

Addressing network congestion involves optimizing data transmission protocols, investing in network infrastructure, and implementing load-balancing techniques. Prioritizing network reliability in critical sectors like healthcare ensures that real-time communication remains unaffected even during periods of high data traffic (Al-Jumaili et al. 2023).

5.8 Special hardware and software

In big data, concerns emerge regarding the need for specialized hardware and software. Access and compatibility challenges risk obstructing positive outcomes. Processing vast data volumes demands specialized tools, resist by limitations in hardware or software and the complexity of multiple data formats. The substantial processing needs also contribute to higher costs, necessitating careful cost management for optimal resource utilization (Badshah et al. 2022a).

To overcome challenges related to specialized hardware and software, organizations should invest in versatile and scalable technologies. Collaborating with technology providers to develop solutions that enhance accessibility and compatibility can facilitate positive outcomes without compromising on processing efficiency (Selmy et al. 2023).

5.9 Dependency on tech experts

One limitation of big data lies in its dependency on technology experts for its collection, filtering, and processing. This reliance poses a challenge in ensuring that the necessary expertise is consistently available for the effective utilization of big data resources (Badshah 2023b).

Reducing the dependency on tech experts requires investing in user-friendly interfaces and tools. Implementing training programs for non-experts and promoting the development of intuitive big data platforms can empower a wider range of professionals to harness the power of big data resources effectively (Selmy et al. 2023).

6 Comparative analysis

This section compares the current literature study with related surveys. Scholars have extensively studied and investigated big data and its applications. However, existing reviews often focus on a single application of big data, failing to explore it comprehensively. Big data has potential and challenges in every domain, necessitating a thorough investigation. Additionally, no studies have categorized big data applications or comprehensively discussed their future potentials and concerns.

To evaluate and compare our study with similar ones, we applied the criteria outlined in Table 14. The criteria included examining challenges (C1), future potentials (C2), domain categorization (C3), privacy concerns (C4), and specific domains such as healthcare (C5), supply chain and logistics (C6), marketing and advertising (C7), smart cities (C8), media and entertainment (C9), cybersecurity (C10), climate and earth science (C11), industry (C12), and education (C13). Table 15 shows the overall comparison of the related surveys literature.

Big data applications in healthcare have been extensively reviewed, focusing on the benefits and challenges in this domain. The study in Hong et al. (2018) offers a comprehensive overview of big data in healthcare, addressing challenges (C1) and exploring applications (C5). The authors emphasize the importance of privacy (C4) and regulatory frameworks. Subsequently, a study (Abouelmehdi et al. 2018) investigate the transformative potential of big data within the healthcare domain (C2), highlighting privacy concerns (C4). This study provides valuable insights into disease prediction and cost reduction (C5). Furthermore, authors in Rajabion et al. (2019) contribute to understanding data processing mechanisms in healthcare (C3). Lastly, study in Galetsi et al. (2019) emphasizes the value of personalized services in healthcare (C5), acknowledging privacy concerns (C4).

Big data applications within the supply chain and logistics domain have shown significant potential for optimization and efficiency improvements. The study in Torre-Bastida et al. (2018) offers a comprehensive overview of big data applications within the transportation industry, addressing challenges (C1) and exploring opportunities in routing, planning, monitoring, and network design (C6). Building upon this, authors in Nguyen et al. (2018) extend the analysis to the broader supply chain management domain (C6), proposing a classification framework and identifying research gaps (C2). Focusing on the railway sector, study (Ghofrani et al. 2018) contributes to the understanding of big data applications in operations, maintenance, and safety, leveraging Mayring's framework (C6). A broader perspective is offered by Mishra et al. (2018), which provides a bibliometric analysis of big data in supply chain management (C6), identifying key research clusters and managerial insights.

The application of big data in marketing and advertising has been explored to understand its impact on digital marketing strategies and customer engagement. The study by Miklosik and Evans (2020) delves into the application of big data and ML in the realm of digital marketing (C7), uncovering unexplored avenues for future research. Subsequently, authors in Anshari et al. (2019) explore the integration of big data into CRM, emphasizing

Table 14 Standards for evaluating studies

Symbol	Criteria	Criteria definition
C1	Addressing challenges	Study covering challenges within the relevant domain
C2	Future of big data	Examination of the study's coverage on the future trends and developments in big data
C3	Categorization of domains	Discussion on how the study categorizes and classifies different domains related to the big data
C4	Privacy concerns	Investigation into the privacy concerns within the relevant domain addressed by the study
C5	Healthcare	In-depth analysis and coverage of the healthcare sector within the study
C6	Supply chain and logistics	Addressing the role of big data in supply chain and logistics management
C7	Marketing and advertising	Focus on the study's exploration of big data's impact on marketing and advertising
C8	Smart cities	Exploration of the study's insights into the application of big data in smart city initiatives
C9	Media and entertainment	Encompassing coverage of big data's influence on media and entertainment industries
C10	Cybersecurity	Examination of the study's insights into the role of big data in cybersecurity measures
C11	Climate and earth science	Delving into the study's exploration of big data applications in climate and earth science
C12	Industry	Coverage of the study's insights into big data's impact on various industries
C13	Education	Addressing the study's coverage of big data applications in the field of education

Table 15 Comparative analysis of related studies

Paper	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
Hong et al. (2018)	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Abouelmehdi et al. (2018)	✓	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
Rajabion et al. (2019)	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
Galetsi et al. (2019)	✓	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
Nguyen et al. (2018)	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
Torre-Bastida et al. (2018)	✓	✓	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗
Ghofrani et al. (2018)	✓	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗
Mishra et al. (2018)	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
Karimi et al. (2021)	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
Mohammadi and Al (2018)	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
Huang et al. (2021)	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
Alani (2021)	✓	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗
Ullah and Babar (2019)	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗
Srivastava and Jaiswal (2019)	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗
Akter and Wamba (2019)	✓	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
Amani et al. (2020)	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗
Huang et al. (2018)	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
Qi (2020)	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗
Misra et al. (2020)	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
Mosavi et al. (2018)	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
Luan et al. (2020)	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Baig et al. (2020)	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Li and Jiang (2021)	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
Abkenar et al. (2021)	✓	✓	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗
Sebei et al. (2018)	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
Miklosik and Evans (2020)	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗
Anshari et al. (2019)	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
Badshah et al. (2023a, 2023b, 2023c)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

its role in personalized marketing strategies (C7). While survey papers like Kushwaha et al. (2021); Sestino et al. (2020), and Lee et al. (2023) offer comprehensive overviews of big data in marketing and advertising (C7).

Big data plays a crucial role in the development and management of smart cities, enhancing sustainability and livability. The study (Karimi et al. 2021) delves into the urban potential of AI within Smart Cities, emphasizing the integration of culture, metabolism, and governance for sustainability and livability (C8). It prioritizes the livability of the urban fabric alongside economic growth, showcasing the potential of AI and Big Data integration. In alignment with this perspective, authors in Mohammadi and Al (2018) conduct a comprehensive review of big data handling in smart cities, categorizing techniques and exploring key ideas (C3). The study introduces crucial factors such as scalability, time, availability, and accuracy, contributing to the understanding of big data's role in smart city development (C8). Similarly, the study in Huang et al. (2021) addresses the underutilized data in smart cities by proposing a three-level framework employing semi-supervised deep reinforcement learning to optimize control policies (C8). The interconnected studies

collectively contribute to a more holistic understanding and advancement of AI and big data applications in the context of smart cities.

The media and entertainment industry has been significantly impacted by big data, especially through social media platforms. Big data is revolutionizing the media and entertainment industry. A significant portion of this data is generated by social media platforms. The study in Abkenar et al. (2021) explores the types of SBD, laying the groundwork for understanding its potential applications in this domain (C9). While studies (Sebei et al. 2018) and Muhammad et al. (2018) contribute to the growing body of knowledge in this area (C9).

The application of BDA in cybersecurity is critical for enhancing security measures and protecting against cyber threats. The study in Alani (2021) surveys the applications of BDA in cybersecurity, covering areas such as intrusion detection, spamming detection, and cloud security (C10). It highlights the rapid increase in data generation due to the growing number of internet users. Building on this foundation, authors in Ullah and Babar (2019) and Srivastava and Jaiswal (2019) further explore the role of big data in cybersecurity, expanding the knowledge base in this domain (C10).

Big data has significant applications in earth sciences and disaster management, aiding in visualization, analysis, and prediction. The study by Akter and Wamba (2019) examines the application of big data in natural disaster management (C11), emphasizing visualization, analysis, and prediction. It highlights the role of emerging technologies in enhancing disaster response and recovery strategies. Expanding on this, Amani et al. (2020) delves into the utilization of Google Earth Engine (GEE) in various domains, including land classification, hydrology, and climate analysis (C11). Shifting focus to agriculture, authors in Huang et al. (2018) explore the application of big data in precision agriculture, addressing challenges and proposing a management framework. A comprehensive overview of big data in disaster management is presented in Akter and Wamba (2019), providing valuable insights into research trends, challenges, and future directions (C11).

BDA is revolutionizing industries, offering advanced analytics, optimization, decision-making, modelling, and predictions. BDA is revolutionizing industries, offering advanced analytics, optimization, decision-making, modelling, and predictions. The study in Mosavi et al. (2018) explores the adoption of big data technologies in the engineering domain, highlighting its role in enhancing competitiveness. It reviews academic literature on big data applications within the engineering field (C12). Expanding the focus to industry-specific challenges, Qi (2020) delves into the mining industry, addressing hurdles in implementing big data management (BDM) (C12). The study outlines data sources, challenges, and future prospects for the mining industry (C1, C2). Furthermore, Misra et al. (2020) explores the impact of IoT, big data, and AI on agri-food systems (C12). It covers applications across the supply chain, from agriculture to food quality assessment, emphasizing commercialization and translational research outcomes.

The exploration of big data applications in education reveals a growing body of research. Authors in Luan et al. (2020) delve into challenges and trends (C1, C2), advocating for a balanced approach to technology integration (C13). A study in Baig et al. (2020) contributes by analyzing 40 studies, focusing on learner behaviour and performance (C13), while the investigation in Li and Jiang (2021) examines the impact of COVID-19 on educational big data, highlighting the role of educational psychology (C13).

In the context of these contributions and the existing literature, this study represents a pioneering investigation that deeply probes big data applications, categorization, challenges, and potential futures. This collective exploration paints a comprehensive picture of the diverse applications and impacts of big data across various domains.

7 Conclusion

This research explored the dynamic landscape of Big Data applications, unveiling their profound impact across diverse domains. The literature is meticulously categorized into distinct segments: healthcare, supply chain and logistics, marketing and advertising, smart cities, media and entertainment, cybersecurity, climate and earth science, industry, and education. Furthermore, it examined the transformative effects on decision-making processes, emphasizing the role of data-driven insights in various domains. Challenges and issues related to Big Data are thoroughly investigated, and recommendations are presented to overcome these hurdles. Additionally, core technologies for storing, processing, and analyzing large datasets are explored. The study also identifies and addresses potential concerns within Big Data, offering robust solutions and effective mitigation strategies.Through a comprehensive comparative analysis with related surveys, this research highlights its unique contributions and superiority. These contributions collectively bridge the existing gap in collective analysis, providing a holistic perspective on multifaceted Big Data applications.

Declarations

References

Abbasi RA, Maqbool O, Mushtaq M, Aljohani NR, Daud A, Alowibdi JS, Shahzad B (2018) Saving lives using social media: analysis of the role of twitter for personal blood donation requests and dissemination. Telemat Inform 35(4):892–912

Abkenar SB, Kashani MH, Mahdipour E, Jameii SM (2021) Big data analytics meets social media: a systematic review of techniques, open issues, and future directions. Telemat Inform 57:101517

Abouelmehdi K, Beni-Hessane A, Khaloufi H (2018) Big healthcare data: preserving security and privacy. J Big Data 5(1):1–18

Adeghe EP, Okolo CA, Ojeyinka OT (2024) The role of big data in healthcare: a review of implications for patient outcomes and treatment personalization. World J Biol Pharm Health Sci 17(3):198–204

Ahmed A, Xi R, Hou M, Shah SA, Hameed S (2023) Harnessing big data analytics for healthcare: a comprehensive review of frameworks, implications, applications, and impacts. IEEE Access 11:112891–112928

Ajah IA, Nweke HF (2019) Big data and business analytics: trends, platforms, success factors and applications. Big Data Cognitive Comput 3(2):32

Akter S, Wamba SF (2019) Big data and disaster management: a systematic review and agenda for future research. Ann Oper Res 283:939–959

Al-Jumaili AHA, Muniyandi RC, Hasan MK, Paw JKS, Singh MJ (2023) Big data analytics using cloud computing based frameworks for power management systems: Status, constraints, and future recommendations. Sensors 23(6):2952

Alahakoon D, Nawaratne R, Xu Y, De Silva D, Sivarajah U, Gupta B (2020) Self-building artificial intelligence and machine learning to empower big data analytics in smart cities. Int Syst Front. https://doi.org/10.1007/s10796-020-10056-x

Alani MM (2021) Big data in cybersecurity: a survey of applications and future trends. J Reliab Intell Environ 7(2):85–114

Albqowr A, Alsharairi M, Alsoussi A (2024) Big data analytics in supply chain management: a systematic literature review. VINE J Inform Knowled Manag Syst 54(3):657–682

Allam Z, Dhunny ZA (2019) On big data, artificial intelligence and smart cities. Cities 89:80–91

Amaithi Rajan A, V V (2023) Systematic survey: secure and privacy-preserving big data analytics in cloud. J Comput Inform Syst 64:1–21

Amalina F, Hashem IAT, Azizul ZH, Fong AT, Firdaus A, Imran M, Anuar NB (2019) Blending big data analytics: review on challenges and a recent study. IEEE Access 8:3629–3645

Amani M, Ghorbanian A, Ahmadi SA, Kakooei M, Moghimi A, Mirmazloumi SM, Moghaddam SHA, Mahdavi S, Ghahremanloo M, Parsian S, Wu Q, Brisco B (2020) Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. IEEE J Select Topics Appl Earth Observ Remote Sens 13:5326–5350

Amazone (2023). Amazon redshift. https://aws.amazon.com/redshift/

Ameer S, Shah MA (2018) Exploiting big data analytics for smart urban planning. In 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), (pp. 1–5). IEEE

Amjad T, Sher M, Daud A (2012) A survey of dynamic replication strategies for improving data availability in data grids. Futur Gener Comput Syst 28(2):337–349

Amjad T, Daud A, Aljohani NR (2018) Ranking authors in academic social networks: a survey. Library Hi Tech 36(1):97–128

Anitha P, Vimala H, Shreyas J (2023) Comprehensive review on congestion detection, alleviation, and control for iot networks. J Netw Comput Appl. https://doi.org/10.1016/j.jnca.2023.103749

Anshari M, Almunawar MN, Lim SA, Al-Mudimigh A (2019) Customer relationship management and big data enabled: personalization & customization of services. Appl Comput Inf 15(2):94–101

Apache (2023a). Apache hadoop. https://hadoop.apache.org/

Apache (2023b). Apache nifi. https://Nifi.apache.org/

Apache (2023c). Apache spark. https://spark.apache.org/

Apache (2023d). Mongodb. https:www.mongodb.com

Badshah A (2023a) Cloud storage future and its opportunities

Badshah A (2023b) The raceof lead in ai: micrososft, google and openai

Badshah A (2023c) Why edge computing. https://afzalbadshah.com/index.php/2022/12/11/why-edge-computing/

Badshah A, Ghani A, Daud A, Chronopoulos AT, Jalal A (2022a) Revenue maximization approaches in IAAS clouds: research challenges and opportunities. Trans Emerg Telecommun Technol 33(7):e4492

Badshah A, Iwendi C, Jalal A, Hasan SSU, Said G, Band SS, Chang A (2022b) Use of regional computing to minimize the social big data effects. Comput Ind Eng 171:108433

Badshah, A., Nasralla, M.M., Jalal, A. and Farman, H., (2023a) September. smart education in smart cities: challenges and solution. In 2023 IEEE International Smart Cities Conference (ISC2) (pp. 01-08). IEEE. https://doi.org/10.1109/ISC257844.2023.10293615

Badshah A, Ghani A, Daud A, Jalal A, Bilal M, Crowcroft J (2023b) Towards smart education through internet of things: a survey. ACM Comput Surv 56(2):1–33

Badshah A, Daud A, Khan HU, Alghushairy O, Bukhari A (2024) Optimizing the over and underutilization of network resources during peak and off-peak hours. IEEE Access

Bag S, Rahman MS, Srivastava G, Shore A, Ram P (2023) Examining the role of virtue ethics and big data in enhancing viable, sustainable, and digital supply chain performance. Technol Forecasting and Social Change 186:122154

Baig MI, Shuib L, Yadegaridehkordi E (2020) Big data in education: a state of the art, limitations, and future research directions. Int J Educ Technol Higher Educ 17(1):1–23

Bansal S, Kumar P, Rawat S, Choudhury T (2018) Analysis and impact of social media and it's privacy on big data. In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), pages 248–253. IEEE

Beauvisage T, Beuscart J-S, Coavoux S, Mellet K (2023) How online advertising targets consumers: the uses of categories and algorithmic tools by audience planners. New Med Soc. https://doi.org/10.1177/14614448221146174

Bhure A, Desai S (2023) Exploring the intersection of big data and legal analytics: a survey of its application in the legal industry. J Big Data Technol Bus Anal (e-ISSN: 2583-7834) 2(1):5–14

Bibri SE (2019) On the sustainability of smart and smarter cities in the era of big data: an interdisciplinary and transdisciplinary literature review. J Big Data 6(1):1–64

Cassandra (2023). Apache cassandra. https://Cassandra.apache.org/

Chang V (2021) An ethical framework for big data and smart cities. Technol Forecast Social Change 165:120559

Chen B, Nie G, Jiang S, Hu N (2022a) Research on the big data-based product quality data package construction and application. In 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC), pages 1–6. IEEE

Chen L, Zhang Y, Wang Z (2022b) Logistics service supply chain model applying artificial intelligence and big data analysis. Secur Commun Netw 2022:1575813

Chen X (2022) High-concurrency big data precision marketing and advertising recommendation under 5g wireless communication network environment. J Sens 2022:7609555

Cockcroft S, Russell M (2018) Big data opportunities for accounting and finance practice and research. Aust Account Rev 28(3):323–333

Coursera (2023). Introduction to big data with spark hadoop. https://www.coursera.org/learn/introduction-to-big-data-with-spark-hadoop/home/week/1

Craig T, Ludloff ME (2011) Privacy and big data: the players, regulators, and stakeholders. " O'Reilly Media, Inc."

Dash S, Shakyawar SK, Sharma M, Kaushik S (2019) Big data in healthcare: management, analysis and future prospects. J Big Data 6(1):1–25

Daud A, Abbasi R, Muhammad F (2013) Finding rising stars in social networks. In Database Systems for Advanced Applications: 18th International Conference, DASFAA 2013, Wuhan, China, April 22-25, 2013. Proceedings, Part I 18, pp. 13–24. Springer

Del Vecchio P, Di Minin A, Petruzzelli AM, Panniello U, Pirri S (2018) Big data for open innovation in smes and large corporations: trends, opportunities, and challenges. Creativity Innov Manag 27(1):6–22

Del Vecchio P, Mele G, Siachou E, Schito G (2022) A structured literature review on big data for customer relationship management (crm): toward a future agenda in international marketing. Int Market Rev 39(5):1069–1092

Demirbaga U, Aujla GS (2022) Mapchain: a blockchain-based verifiable healthcare service management in iot-based big data ecosystem. IEEE Trans Netw Service Manag 19(4):3896–3907

Ding X, Gan Q, Shaker MP (2023) Optimal management of parking lots as a big data for electric vehicles using internet of things and long-short term memory. Energy 268:126613

Ducange P, Pecori R, Mezzina P (2018) A glimpse on big data analytics in the framework of marketing strategies. Soft Comput 22(1):325–342

Elhoseny H, Elhoseny M, Riad AM, Hassanien AE (2018) A framework for big data analysis in smart cities. In The international conference on advanced machine learning technologies and applications (AMLTA2018), pages 405–414. Springer

EOS (2023) Analyzing big earth data progress challenges opportunities. https://eos.org/editors-vox/analyzing-big-earth-data-progress-challenges-opportunities

Esfahani H, Tavasoli K, Jabbarzadeh A (2019) Big data and social media: a scientometrics analysis. Int J Data Netw Sci 3(3):145–164

Farchi F, Farchi C, Touzi B, Mabrouki C (2023) A comparative study on ai-based algorithms for cost prediction in pharmaceutical transport logistics. Acadlore Trans Mach Learn 2(3):129–141

Farley SS, Dawson A, Goring SJ, Williams JW (2018) Situating ecology as a big-data science: current advances, challenges, and solutions. BioScience 68(8):563–576

Favaretto M, De Clercq E, Elger BS (2019) Big data and discrimination: perils, promises and solutions: a systematic review. J Big Data 6(1):1–27

Fischer C, Pardos ZA, Baker RS, Williams JJ, Smyth P, Yu R, Slater S, Baker R, Warschauer M (2020) Mining big data in education: affordances and challenges. Rev Res Educ 44(1):130–160

Fosso Wamba S, Gunasekaran A, Papadopoulos T, Ngai E (2018) Big data analytics in logistics and supply chain management. Int J Logist Manag 29(2):478–484

França RP, Monteiro ACB, Arthur R, Iano Y (2021) The fundamentals and potential for cybersecurity of big data in the modern world. In: Maleh Y, Shojafar M, Alazab M, Baddi Y (eds) Machine intelligence and big data analytics for cybersecurity applications. Studies in computational intelligence. Springer, Cham. https://doi.org/10.1007/978-3-030-57024-8_3

Galetsi P, Katsaliaki K, Kumar S (2019) Values, challenges and future directions of big data analytics in healthcare: a systematic review. Social Sci Med 241:112533

Galletta A, Carnevale L, Bramanti A, Fazio M (2018) An innovative methodology for big data visualization for telemedicine. IEEE Trans Indus Inform 15(1):490–497

Ghani NA, Hamid S, Hashem IAT, Ahmed E (2019) Social media big data analytics: a survey. Comput Human Behav 101:417–428

Ghofrani F, He Q, Goverde RM, Liu X (2018) Recent applications of big data analytics in railway transportation systems: a survey. Transport Res Part C: Emerging Technol 90:226–246

Google (2023) Google big data query. https://cloud.google.com/bigquery

Hao W (2021) Empirical study on the application of flipper classroom innovation teaching under the context of big data. In 2021 International Conference on Computer Technology and Media Convergence Design (CTMCD), pages 138–142

Hariri RH, Fredericks EM, Bowers KM (2019) Uncertainty in big data analytics: survey, opportunities, and challenges. J Big Data 6(1):1–16

Hassani H, Huang X, Silva E (2018) Digitalisation and big data mining in banking. Big Data Cognit Comput 2(3):18

Hassani H, Huang X, Silva E (2019) Big data and climate change. Big Data Cognit Comput 3(1):12

Hayat MK, Daud A, Alshdadi AA, Banjar A, Abbasi RA, Bao Y, Dawood H (2019) Towards deep learning prospects: insights for social media analytics. IEEE Access 7:36958–36979

He X, Wang K, Huang H, Liu B (2018) Qoe-driven big data architecture for smart city. IEEE Commun Mag 56(2):88–93

Himeur Y, Elnour M, Fadli F, Meskin N, Petri I, Rezgui Y, Bensaali F, Amira A (2023) Ai-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artif Intell Rev 56(6):4929–5021

Hong L, Luo M, Wang R, Lu P, Lu W, Lu L (2018) Big data in health care: applications and challenges. Data Inform Manag 2(3):175–197

Hong W, Xiong Z, Zheng N, Weng Y (2019) A medical-history-based potential disease prediction algorithm. IEEE Access 7:131094–131101

Hu G, Liu W, Xu H (2022) Research on hybrid teaching assessment driven by big data. In 2022 2nd International Conference on Big Data Engineering and Education (BDEE), pages 206–209

Huang YanBo HY, Chen ZhongXin CZ, Yu Tao YT, Huang XiangZhi HX, Gu XingFa GX (2018) Agricultural remote sensing big data: management and applications. J Integr Agricul 17(9):1915–1931

Huang H, Yao XA, Krisp JM, Jiang B (2021) Analytics of location-based big data for smart cities: opportunities, challenges, and future directions. Comput Environ Urban Syst 90:101712

Ikegwu AC, Nweke HF, Anikwe CV, Alo UR, Okonkwo OR (2022) Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. Cluster Comput 25(5):3343–3387

Ikegwu AC, Nweke HF, Mkpojiogu E, Anikwe CV, Igwe SA, Alo UR (2024) Recently emerging trends in big data analytic methods for modeling and combating climate change effects. Energy Inform 7(1):6

Ilieva RT, McPhearson T (2018) Social-media data for urban sustainability. Natu Sustain 1(10):553–565

Inayatulloh Prabowo H, Warnars H LHS, Napitupulu TA, Khairil, Deviarti H (2022) Extended e-learning model to support home schooling with collaboration between teacher, parents and student. In 2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM), (pp. 1–6)

Insider (2023) Last mile delivery shipping explained. https://www.insiderintelligence.com/insights/last-mile-delivery-shipping-explained/

Jahani H, Jain R, Ivanov D (2023) Data science and big data analytics: a systematic review of methodologies used in the supply chain and logistics research. Ann Oper Res. https://doi.org/10.1007/s10479-023-05390-7

Jia Y, Chao K, Cheng X, Xu L, Zhao X, Yao L (2019) Telecom big data based precise user classification scheme. In 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of

People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), (pp. 1517–1520). IEEE

Jiang W (2019) An intelligent supply chain information collaboration model based on internet of things and big data. IEEE Access 7:58324–58335

Jieyu L (2020) Research on network advertisement precise delivery system based on big data technology. In 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), pages 794–797. IEEE

Jimenez-Marquez JL, Gonzalez-Carrasco I, Lopez-Cuadrado JL, Ruiz-Mezcua B (2019) Towards a big data framework for analyzing social media content. Int J of Inform Manag 44:1–12

Jiyang Y, Yanbin Z, Jian G (2020) Research on the development of intelligent chemical manufacturing industry in shandong province based on big data analysis. In 2020 2nd International Conference on Industrial Artificial Intelligence (IAI), (pp. 1–6). IEEE

Kandt J, Batty M (2021) Smart cities, big data and urban policy: Towards urban analytics for the long run. Cities 109:102992

Kantarcioglu M, Xi B (2016) Adversarial data mining: Big data meets cyber security. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, (pp. 1866–1867)

Kanth R, Laakso M-J, Nevalainen P, Heikkonen J (2018) Future educational technology with big data and learning analytics. In 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE), (pp. 906–910)

Karimi Y, Haghi Kashani M, Akbari M, Mahdipour E (2021) Leveraging big data in smart cities: a systematic review. Concurr Comput: Pract Exp 33(21):e6379

Khan J, Ahmad N (2023) Security and privacy technique in big data: A review. In 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), (pp. 1575–1579)

Khan W, Daud A, Nasir JA, Amjad T (2016) A survey on the state-of-the-art machine learning models in the context of nlp. Kuwait J Sci, 43(4)

Knüsel B, Zumwald M, Baumberger C, Hirsch Hadorn G, Fischer EM, Bresch DN, Knutti R (2019) Applying big data beyond small problems in climate research. Natu Climate Change 9(3):196–202

Kushwaha AK, Kar AK, Dwivedi YK (2021) Applications of big data in emerging management disciplines: a literature review using text mining. Int J Inf Manage Data Insights 1(2):100017

Laney D et al (2001) 3d data management: controlling data volume, velocity and variety. META Group Res Note 6(70):1

Learning S (2023) Scikit learning. https://scikit-learn.org/stable/

Lee SE, Ju N, Lee K-H (2023) Service chatbot: Co-citation and big data analysis toward a review and research agenda. Int J Inf Manage Data Insights 194:122722

Leng P, Xiang L, Lin Y, Xiao W, Yang Z, Li D, Nai W (2020) Logistic regression based on artificial fish swarm algorithm with t-distribution parameters. In 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), volume 9, pages 1912–1915

Leung, C.K., Braun, P., Hoi, C.S., Souza, J. and Cuzzocrea, A., 2019. Urban analytics of big transportation data for supporting smart cities. In Big Data Analytics and Knowledge Discovery: 21st International Conference, DaWaK 2019, Linz, Austria, August 26–29, 2019, Proceedings 21 (pp. 24-33). Springer International Publishing.

Li B, Zhao S, Zhang R, Shi Q, Yang K (2019a) Anomaly detection for cellular networks using big data analytics. IET Commun 13(20):3351–3359

Li F, Xie R, Wang Z, Guo L, Ye J, Ma P, Song W (2019b) Online distributed iot security monitoring with multidimensional streaming big data. IEEE Interne of Things J 7(5):4387–4394

Li C, Chen Y, Shang Y (2022) A review of industrial big data for decision making in intelligent manufacturing. Eng Sci Technol Int J 29:101021

Li J, Jiang Y (2021) The research trend of big data in education and the impact of teacher psychology on educational development during covid-19: a systematic review and future perspective. Front Psychol 12:753388

Liang J (2020) Research on the application of big data in the informatization of higher education management mode. In 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), (pp. 799–802)

Lies J (2019) Marketing intelligence and big data: digital marketing techniques on their way to becoming social engineering techniques in marketing

Lin L, Zhou D, Wang J, Wang Y (2024) A systematic review of big data driven education evaluation. SAGE Open 14(2):21582440241242180

Liu X, You J (2021) Research on the impact of big data application on technological innovation of chinese new energy vehicle industry. In 2021 2nd International Conference on Big Data Economy and Information Management (BDEIM), pages 323–327. IEEE

Liu X, Shin H, Burns AC (2021) Examining the impact of luxury brands' social media marketing on customer engagement: using big data analytics and natural language processing. J Bus Res 125:815–826

Luan H, Geczy P, Lai H, Gobert J, Yang SJ, Ogata H, Baltes J, Guerra R, Li P, Tsai C-C (2020) Challenges and future directions of big data and artificial intelligence in education. Front Psychol 11:580820

Lwin KK, Sekimoto Y, Takeuchi W, Zettsu K (2019) City geospatial dashboard: Iot and big data analytics for geospatial solutions provider in disaster management. In 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), pages 1–4

Lyu X, Zhao J (2019) Compressed sensing and its applications in risk assessment for internet supply chain finance under big data. IEEE Access 7:53182–53187

Makkie M, Li X, Quinn S, Lin B, Ye J, Mon G, Liu T (2018) A distributed computing platform for fmri big data analytics. IEEE Trans Big Data 5(2):109–119

Mani Z, Chouk I (2022) Impact of privacy concerns on resistance to smart services: does the 'big brother effect'matter? In the role of smart technologies in decision making, (pp. 94–113). Routledge

Masood I, Wang Y, Daud A, Aljohani NR, Dawood H (2018a) Privacy management of patient physiological parameters. Telem Inform 35(4):677–701

Masood I, Wang Y, Daud A, Aljohani NR, Dawood H (2018b) Towards smart healthcare: patient data privacy and security in sensor-cloud infrastructure. Wireless Commun Mobile Comput 2018:1–23

Mehta N, Pandit A (2018) Concurrence of big data analytics and healthcare: a systematic review. Int J Med Inform 114:57–65

Microsoft (2023) Power bi. https://www.microsoft.com/en-us/power-platform/products/power-bi

Miklosik A, Evans N (2020) Impact of big data and machine learning on digital transformation in marketing: a literature review. IEEE Access 8:101284–101292

Mishra D, Gunasekaran A, Papadopoulos T, Childe SJ (2018) Big data and supply chain management: a review and bibliometric analysis. Ann Oper Res 270:313–336

Misra N, Dixit Y, Al-Mallahi A, Bhullar MS, Upadhyay R, Martynenko A (2020) Iot, big data, and artificial intelligence in agriculture and food industry. IEEE Internet Things J 9(9):6305–6324

Mohammadi M, Al A (2018) Enabling cognitive smart cities using big data and machine learning: approaches and challenges. IEEE Commun Mag 56(2):94–101

Mohammadpoor M, Torabi F (2020) Big data analytics in oil and gas industry: an emerging trend. Petroleum 6(4):321–328

Mosavi A, Lopez A, Varkonyi-Koczy AR (2018) Industrial applications of big data: state of the art survey. In Recent Advances in Technology Research and Education: Proceedings of the 16th International Conference on Global Research and Education Inter-Academia 2017 16, pages 225–232. Springer

Muhammad SS, Dey BL, Weerakkody V (2018) Analysis of factors that influence customers' willingness to leave big data digital footprints on social media: a systematic review of literature. Inf Syst Front 20:559–576

Munshi AA, Alhindi A (2021) Big data platform for educational analytics. IEEE Access 9:52883–52890

Nguyen T, Zhoul L, Spiegler V, Ieromonachou P, Lin Y (2018) Big data analytics in supply chain management: a state-of-the-art literature review. Comput Oper Res 98:254–264

Nguyen T, Gosine RG, Warrian P (2020) A systematic review of big data analytics for oil and gas industry 4.0. IEEE Access 8:61183–61201

Patel D, Shah D, Shah M (2020) The intertwine of brain and body: a quantitative analysis on how big data influences the system of sports. Ann Data Sci 7:1–16

Philip NY, Razaak M, Chang J, M, S., O'Kane, M., and Pierscionek, B. K. (2022) A data analytics suite for exploratory predictive, and visual analysis of type 2 diabetes. IEEE Access 10:13460–13471

Price WN, Cohen IG (2019) Privacy in the age of medical big data. Natu Med 25(1):37–43

Qi C-C (2020) Big data management in the mining industry. Int J Mineral Metall Mater 27:131–139

Qian R, Sengan S, Juneja S (2022) English language teaching based on big data analytics in augmentative and alternative communication system. Int J Speech Technol 25(2):409–420

Rahman MS, Reza H (2022) A systematic review towards big data analytics in social media. Big Data Min Anal 5(3):228–244

Rajabion L, Shaltooki AA, Taghikhah M, Ghasemi A, Badfar A (2019) Healthcare big data processing mechanisms: the role of cloud computing. Int J Inf Manage 49:271–289

Rao SUM, Lakshmanan L (2024) Securing communicating networks in the age of big data: an advanced detection system for cyber attacks. Opt Quantum Electron 56(1):116

Rassam MA, Maarof M, Zainal A, et al (2017) Big data analytics adoption for cybersecurity: A review of current solutions, requirements, challenges and trends. J Inf Assur Secur, 12(4)

Rawat DB, Doku R, Garuba M (2019) Cybersecurity in big data era: from securing big data to data-driven security. 14:2055–2072. IEEE

Rawat R, Oki OA, Sankaran KS, Olasupo O, Ebong GN, Ajagbe SA (2023) A new solution for cyber security in big data using machine learning approach. In Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2023, pages 495–505. Springer

Ray S, Saeed M (2018) Applications of educational data mining and learning analytics tools in handling big data in higher education. In: Alani M, Tawfik H, Saeed M, Anya O (eds) Applications of big data analytics. Springer, Cham. https://doi.org/10.1007/978-3-319-76472-6_7

Rehman GU, Zubair M, Qasim I, Badshah A, Mahmood Z, Aslam M, Jilani SF (2022) Ems: Efficient monitoring system to detect non-cooperative nodes in iot-based vehicular delay tolerant networks (vdtns). Sensors 23(1):99

Research A, Consulting (2023) Big data market size to reach usd 473.6 billion by 2030. https://www.acumenresearchandconsulting.com/press-releases/big-data-market

Saravanan S, Prakash G (2021) A comprehensive survey on big data technology based cybersecurity analytics systems. Applied soft computing and communication networks: proceedings of ACN 2020:123–143

Sarker MNI, Peng Y, Yiran C, Shouse RC (2020a) Disaster resilience through big data: way to environmental sustainability. Int J Disaster Risk Reduct 51:101769

Sarker MNI, Yang B, Yang L, Huq ME, Kamruzzaman M (2020b) Climate change adaptation and resilience through big data. Int J Adv Comput Sci Appl. https://doi.org/10.14569/ijacsa.2020.0110368

Sebei H, Hadj Taieb MA, Ben Aouicha M (2018) Review of social media analytics process and big data pipeline. Social Netw Anal Min 8(1):30

Sebestyén V, Czvetkó T, Abonyi J (2021) The applicability of big data in climate change research: the importance of system of systems thinking. Front Environ Sci 9:70

Selmy HA, Mohamed HK, Medhat W (2023) Big data analytics deep learning techniques and applications: a survey. Inform Syst. https://doi.org/10.1016/j.is.2023.102318

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

Shah SA, Seker DZ, Rathore MM, Hameed S, Yahia SB, Draheim D (2019) Towards disaster resilient smart cities: can internet of things and big data analytics be the game changers? IEEE Access 7:91885–91903

Silva BN, Khan M, Jung C, Seo J, Muhammad D, Han J, Yoon Y, Han K (2018) Urban planning and smart city decision management empowered by real-time data processing using big data analytics. Sensors 18(9):2994

Sivarajah U, Irani Z, Gupta S, Mahroof K (2020) Role of big data and social media analytics for business to business sustainability: a participatory web context. Indus Market Manag 86:163–179

Srivastava N, Jaiswal UC (2019) Big data analytics technique in cyber security: a review. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), (pp. 579–585). IEEE

Statista (2023). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. https://www.statista.com/statistics/871513/worldwide-data-created/

Stegenga SM, Steltenpohl CN, Lustick H, Meyer MS, Renbarger R, Standiford Reyes L, Lee LE (2024) Qualitative research at the crossroads of open science and big data: ethical considerations. Social Personal Psychol Compass 18(1):12912

Subroto A, Apriyana A (2019) Cyber risk prediction through social media big data analytics and statistical machine learning. J Big Data 6(1):50

Sudmanns M, Tiede D, Lang S, Bergstedt H, Trost G, Augustin H, Baraldi A, Blaschke T (2019) Big earth data: disruptive changes in earth observation data management and analysis? Int J Digital Earth 13(7):832–850

Tableau (2023). Tableau. https://www.tableau.com/

Talaoui Y, Kohtamäki M, Ranta M, Paroutis S (2023) Recovering the divide: a review of the big data analytics—strategy relationship. Long Range Plan 56(2):102290

Talend (2023). Talent. https://www.talend.com/

Tang L, Li J, Du H, Li L, Wu J, Wang S (2022) Big data in forecasting research: a literature review. Big Data Res 27:102290

Tao H, Bhuiyan MZA, Rahman MA, Wang G, Wang T, Ahmed MM, Li J (2019) Economic perspective analysis of protecting big data security and privacy. Futur Gener Comput Syst 98:660–671

Teneiji AL, Salim TYA, Riaz Z (2024) Factors impacting the adoption of big data in healthcare: a systematic literature review. Int J Med Inform. https://doi.org/10.1016/j.ijmedinf.2024.105460

Tensor (2023). Tensorflow. https://www.tensorflow.org/

Thilagavathi C, Rajeswari M, Sheethal M, Devassy D, Priya K, Divya R (2019) Security issues on inter-net of things in smart cities. In Handbook of Research on Implementation and Deployment of IoT Projects in Smart Cities, (pp. 149–164). IGI Global

Tian L, Wang H, Zhou Y, Peng C (2018) Video big data in smart city: background construction and opti-mization for surveillance video processing. Futur Gener Comput Syst 86:1371–1382

Tohka J, Van Gils M (2021) Evaluation of machine learning algorithms for health and wellness applica-tions: a tutorial. Comput Biol Med 132:104324

Torre-Bastida AI, Del Ser J, Laña I, Ilardia M, Bilbao MN, Campos-Cordobés S (2018) Big data for transportation and mobility: recent advances, trends and challenges. IET Intell Trans Syst 12(8):742–755

Ullah F, Babar MA (2019) Architectural tactics for big data cybersecurity analytics systems: a review. J Syst Softw 151:81–118

Ullah F, Babar MA, Aleti A (2022) Design and evaluation of adaptive system for big data cyber security analytics. Expert Syst Appls 207:117948

Vargo CJ, Guo L, Amazeen MA (2018) The agenda-setting power of fake news: a big data analysis of the online media landscape from 2014 to 2016. New Med Society 20(5):2028–2049

Vassakis K, Petrakis E, Kopanakis I (2018) Big data analytics: Applications, prospects and challenges. A roadmap from models to technologies, Mobile big data, pp 3–20

Ved M, B, R (2019) Big data analytics in telecommunication using state-of-the-art big data framework in a distributed computing environment: A case study. In 2019 IEEE 43rd Annual Computer Soft-ware and Applications Conference (COMPSAC), volume 1, pages 411–416

Walters R, Novak M (2021) Cyber Security, Artificial Intelligence. Springer, Data Protection & the Law

Wang L, Jones R (2021) Big data analytics in cyber security: network traffic and attacks. J Comput Inform Syst 61(5):410–417

Wang X, Yang LT, Chen X, Deen MJ, Jin J (2018) Improved multi-order distributed hosvd with its incre-mental computing for smart city services. IEEE Trans Sustain Comput 6(3):456–468

Wu SM, Chen T-C, Wu YJ, Lytras M (2018) Smart cities in Taiwan: a perspective on big data applica-tions. Sustainability 10(1):106

Wu M, Hong L, Zhao Y, Chen L, Wang J (2019) Dosage prediction in pediatric medication leveraging prescription big data. IEEE Access 7:94285–94292

Yadav SS, Jadhav SM (2019) Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data 6(1):1–18

Yan L, Huang W, Wang L, Feng S, Peng Y, Peng J (2019) Data-enabled digestive medicine: a new big data analytics platform. IEEE/ACM Trans Comput Biol Bioinform 18(3):922–931

Yang C, Yu M, Li Y, Hu F, Jiang Y, Liu Q, Sha D, Xu M, Gu J (2019) Big earth data analytics: a survey. Big Earth Data 3(2):83–107

Yang C, Lan S, Zhao Z, Zhang M, Wu W, Huang GQ (2022) Edge-cloud blockchain and ioe-enabled quality management platform for perishable supply chain logistics. IEEE Internet Things J 10(4):3264–3275

Yin P, Huang H, Zhao M, Zhu Y (2021) Application of big data marketing in customer relationship man-agement. In Proceedings of the 2021 5th International Conference on E-Education, E-Business and E-Technology, (p. 1)

Yu M, Yang C, Li Y (2018) Big data in natural disaster management: a review. Geosciences 8(5):165

Yuwen Z, Changqin H, Qintai H, Jia Z, Yong T (2018) Personalized learning full-path recommendation model based on lstm neural networks. Inform Sci 444:135–152

Zhang J (2022) Application of computer big data and cloud computing technology in the promotion of e-commerce advertising. In 2022 IEEE 2nd International Conference on Data Science and Com-puter Application (ICDSCA), pages 834–837. IEEE

Zhang X, Ghorbani AA (2021) Human factors in cybersecurity: issues and challenges in big data. Res Anthol Privat Secur Data, (pp. 1695–1725)

Zhang D, Wang D, Vance N, Zhang Y, Mike S (2018) On scalable and robust truth discovery in big data social media sensing applications. IEEE Trans Big Data 5(2):195–208

Zhang N, Geng B, Hu W, Wen R (2021) The applications of big data analysis in student management education. In 2021 2nd International Conference on Big Data and Informatization Education (ICBDIE), pages 55–58

Zhang H, Zang Z, Zhu H, Uddin MI, Amin MA (2022) Big data-assisted social media analytics for business model for business decision making system competitive analysis. Inform Process Manag 59(1):102762

Zhang Y, Hong J, Chen S (2023) Medical big data and artificial intelligence for healthcare. Appl Sci 13(6):3745

Zhili Wang, M., Huang, J., Lin, S., and Lv, Z. (2021) Blockchain in big data security for intelligent transportation with 6g. IEEE Transactions on Intelligent Transportation Systems 23(7):9736–9746

Zhou S, He J, Yang H, Chen D, Zhang R (2020) Big data-driven abnormal behavior detection in healthcare based on association rules. IEEE Access 8:129002–129011

Zhou R, Zhang X, Wang X, Yang G, Guizani N, Du X (2021) Efficient and traceable patient health data search system for hospital management in smart cities. IEEE Internet Things J 8(8):6425–6436

Zhu S, Du G (2022) Evaluation of the service capability of maritime logistics enterprises based on the big data of the internet of things supply chain system. IEEE Consum Electron Mag 12(2):100–108

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.