

SURVEY

Open Access



# 15 years of Big Data: a systematic literature review

Davide Tosi<sup>1\*</sup>, Redon Kokaj<sup>1</sup> and Marco Rocchetti<sup>2</sup>

\*Correspondence:  
davide.tosi@uninsubria.it

<sup>1</sup> Department of Theoretical and Applied Sciences, Università degli Studi dell'Insubria, Via Mazzini 5, 21100 Varese, Italy

<sup>2</sup> Department of Computer Science and Engineering, University of Bologna, Mura Anteo Zamboni 7, 40126 Bologna, Italy

## Abstract

Big Data is still gaining attention as a fundamental building block of the Artificial Intelligence and Machine Learning world. Therefore, a lot of effort has been pushed into Big Data research in the last 15 years. The objective of this Systematic Literature Review is to summarize the current state of the art of the previous 15 years of research about Big Data by providing answers to a set of research questions related to the main application domains for Big Data analytics; the significant challenges and limitations researchers have encountered in Big Data analysis, and emerging research trends and future directions in Big Data. The review follows a predefined procedure that automatically searches five well-known digital libraries. After applying the selection criteria to the results, 189 primary studies were identified as relevant, of which 32 were Systematic Literature Reviews. Required information was extracted from the 32 studies and summarized. Our Systematic Literature Review sketched the picture of 15 years of research in Big Data, identifying application domains, challenges, and future directions in this research field. We believe that a substantial amount of work remains to be done to align and seamlessly integrate Big Data into data-driven advanced software solutions of the future.

**Keywords:** Big Data, Systematic literature review, SLR, Data analysis, Artificial intelligence

## Introduction

Over the past 15 years, Big Data has emerged as a foundational pillar providing support to an extensive range of different scientific fields, from medicine and healthcare [1] to engineering [2], finance and marketing [3–5], politics [6], social networks analysis [7, 8], and telecommunications [9], to cite only a few examples. This 15-year period has witnessed a significant increase in research efforts aimed at unraveling the major problems in Big Data, with an almost innumerable array of potential solutions and data sources [10–13]. This has resulted in a boundless world of scientific papers that, in the end, have demonstrated the twofold, ambivalent nature of Big Data. On one side, in fact, we have had a confirmation of the pivotal role played by this scientific field in shaping the technological advancements of our time. On the other side, an approach to the comprehension of Big Data, based on this endless universe of ten of thousand technical papers, each specializing in its specific sector, however natural it might

seem, has become not sustainable because it has often made researchers confuse (or mixing) the theory (of Big Data) with the practice or use of it. We cannot ignore that there have also been numerous active attempts to describe the general landscape of Big Data through survey papers. Nonetheless, again, given the vastness of the subject, the majority of them did not shun the trap of pre-formed models and have tried to respond, as closely as possible, to the concrete requirements coming from just one sub-field or from the point of view of a few perspectives. In this complex context, to take at least one step further into the knowledge of the state of the art of Big Data research over the above-mentioned period of time, we have decided to conduct a different form of comprehensive exploration which was not biased by the specificity of some given sectors or confounded by single technical perspectives. To do that, we have adopted the methodology termed systematic literature review (SLR), as proposed by Kitchenham and Charters [14] in the field of software engineering [15, 16]. Although SLR proceeds through a set of well-defined steps, also in this case, an initial choice has to be made regarding the most crucial parameters through which the subject of investigation should be explored. In the case of Big Data, our primary focus has been on gaining insights into the principal application domains of Big Data, unraveling the major challenges and limitations encountered by researchers in the analysis of the typically enormous datasets they manage, and unveiling the emerging trends and directions in future Big Data research.

Guided by the structured methodology imposed by SLR, we hence started with three research questions that matched the points raised before: essentially, (i) most common application domains, (ii) current research challenges and limitations, and (iii) emerging future trends and directions. From this point on, we proceeded following the SLR steps. Basically: first, we translated the three research questions above into specific search terms, through which five different digital libraries were investigated, namely: *Scopus*, *IEEE Explore*, *ACM Digital Library*, *SpringerLink* and *Google Scholar*. Upon completion of the search activity (detailed in the following section of this paper), 189 primary studies that matched our generic search criteria were identified. Of these 189, only 32 of these studies were actually reviews. Since the target of our study was to provide a panoramic view of this 15-year Big Data research period, (a) shedding light on the prevalent application domains, (b) highlighting the hurdles faced by researchers, and (c) finally outlining the potential trajectories for future research, we focused on the analysis of just these 32 survey studies.

With this paper, we do not want to conduct a traditional literature review on a very extensive topic like Big Data. Traditional scientific surveys can include many more studies and corresponding papers, and they are mainly built with an eye toward generalizability and inclusion rather than selectivity and relevance. As a consequence, those approaches often bring to us no much more than a mere summary of the topic of interest. SLRs, instead, start from the legitimate presumption to be more than merely a summary of a topic. In essence, they distinguish themselves from ordinary surveys of the available literature because they are specifically built to add to the identification of all publications on a topic also all the following activities: explicit formulation of a search objective, identification and description of a search procedure, definition of criteria for inclusion and exclusion of publications, literature selection, and information extraction

only based on a transparent evaluation of the quality of publications. Not only this, but an SLR should also provide insightful information on the current state of research on a topic, starting from a given set of research questions and following a formal methodological procedure, designed to reduce distortions caused by an overly generous and restrictive selection of the literature, while guaranteeing the reliability of the selected publications. Hence, to pursue these objectives, an SLR should start with the definition of the criteria for determining what should be included/excluded before conducting the search. Not to mention that, typically, an SLR should be performed mainly using electronic literature databases. It should be also noticed that such a structured approach should document all the information gathered (and the steps taken as part of this process), with the aim of making the paper selection process completely visible and reproducible [17].

In the end, we know very well that a point-to-point analysis of the set of almost 320 papers from which we have started our SLR could have brought more (generic) information than that provided by the circa 30 papers finally selected by our SLR. Nonetheless, it is highly likely that this information would have been somewhat redundant, more prone to defects and personal biases, and finally, also more boring to read.

With this SLR, we aim to contribute, in a focused and structured way, to Big Data research in several ways: from one side, we provide researchers with a clear picture of how Big Data application domains changed over time; then, we highlight challenges faced by academia and industry and their evolution in the last 15 years; finally, we sketch a set of open points that researchers will take into consideration in the next future.

We can conclude that, while our collective understanding of Big Data has grown after this investigation, this analysis has underscored again the fact that in this field, a kind of optimal stability emerges in terms of research interests through the even distribution among applications domains/challenges/future trends. From one side, we observe a pervasive adoption of Big Data solutions in all everyday life domains (such as Energy [18], Smart Cities [9], and Healthcare [19].) On the other hand, researchers have spent a lot of effort managing data quality, designing and developing advanced frameworks to manage Big Data in real-time, focusing on security and privacy. However, many challenges still remain open to seamlessly integrate Big Data into data-driven advanced software solutions of the future, such as mitigating energy consumption, optimizing algorithms, increasing framework security with privacy and ethical focus, intersecting Artificial Intelligence and Machine Learning technologies, opening data sets, improving interoperability among different stakeholders, and considering societal and business changes.

The remainder of this paper is organized as follows: in Sect "[Research method](#)", we run the SLR methodology on our Big Data use case (with the definition of our research questions, the search strategy, the inclusion/exclusion criteria, the study quality assessment questionnaire, and the data extraction from primary studies). All this is in the dual attempt to explain the abstract methodology, as well as its application in our field. Section "[SLR: implementation](#)" describes how we conducted the review and the results obtained in each stage and step of our SLR; Section "[SLR: results](#)" shows our findings, briefly summarizing each of our selected primary studies; Section "[Discussion](#)" discusses critically those findings garnering special attention in our analytical process;

Section "Threats to validity" discusses the possible threats to the validity of our study; Section "Conclusion" demonstrates the conclusions we drew for our SLR.

A taxonomy of key concepts for Big Data evolution over the last 15 years is presented in Fig. 1.

## Research method

### Research questions

This SLR has been conducted following the procedure defined by Kitchenham and Charters. As such, in the first step, we defined the research questions (RQ) that will drive the entire review methodology.

As we define the research questions that will guide our SLR, it is crucial to establish a balance between the breadth and depth of our investigation. After careful consideration and to ensure that our review maintains a focused and meaningful scope, it has been decided to narrow down our research questions to the following three:

- RQ1: what are the most common application domains for Big Data analytics, and how have they evolved over time?
- RQ2: what are the major challenges and limitations that researchers have encountered in Big Data analysis, and how have they been addressed?
- RQ3: what are the emerging research trends and directions in Big Data that will likely shape the field in the next 5 to 10 years?

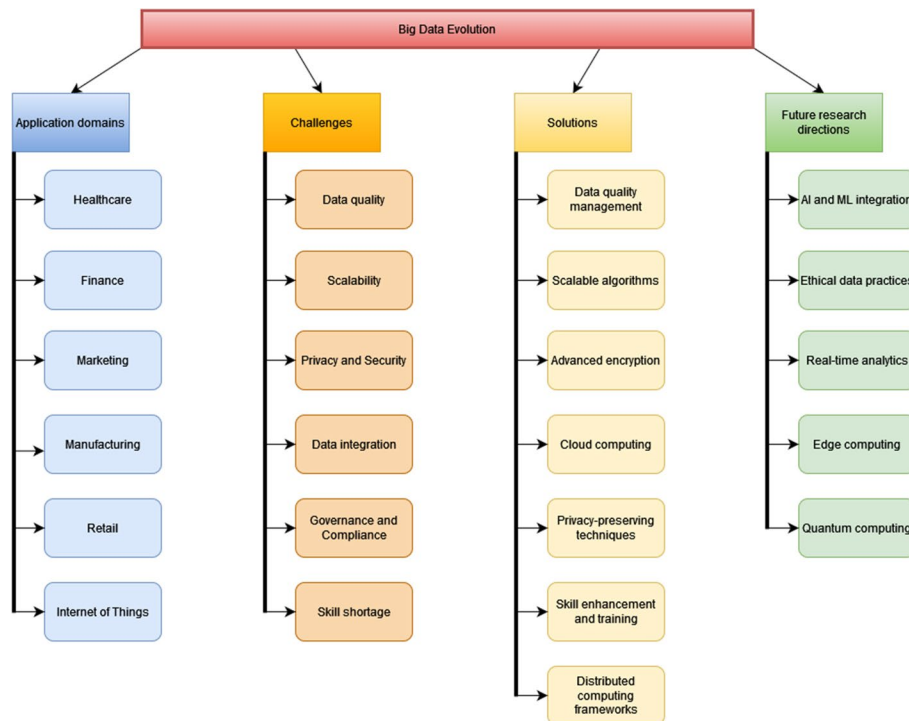


Fig. 1 Taxonomy of Big Data evolution over the last 15 years

### Search strategy

SLR begins by looking for relevant studies related to our research questions. To do this, we find appropriate search terms using the method outlined by Kitchenham and Charters, which suggests to consider three aspects: Population (P), Interventions (I), and Outcomes (O).

We identified the following relevant search terms for each aspect in our review:

- *Population*: Big Data, real-time data analytics, large datasets.
- *Intervention*: methodologies, techniques, domains, architectures, solutions.
- *Outcomes*: research trends, future directions, emerging technologies, challenges, SLR, Systematic Literature Review.

The search string was constructed as follows:

$$(P_1 \text{ OR } P_2 \dots \text{ OR } P_n) \text{ AND } (I_1 \text{ OR } I_2 \dots \text{ OR } I_n) \text{ AND } (O_1 \text{ OR } O_2 \dots \text{ OR } O_n)$$

P refers to population terms, I refers to intervention terms and O refers to outcome terms, all of which are connected through boolean operators AND and OR.

Searches string may take the exemplar form like the following:

("big data" OR "real-time data analytics" OR "large datasets") AND ("methodologies" OR "techniques" OR "domains" OR "architectures" OR "solutions") AND ("research trends" OR "future directions" OR "emerging technologies" OR "challenges" OR "SLR" OR "Systematic Literature Review")

Since we need to find and study primary studies related to our research questions, the selection of appropriate digital libraries/search engines to search for the articles needed is essential. For this reason, it has been decided to use the following state-of-the-art sources:

- *Scopus*: a multidisciplinary database that covers a broad range of research fields.
- *IEEE Xplore*: an invaluable resource for technology and engineering-related SLR.
- *ACM Digital Library*: a comprehensive collection of relevant articles, conference papers, and journals focused on computer science and information technology.
- *SpringerLink*: an extensive collection of academic articles in the fields that align closely with our research interests.
- *Google Scholar*: a freely accessible web search engine that indexes scholarly literature across various disciplines.

We aim to ensure a comprehensive and focused literature search by utilizing these sources, thereby facilitating a thorough and methodical research.

### Inclusion/Exclusion criteria

In this stage of the SLR, we need to make an accurate selection of the studies extracted. To do this, we must define some rigorous inclusion/exclusion criteria, to decide which studies are going to be useful for our purpose. To achieve this, studies were excluded based on the following criteria:

1. Studies published before the 15-year time frame
2. Studies in languages other than English
3. Exclude non-academic sources, including blogs, news articles, marketing materials, and reports from non-academic organizations
4. Studies that are only marginally related to Big Data or the specific topics within our research questions.

In conclusion, all those studies that are not cut off by the exclusion criteria above are to be considered as included. They are called “Primary Studies” (PS).

### **Study quality assessment**

Kitchenham and Charters stresses the necessity of assessing the quality of primary studies to reduce bias and enhance the validity of the evaluation process. In our research, we employ a study quality assessment to make sure that we have only the most relevant results for our research.

To achieve this, we formulated a five question study quality questionnaire, which serves as the foundation for assessing the quality of the primary studies:

- QA1: has the primary study established a well-defined research objective?
- QA2: did the primary study comprehensively describe its research methods and data sources?
- QA3: has the technique or approach undergone a trustworthy validation?
- QA4: has the primary study effectively identified and discussed the significant challenges and limitations encountered in Big Data analysis?
- QA5: are the findings, research trends, and directions clearly presented and directly connected to the study’s objectives or goals?

Hence, we applied the formulated questionnaire to the included PSs to assess their quality. The output of this SLR stage will be discussed in Section 4.

### **Data extraction**

The data extraction process entails gathering relevant information from the chosen primary studies to address the research questions. To facilitate this process, we have created a dedicated data extraction form, as shown in Table 1. As suggested in Kitchenham and Charters, we used the test-retest process to check the consistency and accuracy of the extracted data with respect to the original sources. After finishing the data extraction for all the selected studies, we randomly selected 3 primary studies and performed a second extraction of the data. No inconsistencies were detected.

### **SLR: implementation**

In this section, we describe step-by-step the implementation and execution of the different stages of our SLR. Figure 2 depicts the search stages followed and the resulting number of primary studies for each stage.

In stage 1, an automated search was performed by applying the search string to the digital libraries. The software used for the management of the references is Zotero

**Table 1** Data extraction form

No.	Data extraction category	Description	Addresses
General description			
1	Identifier	A unique identifier for the primary study	
2	Date	Date when data was extracted from the study	
Study description			
3	Title	The title of the primary study	
4	Authors	Authors of the primary study	
5	Publication year	Year of publication	
6	Type	Conference paper, Journal article, book chapter, workshop paper	
7	Publication medium	Name of the publishing conference, journal, book, or workshop (including volume, number, page...)	
Study contents			
8	Objectives	What are the main objectives of the study?	RQ1, RQ2, RQ3
9	Domain analysis	How does the primary study approach the analysis of common application domains for Big Data analytics and their evolution over time?	RQ1
10	Identification of emerging trends	How does the primary study identify and discuss the emerging research trends and directions in Big Data that may shape the field in the future?	RQ3
11	Challenges	What are the challenges encountered in Big Data analysis, and how are they addressed or discussed in the study?	RQ2
12	Tool support/automation	Is the approach described in 9 or 10 supported by tools or fully/partially automated?	RQ1, RQ2, RQ3

([www.zotero.org](http://www.zotero.org)), a popular choice for SLRs. We began the research using the following research string:

“big data” OR “real-time data analytics” OR “large datasets”) AND (“methodologies” OR “techniques” OR “domains” OR “architectures” OR “solutions”) AND (“research trends” OR “future directions” OR “emerging technologies” OR “challenges” OR “SLR” OR “Systematic Literature Review”). As a result, we found a total of 4204 studies. The reason for this many results could be attributed mostly to the main topic of this SLR being “Big Data”, a hugely popular field, especially in the last few years.

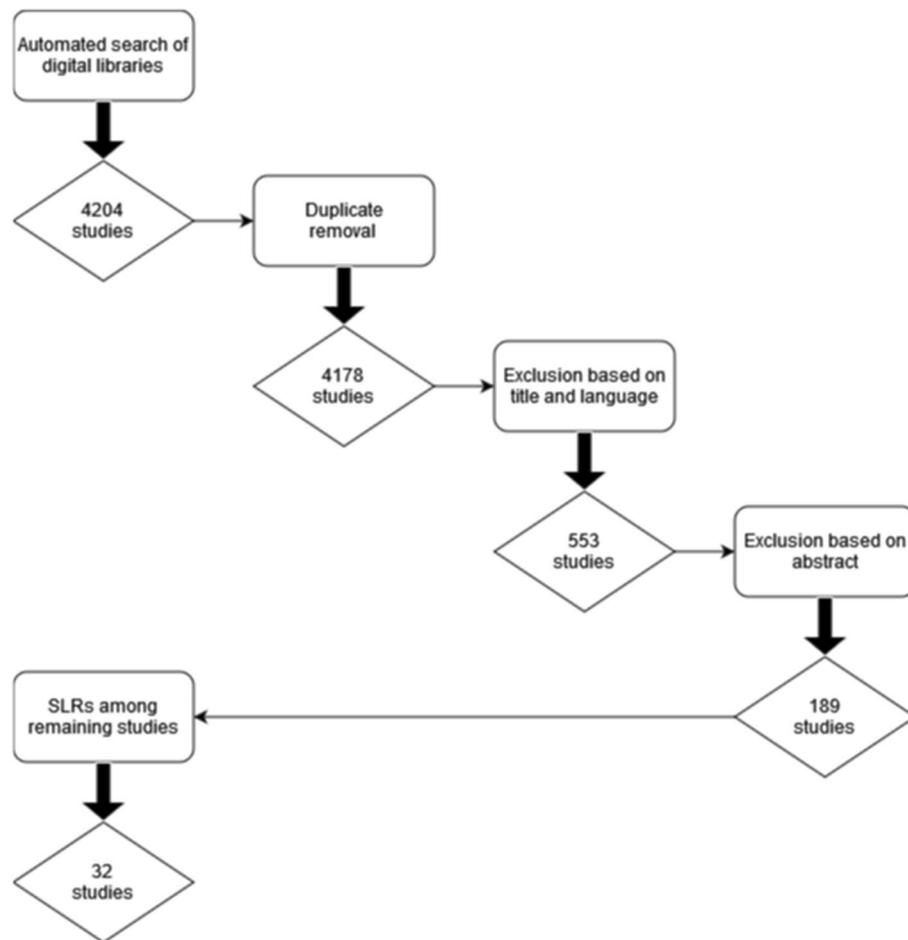
In stage 2, we used the Zotero’s duplicate identification tool, and we found a total of 25 duplicates. Additionally, 1 duplicate was found manually, bringing the total number of results to 4178 articles.

In stage 3, studies were excluded based on the title and the language. Fortunately, all the documents were in English, so we just needed to focus on the title, eliminating what had no use for our research. This cut down the total number to 553.

In stage 4, we eliminated the articles whose abstracts had marginal or no interest at all to us. At the end of the process, 189 Primary Studies were left, 32 of which were SLRs.

To ensure the best quality possible for our SLR, we have collected generic information on all the 189 studies that passed the Primary Study check. This information is depicted in Figs. 3 and 4. We then proceeded with an in-depth full-text review for the 32 PSSs, which are the main subject of our SLR.

Figure 3 depicts the distribution per year for all the 189 studies. Our SLR focuses on the evolution of Big Data in the last 15 years. In any case, no studies before 2012 were



**Fig. 2** Stages of the applied search strategy

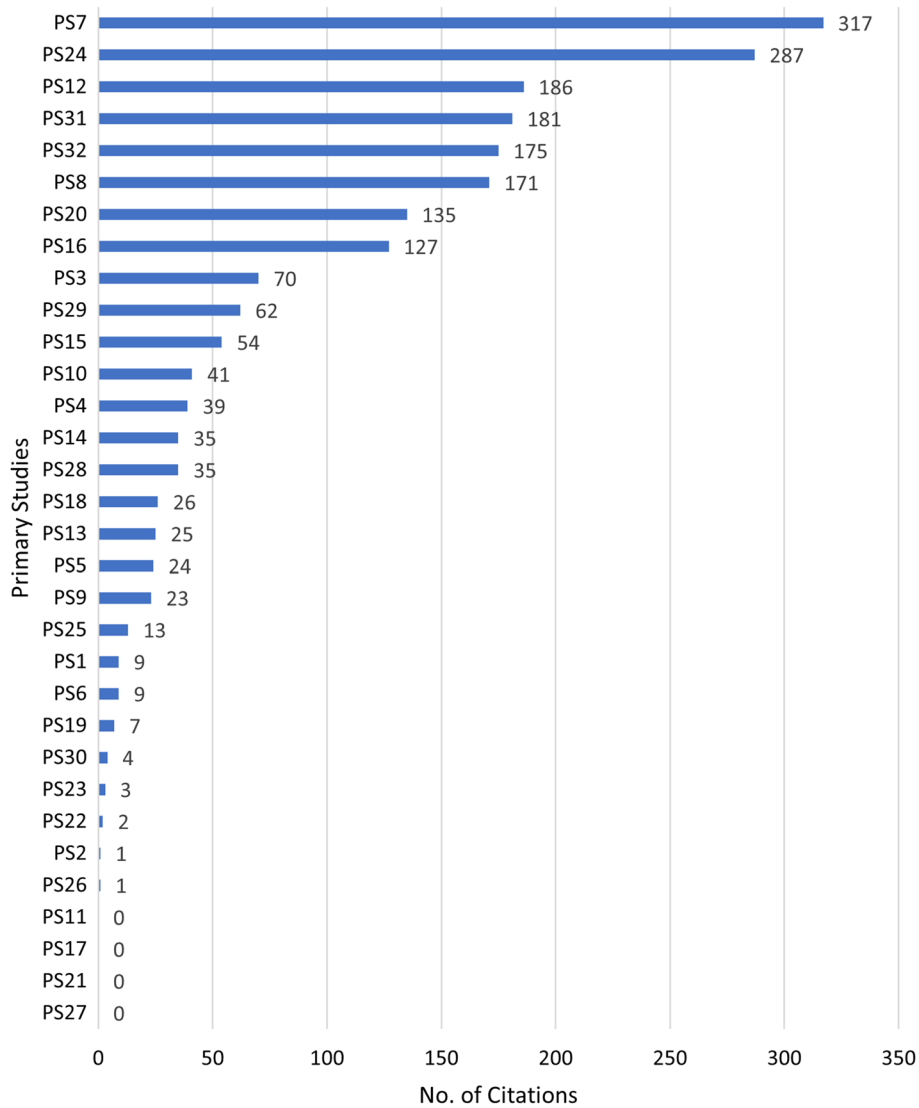
detected. The reason for this could be attributed to the fact that before then Big Data, as a research topic, was not as popular.

Figure 4 represents the total number of citations per year for our selected 32 Primary Studies. The graph clearly shows that the most recent studies have not been cited as much. Particularly, even though the studies released in the last two years compose about one third of our selected primary studies (11 out of 32), we can see that they have not been cited as much in comparison to the previous years. The lower citation rate may indicate that recently, researchers have focused more on understudied areas or more recent emerging trends, suggesting that the field of Big Data is currently undergoing an evolution. However, further analysis of the quality, methodology and context of these studies is necessary for more concrete conclusions.

For further clarity, we elaborated Table 2 to represent the chosen articles by highlighting the first author's family name, the venue, the title of each PS, and a short introduction that highlights the main findings of each PS. Note that the "J" indicates that the article has been published in a journal.

To better understand the influence of the selected Primary Studies over time, we created a bubble chart to show the most cited documents by aggregating the PSs with





**Fig. 3** Number of filtered primary studies and number of total citations

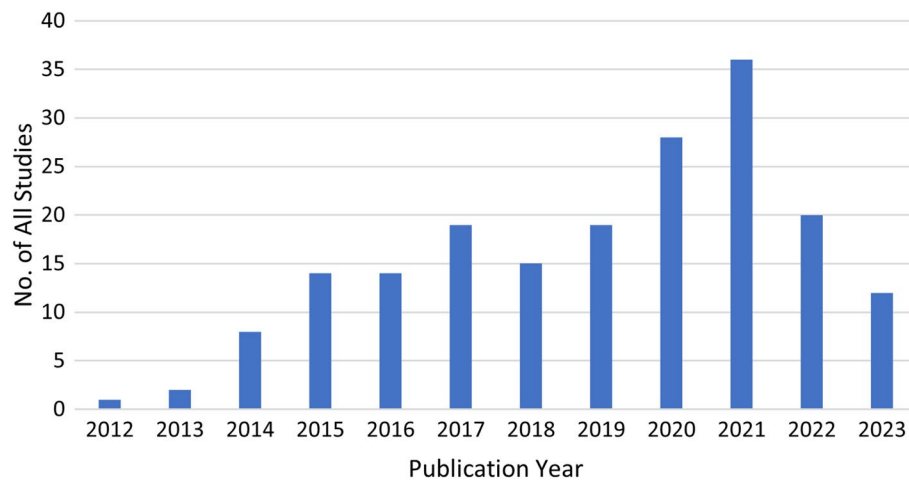
the same publication year (see Fig. 5). The size of each bubble is proportional to the number of citations.

**SLR: results**

The study of the PSs allowed us to pinpoint exactly which research question (RQ1-RQ3) is answered by each primary study. Table 3 summarizes our findings.

As previously stated, it is important to assess the quality of each study. In subsect. "Study quality assessment", we developed a brief questionnaire that would help us determine the quality of a primary study. Table 4 shows the results of this quality check. It uses a simple "Yes," "No," or "na" (used when we don't have enough information to answer) to fill out the Quality Assessment questionnaire.

From now on, we will briefly summarize each study and its findings.



**Fig. 4** Number of total PSs per year

#### **PS1—A comprehensive and systematic literature review on the Big Data management techniques in the internet of things [20]**

In this article, the authors explored the Big Data management techniques applied to the internet of things. Big Data was initially applied for healthcare monitoring, smart cities, and industrial systems. Over time, with the evolution of IoT, it expanded to include broader topics: healthcare applications involved health state monitoring and predictive modeling, smart cities encompassed traffic management, energy efficiency and security, while industrial systems employed Big Data to improve scalability and security. The application landscape broadened emphasizing the importance of quality attributes such as performance, efficiency, reliability, and scalability in ensuring the success of Big Data Analytics systems in IoT across ever-evolving domains.

The challenges and open issues in Big Data Analytics within IoT span various dimensions, including centralized architectures, energy consumption in data collection, blockchain limitations, communication challenges, and diverse data features.

For future research, the exploration of AI for intelligent mobile data collection will take on a more relevant role, combining compressive sensing with AI for communication challenges and utilizing new optimization algorithms for data processing. To ensure security and privacy in IoT, Big Data Analytics could involve cryptography mechanisms, a data perception layer and a lightweight framework with AI. Addressing these challenges is essential for advancing Big Data Analytics in the evolving landscape of IoT applications.

#### **PS2—A comprehensive review on Big Data for industries challenges and opportunities [21]**

The article explores the transformative impact of Big Data Analytics in power systems, mineral industries, and manufacturing. In power systems, it revolutionizes fault detection, enables early warning systems and predicts future electricity demand, enhancing reliability and decision-making. For mineral industries, Big Data improves data storage, processing and analytics, optimizing exploration, extraction, and

**Table 2** Primary studies selected

Primary study		Title	Comments
Id	References		
PS1	Naghib et al. [20] (J) Wireless Networks	A comprehensive and systematic literature review on the Big Data management techniques in the internet of things	The study explores Big Data management in IoT. Challenges include centralized architectures, energy consumption, and communication issues. Future research should focus on AI for mobile data collection and security mechanisms.
PS2	Sarker et al. [21] (J) IEEE Access	A Comprehensive Review on Big Data for Industries: Challenges and Opportunities	The article explores Big Data Analytics in power systems, mineral industries, and manufacturing, improving fault detection, resource management, and decision-making. Challenges include data quality assessment and security. Future research should focus on data quality assessment and secure communication solutions.
PS3	Bansal et al. [22] (J) ACM Computer Survey	A Survey on IoT Big Data: Current Status, 13 V's Challenges, and Future Directions	The document explores Big Data Analytics integration with the IoT, across domains like energy, healthcare, transportation, and smart cities. It identifies 13 major challenges ("13 V's"). Future trends include energy-efficient data acquisition, machine learning integration, and blockchain adoption.
PS4	Hordri et al. [23] (J) International Journal of Advances in Soft Computing and its Applications	A systematic literature review on features of deep learning in Big Data analytics	This study discusses the evolution of Big Data and the increasing importance of machine learning, especially Deep Learning, in sectors like healthcare and finance. It proposes a review focused on Deep Learning in Big Data Analytics to address research gaps and highlight ongoing advancements.
PS5	Zhong et al. [24] (J) Journal of Supercomputing	A systematic survey of data mining and Big Data analysis in internet of things	The document discusses the evolution of Big Data Analytics in sectors like healthcare and finance. It identifies three main challenges: societal changes, IoT usage in businesses, and technical issues like security. It anticipates the growing importance of Big Data frameworks in handling IoT-generated data and highlights potential paths for future research in this area.
PS6	Rashid [25] (J) EAI Endorsed Transactions on Scalable Information Systems	Access methods for Big Data: current status and future directions	The study explores diverse applications of Big Data Analytics across various sectors. It addresses challenges such as data heterogeneity, scale, timeliness, and privacy. Future directions include developing systems for real-time processing and advanced databases.

**Table 2** (continued)

Primary study		Title	Comments
Id	References		
PS7 (J)	O'Donovan et al. [26] J. Big Data	An industrial Big Data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities	The research introduces a novel Big Data pipeline for industrial analytics in manufacturing. It excels in integrating legacy and smart devices, ensuring cross-network communication, and adhering to open standards.
PS8 (J)	Kushwaha et al. [27] International Journal of Information Management Data Insights	Applications of Big Data in emerging management disciplines: A literature review using text mining	The study explores various applications of Big Data Analytics across emerging management domains, addressing adoption challenges and proposing a framework for future research.
PS9 (J)	Alkhalil et al. [28] International Journal of Information and Communication Technology Education	Applying Big Data analytics in higher education: A systematic mapping study	The document examines Big Data Analytics in Higher Education, highlighting applications and challenges. It emphasizes the need for practical outcomes, automated tools, and validated frameworks to mature BDA integration in the sector.
PS10 (J)	Rahmani et al. [29] PeerJ Computer Science	Artificial intelligence approaches and mechanisms for Big Data analytics: a systematic study	The document explores AI-driven Big Data Analytics, emphasizing machine learning and decision-making algorithms. Challenges include processing vast, heterogeneous data and ensuring system security. Emerging trends highlight Big Data Analytics for IoT, via fog computing and the need for enhanced algorithms to handle extensive data.
PS11 (J)	Lundberg [30] Journal of Big Data	Bibliometric mining of research directions and trends for Big Data	The research identifies key application domains, and emphasizes emerging directions like Machine Learning and Healthcare. It introduces a semi-automatic method to address challenges in research direction identification.
PS12 (J)	Baig et al. [31] Information Processing and Management	Big Data adoption: State of the art and research challenges	The study explores Big Data Analytics adoption across various sectors. It emphasizes the need for increased research in untapped areas and suggests addressing challenges through a mixed-method approach. It hints at a changing research focus globally, with growing awareness of untapped opportunities in Big Data adoption.
PS13 (J)	Ikegwu et al. [32] Cluster Computing	Big Data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions	The research outlines Big Data Analytics' integration across sectors and addresses challenges like skill shortages and privacy concerns. Emerging trends involve sourcing data from education and IoT devices, refining data processing, and exploring deep learning methods.

**Table 2** (continued)

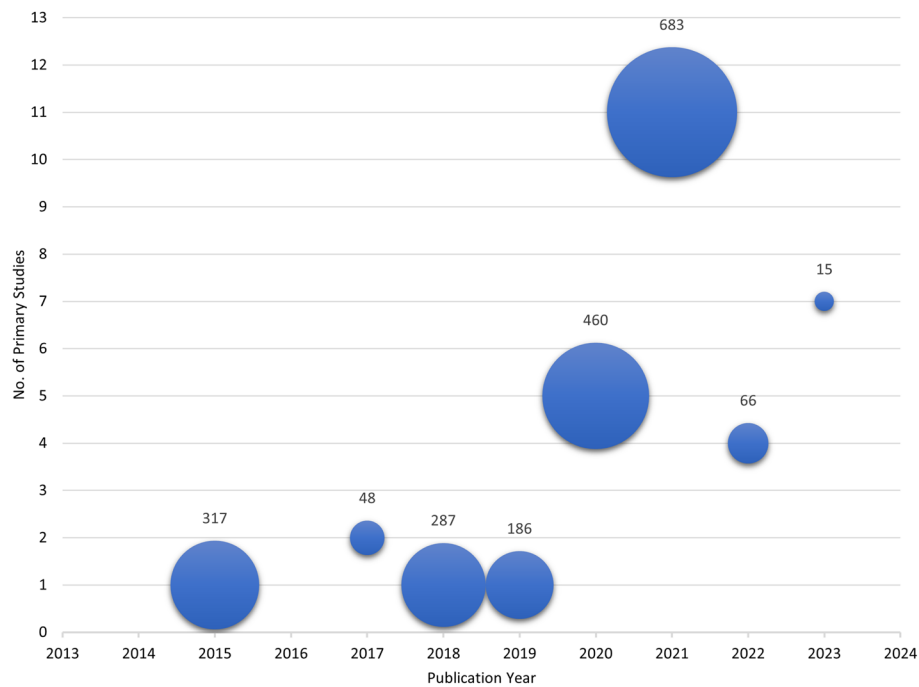
Primary study		Title	Comments
Id	References		
PS14	Imran et al. [33]	Big Data analytics in healthcare. A systematic literature review and roadmap for practical implementation	The paper explores Big Data Analytics applications in healthcare, introducing the Med-BDA architecture. It addresses challenges like increased costs and skill acquisition, presenting solutions. The document concludes with potential extensions to Med-BDA and its applicability to diverse Big Data domains.
(J)	IEEE/CAA Journal of Automatica Sinica		
PS15	Zahid et al. [34]	Big Data analytics in telecommunications: Literature review and architecture recommendations	This paper introduces LambdaTel for Big Data Analytics in TELCO, addressing challenges like lack of standardized architecture. It suggests ongoing adaptation through Dockerized implementation and LambdaTel application for cross-selling/up-selling.
(J)	IEEE/CAA Journal of Automatica Sinica		
PS16	Bazzaz Abkenar et al. [35]	Big Data analytics meets social media: A systematic review of techniques, open issues, and future directions	The document discusses social media's impact on healthcare. Challenges include privacy concerns and accuracy enhancement with incomplete data. Emerging trends in Big Data Analytics include real-time and predictive analysis, with under-explored areas.
(J)	Telematics and Informatics		
PS17	ElSayed et al. [36]	Big Data and its Future in Computational Biology: A Literature Review	The study highlights Big Data's significance in computational biology and healthcare. Challenges include under-utilization of electronic health data and converting it into actionable insights. It outlines emerging trends, focusing on the role of volume, variety, and velocity in bioinformatics. Key technologies like Hadoop and MapReduce are discussed.
(J)	Journal of Computer Science		
PS18	Hajjali [37]	Big Data and sentiment analysis: A comprehensive and systematic literature review	The study explores various applications of Big Data Analytics, focusing on sentiment analysis for marketing and disaster response. Challenges include data quality issues and limitations of centralized data mining algorithms for distributed systems. Future research should examine specific methods and deep learning to understand their strengths and limitations better.
(J)	Concurrency and Computation: Practice and Experience		
PS19	Ahmadova et al. [38]	Big Data applications on the Internet of Things: A systematic literature review	The document explores various applications of Big Data, including customer sentiment analysis and disaster response. Challenges involve data acquisition from IoT devices, security concerns, and system scalability. Future directions are discussed, including enhancing algorithm efficiency, addressing energy consumption, and exploring the synergy of Big Data and machine learning for emergency systems.
(J)	International Journal of Communication Systems		

**Table 2** (continued)

Primary study		Title	Comments
Id	References		
PS20	Baig et al. [39] (J) Int. J. Educ. Technol. High. Educ.	Big Data in education: a state of the art, limitations, and future research directions	The paper discusses the widespread use of Big Data Analytics, particularly in education. Challenges include qualitative methods and data collection techniques, while future research should focus on quantifying Big Data's impact and integrating it into study plans.
PS21	Reshi et al. [40] (J) Scalable Computing	Big Data in Healthcare: a Comprehensive Bibliometric Analysis of Current Research Trends	The document outlines the evolution of Big Data Analytics across various domains, focusing particularly on healthcare. It discusses challenges but lacks specific solutions. Future trends include data analytics and collaborative networks.
PS22	Pulikkottil et al. [41] (J) IEEE Access	Big Data Life Cycle in Shop-Floor-Trends and Challenges	The document explores Big Data Analytics in manufacturing, highlighting applications and challenges. Emerging trends include AI and statistical approaches, with a focus on data privacy.
PS23	Arshad et al. [42] (J) Computers, Materials and Continua	Big Data Testing Techniques: Taxonomy, Challenges and Future Trends	The paper discusses the transition to advanced testing methods in Big Data Analytics to address challenges like ETL processes and node failures. It outlines emerging research trends for the future of Big Data Analytics.
PS24	Gupta et al. [43] (J) International Journal of Information Management	Big Data with cognitive computing: A review for the future	The paper explores the application domains of Big Data Analytics. Challenges include perceived lack of strategic value and data management issues. Emerging trends include rising interest in cognitive computing and a call for more global collaboration.
PS25	Saltz and Krasteva [44] (J) PeerJ Computer Science	Current approaches for executing Big Data science projects-a systematic literature review	This paper explores the landscape of Big Data Analytics, noting increased interest. It addresses challenges in analysis. Emerging trends include the need for integrated frameworks and standardized approaches in data science.
PS26	Liu et al. [45] (J) Symmetry	Data quality affecting Big Data analytics in smart factories: Research themes, issues and methods	This review examines the use of Big Data Analytics in Smart Factories, highlighting empirical case studies and challenges. It emphasizes the need for integrated approaches and calls for future research in frameworks and emerging scenarios.
PS27	Ahmed et al. [46] (J) IEEE Access	Harnessing Big Data Analytics for Healthcare: A Comprehensive Review of Frameworks, Implications, Applications, and Impacts	The study delves into Big Data Analytics in healthcare, covering application domains, challenges, and emerging trends. It emphasizes issues like data quality and privacy concerns.

**Table 2** (continued)

Primary study		Title	Comments
Id	References		
PS28	Karimi et al. [47]	Leveraging Big Data in smart cities: A systematic review	This study investigates Big Data Analytics over diverse domains. Challenges like security and scalability drive innovative solutions. Emerging trends include integrating Big Data with AI, machine learning, and exploring quantum computing.
(J)	Concurrency and Computation: Practice and Experience		
PS29	Gong and Janssen [48]	Roles and capabilities of enterprise architecture in Big Data analytics technology adoption and implementation	The document discusses the evolution and applications of Big Data Analytics, focusing on healthcare and finance. It addresses challenges like data privacy and scalability, proposing innovations. Emerging trends include integrating AI and Machine Learning for enhanced analytics.
(J)	Journal of Theoretical and Applied Electronic Commerce Research		
PS30	Anawar et al. [49]	Security and Privacy Challenges of Big Data Adoption: A Qualitative Study in Telecommunication Industry	The research explores Big Data Analytics applications across various domains. It addresses challenges like data security and privacy, proposing solutions. Emerging trends include explainable AI, ethical data practices, and innovations in handling streaming data and blockchain integration.
(J)	International Journal of Interactive Mobile Technologies		
PS31	Rathore et al. [50]	The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities	The document explores Big Data Analytics applications across industries like healthcare and manufacturing. Challenges include data collection difficulties and the need for accurate AI models. Future trends anticipate integration of AI and Big Data in digital twinning.
(J)	IEEE Access		
PS32	Mohamed et al. [51]	The state of the art and taxonomy of Big Data analytics: view from new Big Data framework	The paper explores Big Data Analytics, highlighting Hadoop's dominance. Major challenges include handling diverse data formats and optimizing algorithms. Future advancements are anticipated in e-commerce and IoT applications.
(J)	Artificial Intelligence Review		



**Fig. 5** Bubble chart showing the number of primary studies and total citations per year of publication

resource management. In manufacturing, it facilitates data-driven decision-making, comprehensive product quality assessment, and streamlined supply chain management for increased operational efficiency.

The study also highlights challenges in implementing Big Data Analytics, emphasizing the crucial need for precise data quality assessment models and secure frameworks. Machine learning and data analytics play a pivotal role in overcoming challenges, particularly in fault detection, load forecasting, and reservoir management. The call for open-source databases and integration with machine learning addresses the scarcity of datasets, reflecting challenges in maximizing Big Data's potential.

Furthermore, the paper recommends future research trends, including advanced data quality assessment models, frameworks for high-dimensional data and solutions for secure communication. Emphasizing open-source databases and integrating machine learning promotes a collaborative and transparent approach. The call for interpretable models reflects a trend toward understanding and optimizing Big Data Analytics. Overall, these recommendations shape the future direction of Big Data applications in diverse industries.

### **PS3—A survey on IoT Big Data current status, 13 V's challenges, and future directions [22]**

The document delves into the landscape of Big Data Analytics, particularly exploring its integration with the Internet of Things. Application domains such as energy, healthcare, transportation, and smart cities emerge prominently. The discussion unfolds how these domains have evolved, signalling a shift towards IoT-driven intelligent applications.

Within this expansive terrain, the study identifies and elucidates 13 major challenges encapsulated by the "13 V's". These challenges span traditional aspects like volume,



**Table 3** Research questions addressed by the primary studies

Primary study ID	Research questions		
	RQ1	RQ2	RQ3
PS1	Yes	Yes	Yes
PS2	Yes	Yes	Yes
PS3	Yes	Yes	Yes
PS4	Yes	No	No
PS5	Yes	Yes	Yes
PS6	Yes	Yes	Yes
PS7	Yes	Yes	Yes
PS8	Yes	Yes	Yes
PS9	Yes	Yes	No
PS10	Yes	Yes	Yes
PS11	Yes	Yes	Yes
PS12	Yes	Yes	No
PS13	Yes	Yes	Yes
PS14	Yes	Yes	Yes
PS15	Yes	Yes	No
PS16	Yes	Yes	Yes
PS17	Yes	Yes	Yes
PS18	Yes	Yes	Yes
PS19	Yes	Yes	Yes
PS20	Yes	Yes	Yes
PS21	Yes	No	Yes
PS22	Yes	Yes	Yes
PS23	No	Yes	Yes
PS24	Yes	Yes	Yes
PS25	Yes	Yes	Yes
PS26	Yes	Yes	Yes
PS27	Yes	Yes	Yes
PS28	Yes	Yes	Yes
PS29	Yes	Yes	Yes
PS30	Yes	Yes	Yes
PS31	Yes	Yes	Yes
PS32	Yes	Yes	No

velocity, and variety, extending to less common concerns like vagueness and location-aware data processing. The document also offers innovative solutions, like edge-based processing and semantic representation, as strategies to manage these complex challenges.

In regards to the future, the document outlines emerging trends anticipated to define the Big Data landscape in the coming 5 to 10 years. These include a focus on energy-efficient data acquisition, the integration of machine learning and deep learning for advanced analytics, a strategic emphasis on edge and fog infrastructures, the evolving paradigm of multi-cloud data management, a shift towards data-oriented network addressing, and the increasing adoption of blockchain technology. These trends collectively indicate a trajectory towards more efficient, scalable, and secure practices in Big Data Analytics, particularly within the realm of IoT applications.

**Table 4** Study quality assessment

Primary study ID	Quality assessment questions				
	QA1	QA2	QA3	QA4	QA5
PS1	Yes	Yes	Na	Yes	Yes
PS2	Yes	Yes	Yes	Yes	Yes
PS3	Yes	Yes	Na	Yes	Yes
PS4	Yes	Yes	No	No	Yes
PS5	Yes	Yes	Na	Yes	Yes
PS6	Yes	Yes	Na	Yes	Yes
PS7	Yes	Yes	Na	Yes	Yes
PS8	Yes	Yes	No	Yes	Yes
PS9	Yes	Yes	Yes	Yes	Yes
PS10	Yes	Yes	Yes	Yes	Yes
PS11	Yes	Yes	Na	Yes	Yes
PS12	Yes	Yes	Na	Yes	Yes
PS13	Yes	Yes	Yes	Yes	Yes
PS14	Yes	Yes	Na	Yes	Yes
PS15	Yes	No	Yes	Yes	Yes
PS16	Yes	Yes	Na	Yes	Yes
PS17	Yes	Yes	Na	Yes	Yes
PS18	Yes	No	Na	Yes	Yes
PS19	Yes	Yes	Na	Yes	Yes
PS20	Yes	Yes	Yes	Yes	Yes
PS21	Yes	Yes	Na	No	Yes
PS22	Yes	Yes	Na	Yes	Yes
PS23	Yes	Yes	Na	Yes	Yes
PS24	Yes	Yes	Na	Yes	Yes
PS25	Yes	Yes	Na	Yes	Yes
PS26	Yes	Yes	Na	Yes	Yes
PS27	Yes	Yes	Yes	Yes	Yes
PS28	Yes	Yes	Na	Yes	Yes
PS29	Yes	Yes	Na	Yes	Yes
PS30	Yes	Yes	Na	Yes	Yes
PS31	Yes	Yes	Na	Yes	Yes
PS32	Yes	Yes	Na	Yes	Yes

**PS4—A systematic literature review on features of deep learning in Big Data analytics [23]**

The document navigates the evolution of Big Data, emphasizing challenges and the rise of machine learning, particularly Deep Learning. Machine learning's widespread use, observed in areas like healthcare and finance, underscores its crucial role. Even in complex data scenarios, its effectiveness is evident, as demonstrated by the U.S. Department of Homeland Security's success in identifying threats.

Recognizing a gap in existing research, the document proposes a review focusing on Deep Learning in Big Data Analytics. The goal is to explore features like hierarchical layers and high-level abstraction. The study emphasizes Deep Learning's strength in handling extensive datasets, its versatility, and its ability to prevent over fitting.

This exploration into Big Data's journey underscores the central role of machine learning. The proposed review, specifically focusing on Deep Learning in Big Data

Analytics, not only captures current advancements but also suggests there's more to discover in the future where Big Data and machine learning intersect.

**PS5—A systematic survey of data mining and Big Data analysis in internet of things [24]**

The document navigates through diverse applications of Big Data Analytics, illustrating its transformative journey across sectors. Notably, it tracks the evolution within healthcare and finance, showcasing how Big Data has become integral to these domains over time.

Going further, the research dives into the various challenges of Big Data analysis. It identifies three main challenges: dealing with societal changes, understanding how businesses use IoT, and solving technical issues like security and connectivity. The study emphasizes the need to adapt to society's changing needs, categorize IoT uses in business and front technical problems for effective Big Data analysis.

Moreover, the research anticipates future trends, in particular the rising importance of Big Data frameworks in handling expansive IoT-generated data. The intersection of these frameworks with data mining in the IoT domain emerges as a pivotal focus, pointing toward exciting possibilities and potential paths for future research in the realm of Big Data.

**PS6—Access methods for Big Data: current status and future directions [25]**

The document explores diverse applications of Big Data Analytics in research, education, urban planning, transportation, environmental modeling, energy conservation, and homeland security, emphasizing its transformative potential.

It addresses challenges like heterogeneity, scale, timeliness, privacy, and the evolving processing paradigms due to data volume surpassing computational resources.

Future directions include the need for systems handling structured and unstructured data, embedded analytics for real-time processing, innovative paradigms, application frameworks, and advanced databases ensuring transactional semantics. The research underscores the importance of tools addressing ethical, security, and privacy concerns.

**PS7—An industrial Big Data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities [26]**

This research introduces an innovative Big Data pipeline designed for industrial analytics in manufacturing.

The pipeline excels in integrating legacy and smart devices, ensuring cross-network communication, and adhering to open standards, marking a significant evolution in the field. The document showcases the pipeline's ability to handle complexities, integrate older systems, ensure reliability, and scale efficiently in industrial data analytics.

The future plan involves implementing the pipeline to validate its architecture, particularly in predictive maintenance for Wind Turbines and Air Handling Units, contributing to the evolving landscape of Big Data Analytics.

**PS8—Applications of Big Data in emerging management disciplines: a literature review using text mining [27]**

This study explores diverse applications of Big Data Analytics across twelve emerging management domains, emphasizing their dynamic nature over time.

It addresses adoption challenges, focusing on data quality, resource management, and distinguishing between the ability and capability of organizations in using Big Data Analytics. The research underscores the thoughtful adoption of Big Data Analytics and the importance of measuring its business value comprehensively. It acknowledges the difficulty of translating insights into real-time actionable items.

Looking forward, the study proposes a framework connecting emerging management domains with conventional practices, suggesting future research areas in human resources, marketing, sales, strategy, and services. The research emphasizes the need for in-depth exploration to integrate emerging domains into established management practices, providing valuable insights for research and practical application.

#### **PS9—Applying Big Data analytics in higher education: a systematic mapping study [28]**

The document conducts a thorough exploration of Big Data Analytics (BDA) in Higher Education Institutions from 2010 to 2020. It uncovers diverse BDA applications in three domains: Educational Quality, Decision-Making Process, and Information Management.

Challenges in BDA adoption include handling large data volumes, addressing privacy concerns, and dealing with resource constraints. The study emphasizes the need for practical outcomes, automated tools, and validated frameworks.

Despite robust research interest, the field exhibits immaturity, with a prevalence of conference papers indicating an early development stage. The study calls for increased empirical research to fortify the evidence base and foster a more mature BDA integration in higher education.

#### **PS10—Artificial intelligence approaches and mechanisms for Big Data analytics: a systematic study [29]**

The SLR explores AI-driven Big Data Analytics, emphasizing machine learning, knowledge-based reasoning, decision-making algorithms, and search methods. Applications, notably in supervised learning, aim to enhance precision and efficiency but grapple with complexity and scalability issues.

Challenges encompass processing vast, heterogeneous data, ensuring system security, and addressing qualitative parameters. Fog computing emerges as a potential solution, yet security concerns remain under-explored.

Emerging trends spotlight Big Data Analytics for IoT through fog computing, the need for enhanced algorithms handling extensive data, and the necessity to address data quality issues in unstructured formats.

#### **PS11—Bibliometric mining of research directions and trends for Big Data [30]**

The research identifies key application domains, with particular focus on China, and emerging directions such as Machine Learning and Healthcare.

Navigating challenges, the study introduces a semi-automatic method, utilizing blacklists and thesauri to enhance precision in identifying research directions. This favors a balance between automation and expert input.

The study forecasts Big Data's future using a growth rate criterion, emphasizing Machine Learning and Deep Learning. Moreover, the study suggests applying its methodology not

only to Big Data but also to various research areas, such as Machine Learning, showcasing its potential applicability in diverse research areas.

**PS12—Big Data adoption: state of the art and research challenges [31]**

The study explores the widespread adoption of Big Data Analytics across diverse sectors such as finance, education, healthcare, and more. It identifies a need for increased research in untapped areas like education and healthcare, suggesting potential transformative effects.

Challenges in current Big Data research include the need for refined theoretical models, adaptable data collection methods, and larger sample sizes to ensure accuracy. The study recommends a mixed-method approach to address these challenges effectively.

The study, although not explicitly stating upcoming trends, suggests a changing research focus in both developing and developed countries. It indicates a growing awareness of untapped opportunities, hinting at a future emphasis on specific situations and new factors in Big Data adoption.

**PS13—Big Data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions [32]**

This research provides a comprehensive overview of Big Data Analytics. Exploring application domains, it traces Big Data's historical integration across education, healthcare, finance, national security, and Industry 4.0 components like IoT and smart cities.

Delving into challenges, the research highlights skill shortages, dataset management, privacy, scalability, and intellectual property issues. Solutions range from software-defined data management to innovative truthfulness and privacy preservation methods.

Looking ahead, the study identifies some emerging trends: sourcing data from education and diverse IoT devices, refining pre-processing, advancing data management, enhancing privacy, and exploring deep learning methods. These trends forecast a dynamic future for Big Data Analytics, shaping the field in the next years.

**PS14—Big Data analytics in healthcare: a systematic literature review and road map for practical implementation [33]**

The paper conducts a thorough examination of Big Data Analytics (BDA) applications in healthcare, introducing the novel Med-BDA architecture.

Notably, the work addresses challenges inherent in BDA (such as increased costs, difficulty in acquiring a relevant skill set, rapidly expanding technology stack, and heightened management overhead), presenting a comprehensive road map to alleviate issues such as cost escalation and skill acquisition hurdles.

The document concludes by outlining the potential for extensions to Med-BDA and its applicability to diverse Big Data domains, showcasing a forward-looking perspective in BDA research and application.

**PS15—Big Data analytics in telecommunications: literature review and architecture recommendations [34]**

The document explores Big Data Analytics in TELCO, introducing LambdaTel as a proposed solution for batch and streaming data processing. It discusses Big Data Analytics applications like CRM and Customer Attrition.

Challenges, such as the lack of standardized architecture, are acknowledged. LambdaTel addresses these challenges through a structured approach, emphasizing security and recommending the usage Python.

While not explicitly talking about future trends, the document suggests a commitment to ongoing adaptation, seen in recommendations like Python usage, Dockerized implementation and the application of LambdaTel in a local Telco company for cross-selling/up-selling.

**PS16—Big Data analytics meets social media: a systematic review of techniques, open issues, and future directions [35]**

The document highlights social media's transformative impact in healthcare, emphasizing its role in patient support and disease tracking. It emphasizes leveraging social platforms for patient support, disease prevention, and real-time tracking of contagious diseases.

The review highlights challenges in both content and network-oriented approaches, such as privacy concerns, scalability limitations, and accuracy enhancement with incomplete data. Comprehensive resolution remains an open frontier, requiring innovative solutions for privacy preservation and accurate predictions.

The paper also highlights emerging trends in Big Data Analytics, emphasizing real-time and predictive analysis, and addressing challenges in sentiment analysis. It identifies under explored areas like political and e-commerce applications, underscoring the expanding trajectory of Big Data Analytics. Furthermore, it emphasizes the evolving complexities of linguistic analysis, underlining the need for domain-dependent sentiment analysis, and addressing challenges like sarcasm detection.

**PS17—Big Data and its future in computational biology: a literature review [36]**

The document underscores the growing significance of Big Data in computational biology and healthcare, particularly in the conversion of healthcare records into digital formats. It highlights the major application domains, focusing on optimizing health and medical care through electronic health data.

Challenges include the under-utilization of electronic health data and the need to convert raw data into actionable information. Despite increasing interest, the field lacks comprehensive literature reviews.

The document outlines emerging trends in Big Data for computational biology and bio informatics. It emphasizes the pivotal role of volume, variety, and velocity in defining Big Data's impact on bio informatics. Key technologies, including Hadoop and MapReduce, are discussed, illustrating their significance in the field. The integration of Big Data technology is shown to enhance biological findings and facilitate real-time identification of high-risk patients. However, limitations, such as narrow study focuses, are noted.

**PS18—Big Data and sentiment analysis: a comprehensive and systematic literature review [37]**

The document delves into the diverse applications of Big Data Analytics, spotlighting its evolution, notably in sentiment analysis for marketing and disaster response.

Challenges identified include data quality issues and the absence of standardized disaster-related datasets. The limitations of centralized data mining algorithms for distributed systems are acknowledged, urging exploration into other platforms (YARN is directly cited as an example). The analysis underscores the need for immediate and improved performance, emphasizing real-time analysis.

In the future, it is important for researchers to carefully look into specific methods like Hadoop, MapReduce, and deep learning. This will help us better understand what these methods are good at and where they might struggle.

**PS19—Big Data applications on the internet of things: a systematic literature review [38]**

This document explores the evolving applications of Big Data, from understanding customer sentiments to enhancing disaster response. Hadoop emerges as a popular framework.

Challenges include robust data acquisition from IoT devices, addressing security concerns and optimizing system scalability.

Future directions involve improving algorithms for efficiency, addressing energy consumption, and exploring the synergy of Big Data and machine learning for emergency systems.

**PS20—Big Data in education: a state of the art, limitations, and future research directions [39]**

The paper talks about how Big Data Analytics is used in various areas, especially in education, with a noticeable increase in publications from 2014 to 2019. It highlights important topics like how students behave, creating models, using data for education, improving systems, and adding Big Data (as a topic) to study plans.

Researchers face challenges in employing qualitative methods and data collection techniques, highlighting the need for quantitative approaches and more robust methodologies.

Future research should emphasize quantifying Big Data's impact, adopting efficient solutions, exploring new tools and developing frameworks for educational applications. Integrating the concept of Big Data into study plans requires significant restructuring and well-designed learning activities.

**PS21—Big Data in healthcare—a comprehensive bibliometric analysis of current research trends [40]**

This document unveils the dynamic evolution of Big Data Analytics across diverse application domains, with a notable surge in research activities within the healthcare sector since 2012.

While the study discusses various related studies and challenges in Big Data analysis, it does not directly address or provide specific solutions to those challenges.

Looking ahead, the document reveals emerging trends and directions shaping the future of Big Data Analytics over the next 5 to 10 years. Key themes include data analytics, predictive analytics, and collaborative networks, providing a glimpse into the evolving landscape of research endeavors.

#### **PS22—Big Data life cycle in shop-floor-trends and challenges [41]**

The document explores Big Data Analytics in manufacturing, emphasizing its application domains like maintenance, automation, and decision-making.

Challenges include data measurement errors, high-frequency sampling issues, and the need for real-time processing. The study notes a shift to scalable storage options and highlights the importance of efficient data management.

Emerging trends involve the prominent role of AI and statistical approaches in data processing, coupled with a growing emphasis on data privacy. The study concludes with a call for future work focused on developing a consolidated framework for the Big Data life cycle in manufacturing.

#### **PS23—Big Data testing techniques: taxonomy, challenges and future trends [42]**

The paper explores the shift from traditional to advanced testing methods to address challenges in ETL processes, data quality, and node failures.

Addressing major challenges in Big Data analysis, the paper emphasizes the inadequacy of traditional testing, highlighting specific difficulties like ETL testing, node failure prevention, and unit-level debugging. It showcases evolving strategies employed by researchers to ensure the quality of Big Data systems.

Looking ahead, the document outlines emerging research trends shaping the future of Big Data Analytics. It identifies trends such as combinatorial testing techniques, fault tolerance testing, and model-driven entity reconciliation testing as key areas for future exploration.

#### **PS24—Big Data with cognitive computing: a review for the future [43]**

The paper explores the application domains of Big Data Analytics, highlighting its early stage in conjunction with cognitive computing, particularly in healthcare.

Challenges in adoption are attributed to a perceived lack of strategic value. The study categorizes issues into data, process, and management challenges, emphasizing the potential of integrating cognitive computing to overcome barriers.

Regarding emerging trends, there's a rising interest in cognitive computing. The research encourages more global collaboration and highlights a gap in understanding how Big Data studies impact decision-making processes.

#### **PS25—Current approaches for executing Big Data science projects—a systematic literature review [44]**

The paper explores the landscape of Big Data Analytics. Regarding the common application domains and their evolution, the study notes a significant increase in articles. Workshops play a crucial role in shaping the trajectory, reflecting a robust and expanding interest in Big Data Analytics, influenced by technological advancements.



It also addresses challenges in Big Data analysis, with a focus on workflows and agility. While acknowledging the conceptual nature of agility papers, a gap between theoretical benefits and practical implementation is underscored, necessitating further exploration to optimize agile frameworks for data science projects.

The study highlights emerging trends in Big Data, emphasizing the need for integrated frameworks in data science. It points out a research gap in standardized approaches, urging further exploration for innovative methodologies.

**PS26—Data quality affecting Big Data analytics in smart factories: research themes, issues and methods [45]**

This review explores the growing applications of Big Data Analytics in Smart Factories, emphasizing an upsurge in empirical case studies on production, process monitoring, and quality tracing.

Challenges involve key data quality issues (missing, anomalous, noisy, and old data), as well as ISO-defined data quality dimensions. While technical methods prevail, an integrated approach combining technical and non-technical methods for comprehensive data quality management is highlighted. Theoretical insights focus on data quality dimensions, issues, and resolutions, while practical implications underscore the need for collaboration and integrated methods.

The study calls for future research in frameworks, data quality requirements, and emerging scenarios, contributing to Big Data Analytics evolution in Smart Factories.

**PS27—Harnessing Big Data analytics for healthcare: a comprehensive review of frameworks, implications, applications, and impacts [46]**

The study meticulously explores the landscape of Big Data Analytics in healthcare. Noteworthy application domains, such as multi modal data analysis and fusion, natural language processing, and electronic health records, emerge from this exploration.

Some challenges faced in Big Data analysis are presented in the document, highlighting issues like data quality, privacy concerns, and a shortage of skilled professionals. It emphasizes the necessity for interoperability and standardization while identifying ongoing challenges in multi modality, ethical considerations, and bias mitigation.

The research outlines emerging trends and directions in Big Data, emphasizing the importance of ongoing exploration in areas like multi modality, data mining, precision medicine, ethical considerations, and the broader understanding of the Big Data Ecosystem.

**PS28—Leveraging Big Data in smart cities: a systematic review [47]**

Big Data Analytics has evolved across diverse domains, expanding from finance and healthcare to smart cities and e-commerce. This evolution has been marked by a transformative impact on industries.

Challenges in Big Data, including security, privacy, and scalability issues, have prompted innovative solutions. Advanced encryption, anonymization techniques, and scalable computing frameworks address these concerns.

Looking ahead, emerging trends highlight the fusion of Big Data with AI, machine learning, and technologies like edge computing. Ethical considerations gain prominence and quantum computing's potential is explored for handling massive datasets.

**PS29—Roles and capabilities of enterprise architecture in Big Data analytics technology adoption and implementation [48]**

The document explores the evolution and current state of Big Data Analytics, highlighting its diverse applications in domains like healthcare and finance.

Researchers have grappled with challenges such as data privacy and scalability, addressing them through innovations like advanced encryption and scalable algorithms.

Looking forward, emerging trends include the integration of Artificial Intelligence and Machine Learning for enhanced analytics and a growing focus on ethics and responsible data use. The intersection of Big Data with edge computing and IoT also opens new frontiers for real-time analytics.

**PS30—Security and privacy challenges of Big Data adoption: a qualitative study in telecommunication industry [49]**

The research investigates the evolution of Big Data Analytics applications across diverse domains, emphasizing healthcare, finance, marketing, and telecommunications.

Challenges include data security and privacy, addressed through advanced encryption and privacy-preserving techniques.

In the future, emerging trends highlight explainable AI, ethical data practices, and innovations in handling streaming data, graph databases, and blockchain integration.

**PS31—The role of AI, machine learning, and Big Data in digital twinning: a systematic literature review, challenges, and opportunities [50]**

The document explores diverse applications of Big Data Analytics across industries like healthcare, energy, and manufacturing. It underscores the evolution of these applications, highlighting a focus on optimization, diagnostics, and predictive analytics.

Challenges include data collection difficulties, picking the right AI models that are both accurate and fast and the ongoing need for standardization in digital twinning.

The document anticipates future trends, emphasizing the integration of AI, Machine Learning, and Big Data, particularly in digital twinning. It sets the stage for ongoing research in optimizing industrial processes, predictive analytics, healthcare, and smart city implementations.

**PS32—The state of the art and taxonomy of Big Data analytics: view from new Big Data framework [51]**

The document extensively explores the landscape of Big Data Analytics, emphasizing the dominant role of Hadoop while acknowledging the rise of Apache Spark in recent years.

Major challenges in the field involve handling diverse data formats, optimizing algorithms for evolving hardware configurations, and bridging the gap between complex systems and end-users through user-friendly visualization techniques.

It anticipates future advancements in applications, specifically in domains like e-commerce and the IoT, while expressing optimism about increased investments in Big Data technology.

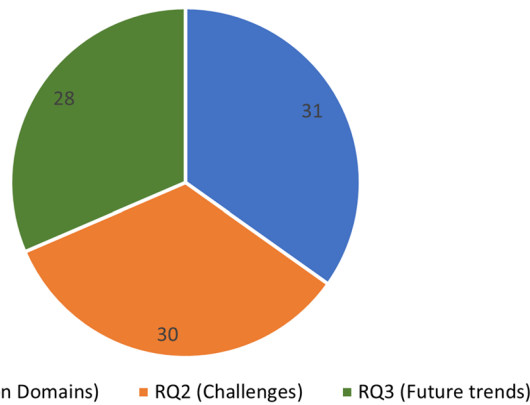
## Discussion

In the last 15 years, Big Data has found applications across various domains, evolving over time in line with the evolution of technologies and new business needs. Some of the most common application domains for Big Data Analytics include:

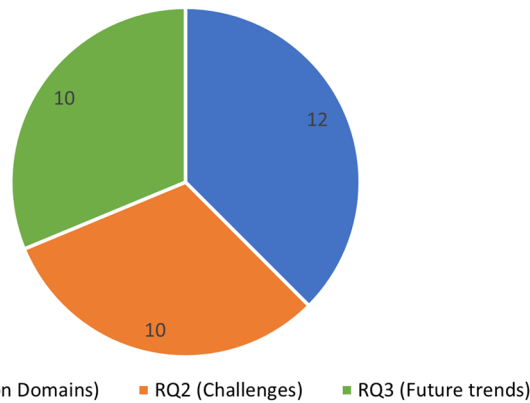
- Business and Finance, for example, to detect fraud detection by analyzing large datasets and identifying patterns indicative of fraudulent activities or to study customer behavior, preferences, and trends to improve marketing strategies.
- Healthcare, for example, to forecast disease outbreaks, patient admission rates, and treatment outcomes, or to personalize medicine with the analysis of genetic data for ad-hoc treatments.
- Retail, for example, to automatically manage and optimize inventories, and stock levels by predicting demands, or to create recommender systems to targeted and segmented customers' profiles.
- Manufacturing, for example, to predict and schedule maintenance needs and potential equipment failures by analyzing sensor data, or to improve product quality by monitoring and analyzing production processes.
- Telecommunications, for example, to optimize at real-time network performance and areas for improvement, or to predict customer churn by identifying factors and customers' behaviors that contribute to customer churn.
- Government, Public Services, and Transportation, for example, to plan efficient urban mobility, traffic management, and resource allocation in Smart Cities, or to predict and prevent criminal activities, or to optimize energy distribution and reduce wastage, or to optimize transportation routes, reduce delivery times, and vehicle fleets for efficiency and cost savings.
- Media, Entertainment, and Education, for example, to recommend movies, music, or articles based on users' behaviors and preferences, or to tailor content and advertising by studying users' behaviors, or to improve educational impact by analyzing student performance.

In Fig. 6, we show the distribution of the studies addressing the three research questions (RQ1-RQ3), from which we have started initially our investigation: 31 PSs discuss common application domains where the use of Big Data solutions is relevant (RQ1); 30 PSs analyze research challenges and limitations of Big Data (RQ2); 28 PSs highlight emerging research trends and directions in Big Data (RQ3). The total number of papers addressing the 3 RQs is different from the number of the selected 32 PSs, since we observed overlaps and intersections (e.g., a PS can address multiple RQs.)

To better understand the main focus of the PSs, Fig. 7 shows the distribution of studies addressing the three research questions, but this time, we made it avoiding intersections (i.e., each primary study can only be part of one of the 3 categories.) We can classify 12 PSs as papers that mainly focus on RQ1, 10 PSs mainly focus on RQ2, and 10 PSs on



**Fig. 6** Distribution of studies addressing the three research questions



**Fig. 7** Distribution of studies mainly addressing the three research questions

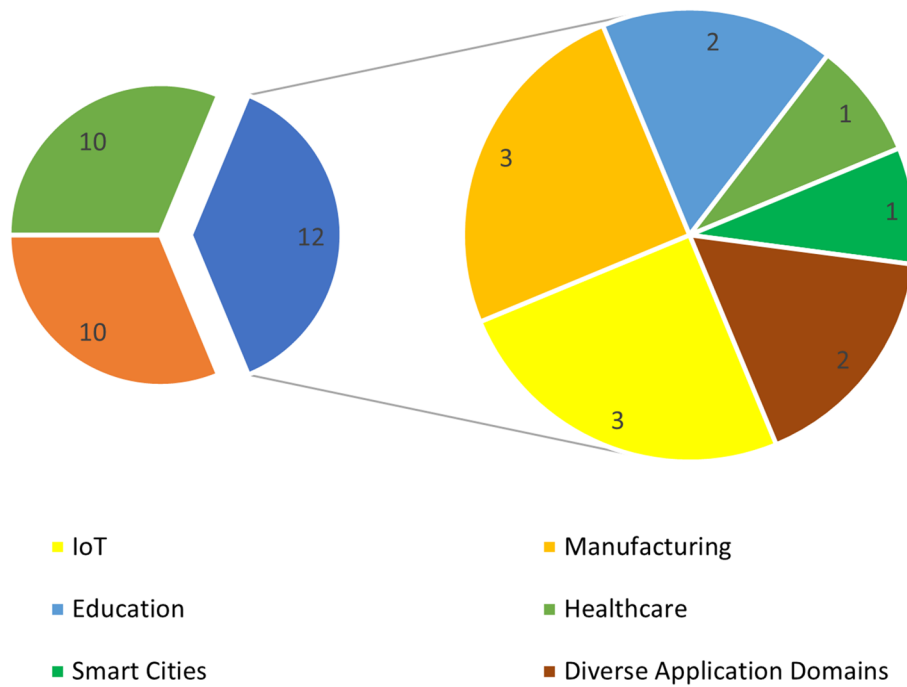
RQ3. The homogeneous distribution of the primary studies allows us to be optimistic about the results of our research since we had a good number of studies to answer each of our research questions.

To further make clear the main focuses of our studies, we decided to categorize each one. Figures 8, 9, and 10 show the focus of the documents for each Research Question (note that the sum of the categorized documents may be greater than the number of studies that answer that RQ, because they may overlap and be part of more than one category).

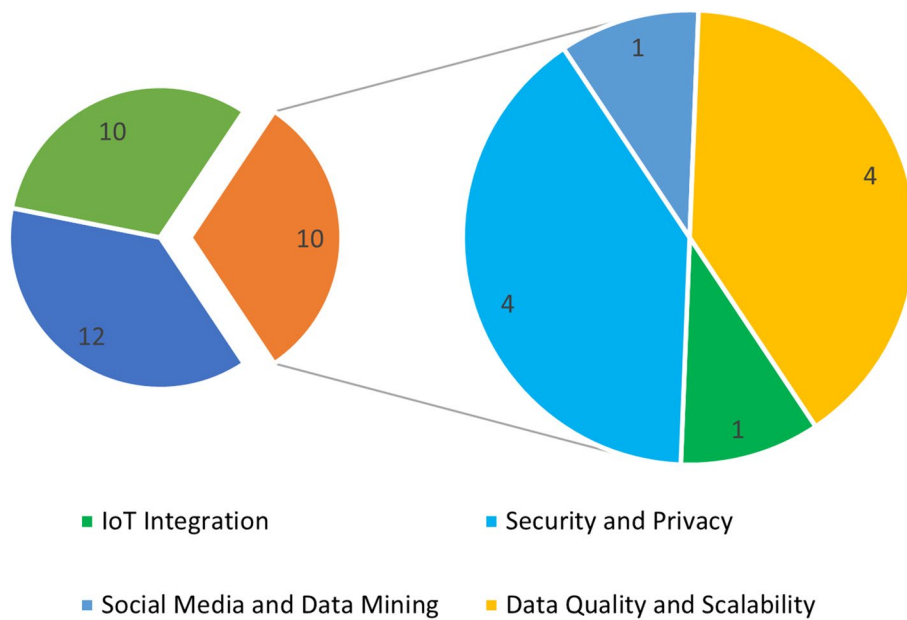
Having clarified this, we now discuss the findings of our SLR. We divided this discussion in three sections, one for each Research Question, so that we could clearly define which elements answer which question.

**RQ1: what are the most common application domains for Big Data analytics, and how have they evolved over time?**

Delving into the realm of Big Data across various sectors over the last 15 years reveals a narrative of evolution and adaptation. Initially rooted in finance, healthcare and marketing, the domain of Big Data analytics has undergone a metamorphosis, embracing applications from computational biology to education and manufacturing,

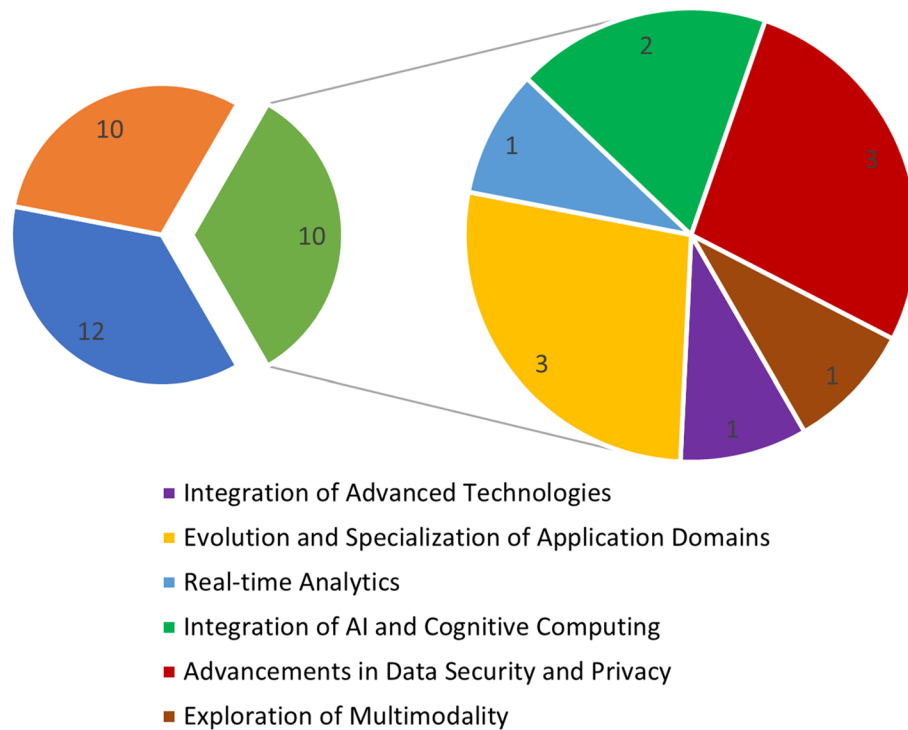


**Fig. 8** Categorization of RQ1 studies



**Fig. 9** Categorization of RQ2 studies

expanding into the avant-garde concept of digital twinning. This dynamic evolution is evident in studies investigating Big Data management techniques on the Internet of Things, where the focus has shifted from basic health state monitoring to sophisticated predictive modeling. This evolution signifies a maturation of Big Data analytics,



**Fig. 10** Categorization of RQ3 studies

with an increased focus on nuanced attributes like performance, efficiency, reliability, and scalability.

**RQ2: what are the major challenges and limitations that researchers have encountered in Big Data analysis, and how have they been addressed?**

Shifting our focus to the challenges within the Big Data analytics landscape, a complex history of persistent hurdles and inventive solutions comes into focus. The studies converge on a common thread, unraveling ongoing challenges encapsulated in the trio of data quality, scalability, and privacy/security concerns. Researchers faced with these challenges have become architects of innovative solutions, leveraging advanced algorithms, distributed frameworks, and privacy-preserving techniques. These solutions reflect a commitment to advancing the field in response to the complexities of handling vast and dynamic datasets.

In the implementation of Big Data Analytics, diverse challenges emerge. A dedicated study on industries points to crucial issues in data quality assessment models and secure frameworks. Here, the role of machine learning and data analytics, particularly in fault detection and reservoir management, becomes pivotal. The interconnected nature of these challenges emphasizes the importance of a comprehensive approach to implementation. Beyond technological challenges, ethical considerations surrounding data privacy and security take center stage. Researchers stress the significance of tools addressing ethical concerns, underlining that responsible deployment is intrinsic to the ethical use of Big Data Analytics.

In response to these challenges, the industry advocates for innovative solutions, emphasizing AI-driven approaches, cryptography mechanisms, and lightweight frameworks with AI. This recognition underscores the need for inventive strategies to navigate the intricate integration of Big Data into rapidly evolving technological landscapes.

### **RQ3: what are the emerging research trends and directions in Big Data that will likely shape the field in the next 5 to 10 years?**

Looking into the next 5 to 10 years, several trends are expected to shape the landscape of Big Data Analytics. One significant trend involves making data acquisition more energy-efficient, a move that aligns with broader sustainability goals. The integration of machine learning and deep learning techniques is anticipated to enhance the analytical capabilities of Big Data systems, enabling more accurate predictions and insights. Another noteworthy trend is the emphasis on edge and fog infrastructures, signifying a shift towards decentralized processing for faster data processing and decision-making, especially relevant in the context of the Internet of Things. Importantly, these trends extend beyond technological advancements to include ethical considerations. As Big Data assumes a pivotal role in decision-making processes, these ethical dimensions must be at the forefront. This involves dealing with the tricky ethical issues that come with having such a big influence through data analytics.

In essence, the trajectory of Big Data analytics in the coming years is a dual journey, one that advances technologically with a keen eye on efficiency and, concurrently, prioritizes ethical practices. It's a future where innovation and responsibility go hand in hand, defining a landscape that reflects both progress and ethical consciousness.

### **Threats to validity**

Ensuring the validity of a SLR is essential for the development of a reliable study. For this reason, in this section, we examine potential threats to construct, internal and external validity, aiming to maintain the robustness of our findings.

Construct validity determines whether the implementation of the SLR aligns with its initial objectives. The efficacy of our search process and the relevance of search terms are crucial concerns. While our search terms were derived from well-defined research questions and adjusted based on that, the completeness and comprehensiveness of these terms may be subject to limitations. Additionally, the use of different keywords might have returned other relevant studies that have not been taken into consideration. A potential language bias may also exist due to the exclusion of non-English articles, representing a limitation that should be acknowledged in the overall validity of the research.

Internal validity assesses the extent to which the design and execution of the study minimize systematic errors. A key focus is on the process of data extraction from the selected primary studies. Some required data may not have been explicitly expressed or were entirely missing, posing a potential threat to internal validity. To minimize this risk, the SLR process has been supervised by another person in order to minimize error into the process.

External validity examines the extent to which the observed effects of the study can be applied beyond its scope. In this SLR, we concentrated on research questions and quality assessments to mitigate the risk of limited generalizability. However, the

study's focus on the specific domain of Big Data research may limit external validity. Moreover, the dynamic nature of Big Data and the predefined time frame (last 15 years) could affect the generalizability of findings. Recognizing these constraints, the outcomes of this SLR are considered generalizable within the specified context of Big Data research.

By acknowledging these potential threats to validity, we strive to enhance the credibility and reliability of our SLR, contributing valuable insights to the evolving landscape of Big Data research.

## **Conclusion**

Over the past 15 years, Big Data has become a crucial player in various fields, adapting to technological shifts and meeting the changing needs of businesses. This review has taken a closer look at how Big Data has been applied, its challenges, and what we can expect in the near future. 189 studies were ultimately found, 32 of which were SLRs analyzed for this study.

Big Data started in areas like Business, Healthcare, and Marketing, but its influence has ultimately grown. Now, it helps predict disease outbreaks, manage retail inventory, forecast equipment failures in manufacturing, improve network performance, optimize urban planning, personalize media content, and enhance education.

Dealing with Big Data hasn't been without challenges. Issues like ensuring data quality, handling scalability, and maintaining privacy and security have been persistent. Researchers have responded with creative solutions, using advanced algorithms and privacy measures.

Looking to the future, the trends suggest exciting developments. Making data acquisition more energy-efficient and integrating advanced machine learning techniques are on the horizon. There is a shift toward decentralized processing, especially with the Internet of Things in mind. Importantly, these trends aren't just about technology; they also emphasize ethical considerations. Ethical issues need careful attention as Big Data becomes more influential in decision-making processes.

To summarize, the future of Big Data is a journey that combines technological progress with a strong ethical stance. It's a path where innovation and responsibility walk hand in hand, shaping a landscape that advances both technologically and ethically. The last 15 years have set the stage and the road ahead invites us to keep exploring and engaging with the ever-evolving world of Big Data.

## **Acknowledgments**

Not applicable.

## **Author contributions**

D.T. designed the SLR and wrote the main manuscript text R.K. conducted the SLR M.R. contributed to the Introduction, Challenges, and Conclusions All authors reviewed the manuscript.

## **Data availability**

No datasets were generated or analysed during the current study.

## **Declarations**

### **Competing interests**

The authors declare no competing interests.



Received: 5 February 2024 Accepted: 7 April 2024

Published online: 14 May 2024

## Bibliography

1. Tosi D, Campi AS. How schools affected the covid-19 pandemic in Italy: data analysis for Lombardy Region, Campania Region, and Emilia Region. *Future Internet*. 2021. <https://doi.org/10.3390/fi13050109>.
2. Davoudian A, Liu M. Big Data systems: a software engineering perspective. *ACM Comput Surv*. 2020. <https://doi.org/10.1145/3408314>.
3. Kushwaha AK, Kar AK. Language model-driven chatbot for business to address marketing and selection of products. In: Sharma SK, Dwivedi YK, Metri B, Rana NP, editors. *Re-imagining diffusion and adoption of information technology and systems: a continuing conversation*. Cham: Springer; 2020. p. 16–28. [https://doi.org/10.1007/978-3-030-64849-7\\_3](https://doi.org/10.1007/978-3-030-64849-7_3).
4. Kushwaha AK, Kar AK. Micro-foundations of artificial intelligence adoption in business: making the shift. In: Sharma SK, Dwivedi YK, Metri B, Rana NP, editors. *Re-imagining diffusion and adoption of information technology and systems: a continuing conversation*. Cham: Springer; 2020. p. 249–60. [https://doi.org/10.1007/978-3-030-64849-7\\_22](https://doi.org/10.1007/978-3-030-64849-7_22).
5. Dong W, Liao S, Zhang Z. Leveraging financial social media data for corporate fraud detection. *J Manag Inf Syst*. 2018;35(2):461–87. <https://doi.org/10.1080/07421222.2018.1451954>.
6. Kushwaha AK, Mandal S, Pharswan R, Kar AK, Ilavarasan PV. Studying online political behaviours as rituals: a study of social media behaviour regarding the CAA. In: Sharma SK, Dwivedi YK, Metri B, Rana NP, editors. *Re-imagining diffusion and adoption of information technology and systems: a continuing conversation*. Cham: Springer; 2020. p. 315–26. [https://doi.org/10.1007/978-3-030-64861-9\\_28](https://doi.org/10.1007/978-3-030-64861-9_28).
7. Fronzetti Colladon A, Gloor P, Iezzi DF. Editorial introduction: the power of words and networks. *Int J Inf Manag*. 2020;51: 102031. <https://doi.org/10.1016/j.ijinfomgt.2019.10.016>.
8. Kushwaha AK, Kar AK, Ilavarasan PV. Predicting retweet class using deep learning. In: Piuri V, Raj S, Genovese A, Srivastava R, editors. *Trends in deep learning methodologies: hybrid computational intelligence for pattern analysis*. Cambridge: Academic Press; 2021. p. 89–112. <https://doi.org/10.1016/B978-0-12-822226-3.00004-0>.
9. Tosi D. Cell phone Big Data to compute mobility scenarios for future smart cities. *Int J Data Sci Anal*. 2017;4:265–84. <https://doi.org/10.1007/s41060-017-0061-2>.
10. Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: from Big Data to big impact. *MIS Q*. 2012;36(4):1165–88. <https://doi.org/10.2307/41703503>.
11. Wamba SF, Ngai E, Riggins F, Akter S. *Big Data and business analytics adoption and use: a step toward transforming operations and production management?* Bingley: Emerald Group Publishing Limited; 2017.
12. George G, Osinga EC, Lavie D, Scott BA. *Big Data and data science methods for management research*. Acad Manag. 2016. <https://doi.org/10.5465/amj.2016.4005>.
13. Curtin J, Kauffman RJ, Riggins FJ. Making the 'MOST' out of RFID technology: a research agenda for the study of the adoption, usage and impact of RFID. *Inf Technol Manag*. 2007;8(2):87–110. <https://doi.org/10.1007/s10799-007-0010-1>.
14. Kitchenham BA, Charters S. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report. 2007. [https://www.elsevier.com/\\_\\_data/promis\\_misc/525444systematicreviewsguide.pdf](https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf). Accessed 15 Jan 2024
15. Tosi D, Morasca S. Supporting the semi-automatic semantic annotation of web services: a systematic literature review. *Inf Softw Technol*. 2015;61:16–32. <https://doi.org/10.1016/j.infsof.2015.01.007>.
16. Tahir A, Tosi D, Morasca S. A systematic review on the functional testing of semantic web services. *J Syst Softw*. 2013;86(11):2877–89. <https://doi.org/10.1016/j.jss.2013.06.064>.
17. Briner RB, Denyer D. 112 systematic review and evidence synthesis as a practice and scholarship tool. In: Rousseau DM, editor. *The Oxford handbook of evidence-based management*. Oxford: Oxford University Press; 2012. <https://doi.org/10.1093/oxfordhb/9780199763986.013.0007>.
18. Tosi D, Marzorati S, La Rosa M, Dondossola G, Terruggia R. Big Data from cellular networks: how to estimate energy demand at real-time. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE. 2015. pp. 1–10. <https://doi.org/10.1109/DSAA.2015.7344881>.
19. Cappi R, Casini L, Tosi D, Rocchetti M. Questioning the seasonality of SARS-COV-2: a Fourier spectral analysis. *BMJ Open*. 2022. <https://doi.org/10.1136/bmjopen-2022-061602>.
20. Naghib A, Jafari Navimipour N, Hosseinzadeh M, Sharifi A. A comprehensive and systematic literature review on the Big Data management techniques in the internet of things. *Wirel Netw*. 2023;29(3):1085–144. <https://doi.org/10.1007/s11276-022-03177-5>.
21. Sarker S, Arefin MS, Kowsher M, Bhuiyan T, Dhar PK, Kwon O-J. A comprehensive review on Big Data for industries: challenges and opportunities. *IEEE Access*. 2023;11:744–69. <https://doi.org/10.1109/ACCESS.2022.3232526>.
22. Bansal M, Chana I, Clarke S. A survey on IoT Big Data: current status, 13 V's challenges, and future directions. *ACM Comput Surv*. 2020. <https://doi.org/10.1145/3419634>.
23. Hordri NF, Samar A, Yuhaniz SS, Shamsuddin SM. A systematic literature review on features of deep learning in Big Data analytics. *Int J Adv Soft Comput Appl*. 2017;9(1):32–49.
24. Zhong Y, Chen L, Dan C, Rezaeiapanah A. A systematic survey of data mining and Big Data analysis in internet of things. *J Supercomput*. 2022;78(17):18405–53. <https://doi.org/10.1007/s11227-022-04594-1>.
25. Rashid ANMB. Access methods for Big Data: current status and future directions. *EAI Endorsed Trans Scal Inf Syst*. 2017;4(15):1–14. <https://doi.org/10.4108/eai.28-12-2017.153520>.

26. O'Donovan P, Leahy K, Bruton K, O'Sullivan DTJ. An industrial Big Data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *J Big Data*. 2015. <https://doi.org/10.1186/s40537-015-0034-z>.
27. Kushwaha AK, Kar AK, Dwivedi YK. Applications of Big Data in emerging management disciplines: a literature review using text mining. *Int J Inf Manag Data Insights*. 2021. <https://doi.org/10.1016/j.jjimei.2021.100017>.
28. Alkhalil A, Abdallah MAE, Alogali A, Aljaloud A. Applying Big Data analytics in higher education: a systematic mapping study. *Int J Inf Commun Technol Educ*. 2021;17(3):29–51. <https://doi.org/10.4018/IJICTE.20210701.oa3>.
29. Rahmani AM, Azhir E, Ali S, Mohammadi M, Ahmed OH, Ghafour MY, Ahmed SH, Hosseinzadeh M. Artificial intelligence approaches and mechanisms for Big Data analytics: a systematic study. *PeerJ Computer Sci*. 2021;7:1–28. <https://doi.org/10.7717/peerj-cs.488>.
30. Lundberg L. Bibliometric mining of research directions and trends for Big Data. *J Big Data*. 2023. <https://doi.org/10.1186/s40537-023-00793-6>.
31. Baig MI, Shuib L, Yadegaridehkordi E. Big Data adoption: state of the art and research challenges. *Inf Process Manag*. 2019. <https://doi.org/10.1016/j.ipm.2019.102095>.
32. Ikegwu AC, Nweke HF, Anikwe CV, Alo UR, Okonkwo OR. Big Data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Clust Comput*. 2022;25(5):343–87. <https://doi.org/10.1007/s10586-022-03568-5>.
33. Imran S, Mahmood T, Morshed A, Sellis T. Big Data analytics in healthcare: a systematic literature review and roadmap for practical implementation. *IEEE/CAA J Autom Sin*. 2021;8(1):1–22. <https://doi.org/10.1109/JAS.2020.1003384>.
34. Zahid H, Mahmood T, Morshed A, Sellis T. Big Data analytics in telecommunications: literature review and architecture recommendations. *IEEE/CAA J Autom Sin*. 2020;7(1):18–38. <https://doi.org/10.1109/JAS.2019.1911795>.
35. Bazzaz Abkenar S, Haghi Kashani M, Mahdipour E, Jameii SM. Big Data analytics meets social media: a systematic review of techniques, open issues, and future directions. *Telemat Inf*. 2021. <https://doi.org/10.1016/j.tele.2020.101517>.
36. ElSayed IA, ElDahshan K, Hefny H, ElSayed EK. Big Data and its future in computational biology: a literature review. *J Computer Sci*. 2021;17(12):1222–8. <https://doi.org/10.3844/jcssp.2021.1222.1228>.
37. Hajjali M. Big Data and sentiment analysis: a comprehensive and systematic literature review. *Concurr Comput Pract Exp*. 2020. <https://doi.org/10.1002/cpe.5671>.
38. Ahmadova U, Mustafayev M, Kiani Kalejahi B, Saeedvand S, Rahmani AM. Big Data applications on the internet of things: a systematic literature review. *Int J Commun Syst*. 2021. <https://doi.org/10.1002/dac.5004>.
39. Baig MI, Shuib L, Yadegaridehkordi E. Big Data in education: a state of the art, limitations, and future research directions. *Int J Educ Technol High Educ*. 2020;17(1):1–23.
40. Reshi AA, Shah ARIF, Shafi S, Qadri MH. Big Data in healthcare a comprehensive bibliometric analysis of current research trends. *Scal Comput*. 2023;24(3):531–49. <https://doi.org/10.12694/scpe.v24i3.2155>.
41. Pulikottil T, Estrada-Jimenez LA, Abadia JJP, Carrera-Rivera A, Torayev A, Rehman HU, Mo F, Nikghadam-Hojjati S, Barata J. Big Data life cycle in shop-floor-trends and challenges. *IEEE Access*. 2023;11:30008–26. <https://doi.org/10.1109/ACCESS.2023.3253286>.
42. Arshad I, Alsamhi SH, Afzal W. Big Data testing techniques: taxonomy, challenges and future trends. *Computers Mater Contin*. 2023;74(2):2739–70. <https://doi.org/10.32604/cmc.2023.030266>.
43. Gupta S, Kar AK, Baabdullah A, Al-Khowaiter WAA. Big Data with cognitive computing: a review for the future. *Int J Inf Manag*. 2018;42:78–89. <https://doi.org/10.1016/j.jjinfomgt.2018.06.005>.
44. Saltz JS, Krasteva I. Current approaches for executing Big Data science projects—a systematic literature review. *PeerJ Computer Sci*. 2022. <https://doi.org/10.7717/PEERJ-CS.862>.
45. Liu C, Peng G, Kong Y, Li S, Chen S. Data quality affecting Big Data analytics in smart factories: research themes, issues and methods. *Symmetry*. 2021. <https://doi.org/10.3390/sym13081440>.
46. Ahmed A, Xi R, Hou M, Shah SA, Hameed S. Harnessing Big Data analytics for healthcare: a comprehensive review of frameworks, implications, applications, and impacts. *IEEE Access*. 2023;11:112891–928. <https://doi.org/10.1109/ACCESS.2023.3323574>.
47. Karimi Y, Haghi Kashani M, Akbari M, Mahdipour E. Leveraging Big Data in smart cities: a systematic review. *Concurr Comput Pract Exp*. 2021. <https://doi.org/10.1002/cpe.6379>.
48. Gong Y, Janssen M. Roles and capabilities of enterprise architecture in Big Data analytics technology adoption and implementation. *J Theor Appl Electron Commer Res*. 2021;16(1):37–51. <https://doi.org/10.4067/S0718-1876021000100104>.
49. Anawar S, Othman NF, Selamat SR, Ayop Z, Harum N, Rahim FA. Security and privacy challenges of Big Data adoption: a qualitative study in telecommunication industry. *Int J Interact Mob Technol*. 2022;16(19):81–97. <https://doi.org/10.3991/ijim.v16i19.32093>.
50. Rathore MM, Shah SA, Shukla D, Bentafat E, Bakiras S. The role of AI, machine learning, and Big Data in digital twinning: a systematic literature review, challenges, and opportunities. *IEEE Access*. 2021;9:32030–52. <https://doi.org/10.1109/ACCESS.2021.3060863>.
51. Mohamed A, Najafabadi MK, Wah YB, Zaman EAK, Maskat R. The state of the art and taxonomy of Big Data analytics: view from new Big Data framework. *Artif Intell Rev*. 2020;53(2):989–1037. <https://doi.org/10.1007/s10462-019-09685-9>.
52. Li Y, Liu Z, Zhu H. Enterprise search in the Big Data era: recent developments and open challenges. *Proc VLDB Endow*. 2014;7(13):1717–8. <https://doi.org/10.14778/2733004.2733071>.
53. Lee D, Camacho D, Jung JJ. Smart mobility with Big Data: approaches, applications, and challenges. *Appl Sci*. 2023. <https://doi.org/10.3390/app13127244>.
54. Himeur Y, Elnour M, Fadli F, Meskin N, Petri I, Rezzgui Y, Bensaali F, Amira A. AI-Big Data analytics for building automation and management systems: a survey, actual challenges and future perspectives. *Artif Intell Rev*. 2023;56(6):4929–5021. <https://doi.org/10.1007/s10462-022-10286-2>.

55. Cesario E. Big Data analytics and smart cities: applications, challenges, and opportunities. *Front Big Data*. 2023. <https://doi.org/10.3389/fdata.2023.1149402>.
56. Zwilling M. Big Data challenges in social sciences: an NLP analysis. *J Computer Inf Syst*. 2023;63(3):537–54. <https://doi.org/10.1080/08874417.2022.2085211>.
57. Rani R, Khurana M, Kumar A, Kumar N. Big Data dimensionality reduction techniques in IoT: review, applications and open research challenges. *Clust Comput*. 2022;25(6):4027–49. <https://doi.org/10.1007/s10586-022-03634-y>.
58. Jagatheesaperumal SK, Rahouti M, Ahmad K, Al-Fuqaha A, Guizani M. The duo of artificial intelligence and Big Data for industry 4.0: applications, techniques, challenges, and future research directions. *IEEE Internet Things J*. 2022;9(15):12861–85. <https://doi.org/10.1109/JIOT.2021.3139827>.
59. Lundberg L, Grahn H. Research trends, enabling technologies and application areas for Big Data. *Algorithms*. 2022. <https://doi.org/10.3390/a15080280>.
60. Ali TAL, Khafagy MH, Farrag MH. Big Data challenges: preserving techniques for privacy violations. *J Theor Appl Inf Technol*. 2022;100(8):2505–17.
61. Latifian A. How does cloud computing help businesses to manage Big Data issues. *Kybernetes*. 2022;51(6):1917–48. <https://doi.org/10.1108/K-05-2021-0432>.
62. Rehman A, Naz S, Razzak I. Leveraging Big Data analytics in healthcare enhancement: trends, challenges and opportunities. *Multimed Syst*. 2022;28(4):1339–71. <https://doi.org/10.1007/s00530-020-00736-8>.
63. Al-Zahrani A, Al-Hebbi M. Big Data major security issues: challenges and defense strategies. *Tehnicki Glasnik*. 2022;16(2):197–204. <https://doi.org/10.31803/tg-20220124135330>.
64. Song X, Zhang H, Akerkar R, Huang H, Guo S, Zhong L, Ji Y, Opdahl AL, Purohit H, Skupin A, Pottathil A, Culotta A. Big Data and emergency management: concepts, methodologies, and applications. *IEEE Trans Big Data*. 2022;8(2):397–419. <https://doi.org/10.1109/TBDDATA.2020.2972871>.
65. Singh N, Singh DP, Pant B. Big Data knowledge discovery as a service: recent trends and challenges. *Wirel Pers Commun*. 2022;123(2):1789–807. <https://doi.org/10.1007/s11277-021-09213-5>.
66. Mohammadi E, Karami A. Exploring research trends in Big Data across disciplines: a text mining analysis. *J Inf Sci*. 2022;48(1):44–56. <https://doi.org/10.1177/0165551520932855>.
67. Ambeth Kumar VD, Varadarajan V, Gupta MK, Rodrigues JJPC, Janu N. AI empowered Big Data analytics for industrial applications. *J Univers Computer Sci*. 2022;28(9):877–81. <https://doi.org/10.3897/jucs.94155>.
68. Kumari S, Muthulakshmi P. Transformative effects of Big Data on advanced data analytics: open issues and critical challenges. *J Computer Sci*. 2022;18(6):463–79. <https://doi.org/10.3844/jcssp.2022.463.479>.
69. Tang S, He B, Yu C, Li Y, Li K. A survey on spark ecosystem: Big Data processing infrastructure, machine learning, and applications. *IEEE Trans Knowl Data Eng*. 2022;34(1):71–91. <https://doi.org/10.1109/TKDE.2020.2975652>.
70. Reyes-Veras PF, Renukappa S, Suresh S. Challenges faced by the adoption of Big Data in the Dominican Republic construction industry: an empirical study. *J Inf Technol Constr*. 2021;26:812–31. <https://doi.org/10.36680/J.ITCON.2021.044>.
71. Bentotahewa V, Hewage C, Williams J. Solutions to Big Data privacy and security challenges associated with COVID-19 surveillance systems. *Front Big Data*. 2021. <https://doi.org/10.3389/fdata.2021.645204>.
72. Escobar CA, McGovern ME, Morales-Menendez R. Quality 4.0: a review of Big Data challenges in manufacturing. *J Intell Manuf*. 2021;32(8):2319–34. <https://doi.org/10.1007/s10845-021-01765-4>.
73. Mwitondi KS, Said RA. Dealing with randomness and concept drift in large datasets. *Data*. 2021. <https://doi.org/10.3390/data6070077>.
74. Kusal S, Patil S, Kotecha K, Aluvalu R, Varadarajan V. Ai based emotion detection for textual Big Data: techniques and contribution. *Big Data Cognit Comput*. 2021. <https://doi.org/10.3390/bdcc5030043>.
75. Lee E, Jang J. Research trend analysis for sustainable QR code use: focus on Big Data analysis. *KSII Trans Internet Inf Syst*. 2021;15(9):3221–42. <https://doi.org/10.3837/tiis.2021.09.008>.
76. Rahlha M, Allegue S, Abdellatif T. Guidelines for GDPR compliance in Big Data systems. *J Inf Secur Appl*. 2021. <https://doi.org/10.1016/j.jisa.2021.102896>.
77. Amović M, Govedarica M, Radulović A, Janković I. Big Data in smart city: management challenges. *Appl Sci*. 2021. <https://doi.org/10.3390/app11104557>.
78. Hoozemans J, Peltenburg J, Nonnemacher F, Hadnagy A, Al-Ars Z, Hofstee HP. FPGA acceleration for Big Data analytics: challenges and opportunities. *IEEE Circuits Syst Mag*. 2021;21(2):30–47. <https://doi.org/10.1109/MCAS.2021.3071608>.
79. Jalali SMJ, Park HW, Vanani IR, Pho K-H. Research trends on Big Data domain using text mining algorithms. *Digit Scholarsh Hum*. 2021;36(2):361–70. <https://doi.org/10.1093/llc/fqaa012>.
80. Almutairi MM. Role of Big Data in education in KSA. *Int J Inf Technol*. 2021;13(1):367–73. <https://doi.org/10.1007/s41870-020-00489-7>.
81. Ardagna D, Barbierato E, Gianniti E, Gribaudo M, Pinto TBM, Silva APC, Almeida JM. Predicting the performance of Big Data applications on the cloud. *J Supercomput*. 2021;77(2):1321–53. <https://doi.org/10.1007/s11227-020-03307-w>.
82. Mkrtrtchian V, Gamidullaeva L, Finogeev A, Chernyshenko S, Chernyshenko V, Amirov D, Potapova I. Big Data and internet of things (IoT) technologies' influence on higher education: current state and future prospects. *Int J Web-Based Learn Teach Technol*. 2021;16(5):137–57. <https://doi.org/10.4018/IJWLTT.20210901.0a8>.
83. Mourtzis D. Towards the 5th industrial revolution: a literature review and a framework for process optimization based on Big Data analytics and semantics. *J Mach Eng*. 2021;21(3):5–39. <https://doi.org/10.36897/jme/141834>.
84. Dias MNR, Hassan S, Shahzad A. The impact of Big Data utilization on Malaysian government hospital healthcare performance. *Int J eBus eGov Stud*. 2021;13(1):50–77. <https://doi.org/10.34111/ijebeg.202113103>.
85. Babar M, Alshehri MD, Tariq MU, Ullah F, Khan A, Uddin MI, Almasoud AS. IoT-enabled Big Data analytics architecture for multimedia data communications. *Wirel Commun Mob Comput*. 2021. <https://doi.org/10.1155/2021/5283309>.
86. Bhat SA, Huang N-F. Big Data and AI revolution in precision agriculture: survey and challenges. *IEEE Access*. 2021;9:110209–22. <https://doi.org/10.1109/ACCESS.2021.3102227>.

87. Zainab A, Ghrayeb A, Syed D, Abu-Rub H, Refaat SS, Bouhali O. Big Data management in smart grids: technologies and challenges. *IEEE Access*. 2021;9:73046–59. <https://doi.org/10.1109/ACCESS.2021.3080433>.
88. Jabir B, Falih N. Big Data analytics opportunities and challenges for the smart enterprise. *Int J Tech Phys Probl Eng*. 2021;13(2):20–6.
89. Zineb EF, Najat R, Jaafar A. An intelligent approach for data analysis and decision making in Big Data: a case study on e-commerce industry. *Int J Adv Computer Sci Appl*. 2021;12(7):723–36. <https://doi.org/10.14569/IJACSA.2021.0120783>.
90. Syed D, Zainab A, Ghrayeb A, Refaat SS, Abu-Rub H, Bouhali O. Smart grid Big Data analytics: survey of technologies, techniques, and applications. *IEEE Access*. 2021;9:59564–85. <https://doi.org/10.1109/ACCESS.2020.3041178>.
91. Talebkah M, Sali A, Marjani M, Gordan M, Hashim SJ, Rokhani FZ. IoT and Big Data applications in smart cities: recent advances, challenges, and critical issues. *IEEE Access*. 2021;9:55465–84. <https://doi.org/10.1109/ACCESS.2021.3070905>.
92. Dubuc T, Stahl F, Roesch EB. Mapping the Big Data landscape: technologies, platforms and paradigms for real-time analytics of data streams. *IEEE Access*. 2021;9:15351–74. <https://doi.org/10.1109/ACCESS.2020.3046132>.
93. Ang KL-M, Seng JKP. Big Data and machine learning with hyperspectral information in agriculture. *IEEE Access*. 2021;9:36699–718. <https://doi.org/10.1109/ACCESS.2021.3051196>.
94. Zeadally S, Siddiqui F, Baig Z, Ibrahim A. Smart healthcare: challenges and potential solutions using internet of things (IoT) and Big Data analytics. *PSU Res Rev*. 2020;4(2):149–68. <https://doi.org/10.1108/PRR-08-2019-0027>.
95. Thudumu S, Branch P, Jin J, Singh JJ. A comprehensive survey of anomaly detection techniques for high dimensional Big Data. *J Big Data*. 2020. <https://doi.org/10.1186/s40537-020-00320-x>.
96. Trang NH. Limitations of Big Data partitions technology. *J Appl Data Sci*. 2020;1(1):11–9. <https://doi.org/10.47738/jads.v1i1.7>.
97. Caíno-Lores S, Lapin A, Carretero J, Kropf P. Applying Big Data paradigms to a large scale scientific workflow: lessons learned and future directions. *Future Gener Computer Syst*. 2020;110:440–52. <https://doi.org/10.1016/j.future.2018.04.014>.
98. Awaysheh FM, Alazab M, Gupta M, Pena TF, Cabaleiro JC. Next-generation Big Data federation access control: a reference model. *Future Gener Computer Syst*. 2020;108:726–41. <https://doi.org/10.1016/j.future.2020.02.052>.
99. Valencia-Parra A, Varela-Vaca AJ, Parody L, Gomez-Lopez MT. Unleashing constraint optimisation problem solving in Big Data environments. *J Comput Sci*. 2020. <https://doi.org/10.1016/j.jocs.2020.101180>.
100. López-Martínez F, Núñez-Valdez ER, García-Díaz V, Bursac Z. A case study for a Big Data and machine learning platform to improve medical decision support in population health management. *Algorithms*. 2020. <https://doi.org/10.3390/A13040102>.
101. Iqbal R, Doctor F, More B, Mahmud S, Yousuf U. Big Data analytics and computational intelligence for cyber-physical systems: recent trends and state of the art applications. *Future Gener Computer Syst*. 2020;105:766–78. <https://doi.org/10.1016/j.future.2017.10.021>.
102. Carnevale L, Celesti A, Fazio M, Villari M. A Big Data analytics approach for the development of advanced cardiology applications. *Information*. 2020. <https://doi.org/10.3390/info11020060>.
103. Shukla AK, Muhuri PK, Abraham A. A bibliometric analysis and cutting-edge overview on fuzzy techniques in Big Data. *Eng Appl Artif Intell*. 2020. <https://doi.org/10.1016/j.engappai.2020.103625>.
104. Karim A, Siddiqa A, Safdar Z, Razaq M, Gillani SA, Tahir H, Kiran S, Ahmed E, Imran M. Big Data management in participatory sensing: issues, trends and future directions. *Future Gener Computer Syst*. 2020;107:942–55. <https://doi.org/10.1016/j.future.2017.10.007>.
105. Humayun M. Role of emerging IoT Big Data and cloud computing for real time application. *Int J Adv Computer Sci Appl*. 2020;11(4):494–506.
106. Rabanal F, Martínez C. Cryptography for Big Data environments: current status, challenges, and opportunities. *Comput Math Methods*. 2020. <https://doi.org/10.1002/cmm4.1075>.
107. Ramesh T, Santhi V. Exploring Big Data analytics in health care. *Int J Intell Netw*. 2020;1:135–40. <https://doi.org/10.1016/j.jjin.2020.11.003>.
108. Gautam A, Chatterjee I. Big Data and cloud computing: a critical review. *Int J Oper Res Inf Syst*. 2020;11(3):19–38. <https://doi.org/10.4018/IJORIS.2020070102>.
109. Bajaber F, Sakr S, Batarfi O, Altalhi A, Barnawi A. Benchmarking Big Data systems: a survey. *Computer Commun*. 2020;149:241–51. <https://doi.org/10.1016/j.comcom.2019.10.002>.
110. Maksimov P, Koironen T. Application of novel Big Data processing techniques in process industries. *Int J Computer Appl Technol*. 2020;62(3):200–15. <https://doi.org/10.1504/IJCAT.2020.106591>.
111. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big Data in healthcare: management, analysis and future prospects. *J Big Data*. 2019. <https://doi.org/10.1186/s40537-019-0217-0>.
112. Nagalakshmi N, Anand Babu GL, Reddy KS, Ashalatha T. Security challenges associated with Big Data in health care system. *Int J Eng Adv Technol*. 2019;9(1):4057–60. <https://doi.org/10.35940/ijeat.A1296.109119>.
113. Dai H-N, Wong RC-W, Wang H, Zheng Z, Vasilakos AV. Big Data analytics for large-scale wireless networks: challenges and opportunities. *ACM Comput Surv*. 2019. <https://doi.org/10.1145/3337065>.
114. Barika M, Garg S, Zomaya AY, Wang L, Moorsel AVAN, Ranjan R. Orchestrating Big Data analysis workflows in the cloud: research challenges, survey, and future directions. *ACM Comput Surv*. 2019. <https://doi.org/10.1145/3332301>.
115. Hariri RH, Fredericks EM, Bowers KM. Uncertainty in Big Data analytics: survey, opportunities, and challenges. *J Big Data*. 2019. <https://doi.org/10.1186/s40537-019-0206-3>.
116. Latif Z, Lei W, Latif S, Pathan ZH, Ullah R, Jianqiu Z. Big Data challenges: prioritizing by decision-making process using analytic network process technique. *Multimed Tools Appl*. 2019;78(19):27127–53. <https://doi.org/10.1007/s11042-017-5161-4>.
117. Kumari A, Tanwar S, Tyagi S, Kumar N. Verification and validation techniques for streaming Big Data analytics in internet of things environment. *IET Netw*. 2019;8(3):155–63. <https://doi.org/10.1049/iet-net.2018.5187>.

118. Singh SP, Nayyar A, Kumar R, Sharma A. Fog computing: from architecture to edge computing and Big Data processing. *J Supercomput.* 2019;75(4):2070–105. <https://doi.org/10.1007/s11227-018-2701-2>.
119. Raufi B, Ismaili F, Ajdari J, Zenuni X. Web personalization issues in Big Data and semantic web: challenges and opportunities. *Turk J Electr Eng Computer Sci.* 2019;27(4):2379–94. <https://doi.org/10.3906/elk-1812-25>.
120. Rahman NA, Nor NM. Healthcare using social media: Big Data analytics perspective. *J Adv Res Dyn Control Syst.* 2019;11(8 Special Issue):1169–79.
121. Mishra S, Pattnaik S, Mishra BB. Application of Big Data analysis in supply chain management: future challenges. *J Adv Res Dyn Control Syst.* 2019;11(8 Special Issue):2541–8.
122. Ivanovic M, Klasnja-Milicevic A. Big Data and collective intelligence. *Int J Embed Syst.* 2019;11(5):573–83. <https://doi.org/10.1504/IJES.2019.102430>.
123. Qolomany B, Al-Fuqaha A, Gupta A, Benhaddou D, Alwajidi S, Qadir J, Fong AC. Leveraging machine learning and Big Data for smart buildings: a comprehensive survey. *IEEE Access.* 2019;7:90316–56. <https://doi.org/10.1109/ACCESS.2019.2926642>.
124. Shah SA, Seker DZ, Hameed S, Draheim D. The rising role of Big Data analytics and IoT in disaster management: recent advances, taxonomy and prospects. *IEEE Access.* 2019;7:54595–614. <https://doi.org/10.1109/ACCESS.2019.2913340>.
125. Lin W, Zhang Z, Peng S. Academic research trend analysis based on Big Data technology. *Int J Comput Sci Eng.* 2019;20(1):31–9. <https://doi.org/10.1504/ijcse.2019.103247>.
126. Hong L, Luo M, Wang R, Lu P, Lu W, Lu L. Big Data in health care: applications and challenges. *Data Inf Manag.* 2018;2(3):175–97. <https://doi.org/10.2478/dim-2018-0014>.
127. Pal D, Triyason T, Padungweang P. Big Data in smart-cities: current research and challenges. *Indones J Electr Eng Inf.* 2018;6(4):351–60. <https://doi.org/10.11591/ijeel.v6i4.543>.
128. Li N, Mahalik NP. A Big Data and cloud computing specification, standards and architecture: agricultural and food informatics. *Int J Inf Commun Technol.* 2019;14(2):159–74. <https://doi.org/10.1504/IJICT.2019.097687>.
129. Chiroma H, Abdullahi UA, Abdulhamid SM, Abdulsalam Alarood A, Gabralla LA, Rana N, Shuib L, Targio Hashem IA, Gbenga DE, Abubakar AI, Zeki AM, Herawan T. Progress on artificial neural networks for Big Data analytics: a survey. *IEEE Access.* 2019;7:70535–51. <https://doi.org/10.1109/ACCESS.2018.2880694>.
130. Waheed H, Hassan S-U, Aljohani NR, Wasif M. A bibliometric perspective of learning analytics research landscape. *Behav Inf Technol.* 2018;37(10–11):941–57. <https://doi.org/10.1080/0144929X.2018.1467967>.
131. Ray J, Johnny O, Trovati M, Sotiriadis S, Bessis N. The rise of Big Data science: a survey of techniques, methods and approaches in the field of natural language processing and network theory. *Big Data Cognit Comput.* 2018;2(3):1–18. <https://doi.org/10.3390/bdcc2030022>.
132. Mantelero A. AI and Big Data: a blueprint for a human rights, social and ethical impact assessment. *Computer Law Secur Rev.* 2018;34(4):754–72. <https://doi.org/10.1016/j.clsr.2018.05.017>.
133. Sultan K, Ali H, Zhang Z. Big Data perspective and challenges in next generation networks. *Future Internet.* 2018. <https://doi.org/10.3390/fi10070056>.
134. Li Q, Chen Y, Wang J, Chen Y, Chen H. Web media and stock markets : a survey and future directions from a Big Data perspective. *IEEE Trans Knowl Data Eng.* 2018;30(2):381–99. <https://doi.org/10.1109/TKDE.2017.2763144>.
135. Jabbar S, Malik KR, Ahmad M, Aldabbas O, Asif M, Khalid S, Han K, Ahmed SH. A methodology of real-time data fusion for localized Big Data analytics. *IEEE Access.* 2018;6:24510–20. <https://doi.org/10.1109/ACCESS.2018.2820176>.
136. Darwish TSJ, Abu Bakar K. Fog based intelligent transportation Big Data analytics in the internet of vehicles environment: motivations, architecture, challenges, and critical issues. *IEEE Access.* 2018;6:15679–701. <https://doi.org/10.1109/ACCESS.2018.2815989>.
137. Zheng S, Chen S, Yang L, Zhu J, Luo Z, Hu J, Yang X. Big Data processing architecture for radio signals empowered by deep learning: concept, experiment, applications and challenges. *IEEE Access.* 2018;6:55907–22. <https://doi.org/10.1109/ACCESS.2018.2872769>.
138. Stefanowski J, Krawiec K, Wrembel R. Exploring complex and Big Data. *Int J Appl Math Computer Sci.* 2017;27(4):669–79. <https://doi.org/10.1515/amcs-2017-0046>.
139. Harerimana G, Jang B, Kim JW, Park HK. Health Big Data analytics: a technology survey. *IEEE Access.* 2018;6:65661–78. <https://doi.org/10.1109/ACCESS.2018.2878254>.
140. Ravi S, Jeyaprakash T. Combined ideas on the necessity of Big Data on internet of things and researchers point of view and its challenges, future directions. *J Adv Res Dyn Control Syst.* 2018;10(9 Special Issue):2140–4.
141. Neggers J, Allix O, Hild F, Roux S. Big Data in experimental mechanics and model order reduction: today's challenges and tomorrow's opportunities. *Arch Comput Methods Eng.* 2018;25(1):143–64. <https://doi.org/10.1007/s11831-017-9234-3>.
142. Khan S, Liu X, Shakil KA, Alam M. A survey on scholarly data: from Big Data perspective. *Inf Process Manag.* 2017;53(4):923–44. <https://doi.org/10.1016/j.ipm.2017.03.006>.
143. Costa C, Santos MY. Big Data: state-of-the-art concepts, techniques, technologies, modeling approaches and research challenges. *IAENG Int J Computer Sci.* 2017;44(3):285–301.
144. Lv Z, Song H, Basanta-Val P, Steed A, Jo M. Next-generation Big Data analytics: state of the art, challenges, and future research topics. *IEEE Trans Ind Inf.* 2017;13(4):1891–9. <https://doi.org/10.1109/TII.2017.2650204>.
145. Memon MA, Soomro S, Jumani AK, Kartio MA. Big Data analytics and its applications. *Ann Emerg Technol Comput.* 2017;1(1):45–54. <https://doi.org/10.33166/AETIC.2017.01.006>.
146. Mantelero A. Regulating Big Data The guidelines of the Council of Europe in the context of the European data protection framework. *Computer Law Secur Rev.* 2017;33(5):584–602. <https://doi.org/10.1016/j.clsr.2017.05.011>.
147. Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on Big Data: opportunities and challenges. *Neurocomputing.* 2017;237:350–61. <https://doi.org/10.1016/j.neucom.2017.01.026>.
148. Yan J, Meng Y, Lu L, Li L. Industrial Big Data in an industry 4.0 environment: challenges, schemes, and applications for predictive maintenance. *IEEE Access.* 2017;5:23484–91. <https://doi.org/10.1109/ACCESS.2017.2765544>.

149. Gonçalves ME. The EU data protection reform and the challenges of Big Data: remaining uncertainties and ways forward. *Inf Commun Technol Law*. 2017;26(2):90–115. <https://doi.org/10.1080/13600834.2017.1295838>.
150. Peng S, Wang G, Xie D. Social influence analysis in social networking Big Data: opportunities and challenges. *IEEE Netw*. 2017;31(1):11–7. <https://doi.org/10.1109/MNET.2016.1500104NM>.
151. L'Heureux A, Grolinger K, Elyamany HF, Capretz MAM. Machine learning with Big Data: challenges and approaches. *IEEE Access*. 2017;5:7776–97. <https://doi.org/10.1109/ACCESS.2017.2696365>.
152. Manikyam NRH, Mohan Kumar S. Methods and techniques to deal with Big Data analytics and challenges in cloud computing environment. *Int J Civil Eng Technol*. 2017;8(4):669–78.
153. El-Seoud SA, El-Sofany HF, Abdelfattah M, Mohamed R. Big Data and cloud computing: trends and challenges. *Int J Interact Mob Technol*. 2017;11(2):34–52. <https://doi.org/10.3991/ijim.v11i2.6561>.
154. Zúñiga H, Diehl T. Citizenship, social media, and Big Data: current and future research in the social sciences. *Soc Sci Computer Rev*. 2017;35(1):3–9. <https://doi.org/10.1177/0894439315619589>.
155. Wang H, Xu Z, Pedrycz W. An overview on the roles of fuzzy set techniques in Big Data processing: trends, challenges and opportunities. *Knowl-Based Syst*. 2017;118:15–30. <https://doi.org/10.1016/j.knsys.2016.11.008>.
156. Choi T-M, Chan HK, Yue X. Recent development in Big Data analytics for business operations and risk management. *IEEE Trans Cybern*. 2017;47(1):81–92. <https://doi.org/10.1109/TCYB.2015.2507599>.
157. Zhong RY, Newman ST, Huang GQ, Lan S. Big Data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives. *Comput Ind Eng*. 2016;101:572–91. <https://doi.org/10.1016/j.cie.2016.07.013>.
158. Bajaber F, Elshawi R, Batarfi O, Altalhi A, Barnawi A, Sakr S. Big Data 2.0 processing systems: taxonomy and open challenges. *J Grid Comput*. 2016;14(3):379–405. <https://doi.org/10.1007/s10723-016-9371-1>.
159. De Gennaro M, Paffumi E, Martini G. Big Data for supporting low-carbon road transport policies in Europe: applications, challenges and opportunities. *Big Data Res*. 2016;6:11–25. <https://doi.org/10.1016/j.bdr.2016.04.003>.
160. Wang H, Xu Z, Fujita H, Liu S. Towards felicitous decision making: an overview on challenges and trends of Big Data. *Inf Sci*. 2016;367–368:747–65. <https://doi.org/10.1016/j.ins.2016.07.007>.
161. Rodríguez-Mazahua L, Rodríguez-Enríquez C-A, Sánchez-Cervantes JL, Cervantes J, García-Alcaraz JL, Alor-Hernández G. A general perspective of Big Data: applications, tools, challenges and trends. *J Supercomput*. 2016;72(8):3073–113. <https://doi.org/10.1007/s11227-015-1501-1>.
162. Bello-Ortiz G, Jung JJ, Camacho D. Social Big Data: recent achievements and new challenges. *Inf Fus*. 2016;28:45–59. <https://doi.org/10.1016/j.inffus.2015.08.005>.
163. Zheng X, Chen W, Wang P, Shen D, Chen S, Wang X, Zhang Q, Yang L. Big Data for social transportation. *IEEE Trans Intell Transp Syst*. 2016;17(3):620–30. <https://doi.org/10.1109/TITS.2015.2480157>.
164. Sahay S. Big Data and public health: challenges and opportunities for low and middle income countries. *Commun Assoc Inf Syst*. 2016;39(1):419–38. <https://doi.org/10.17705/1cais.03920>.
165. Sharma N, Namratha B. Towards addressing the challenges of data intensive computing in Big Data analytics. *Int J Control Theor Appl*. 2016;9(23):57–62.
166. Yu S. Big privacy: challenges and opportunities of privacy study in the age of Big Data. *IEEE Access*. 2016;4:2751–63. <https://doi.org/10.1109/ACCESS.2016.2577036>.
167. Chen C-M. Use cases and challenges in telecom Big Data analytics. *APSIPA Trans Signal Inf Process*. 2016. <https://doi.org/10.1017/ATSIP.2016.20>.
168. Anagnostopoulos I, Zeadally S, Exposito E. Handling Big Data: research challenges and future directions. *J Supercomput*. 2016;72(4):1494–516. <https://doi.org/10.1007/s11227-016-1677-z>.
169. Jothi B, Pushpalatha M, Krishnaveni S. Significance and challenges in Big Data: a survey. *Int J Control Theor Appl*. 2016;9(34):235–43.
170. Huang Y, Schuehle J, Porter AL, Youtie J. A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data'. *Scientometrics*. 2015;105(3):2005–22. <https://doi.org/10.1007/s11192-015-1638-y>.
171. Xu Z, Shi Y. Exploring Big Data analysis: fundamental scientific problems. *Ann Data Sci*. 2015;2(4):363–72. <https://doi.org/10.1007/s40745-015-0063-7>.
172. Olshannikova E, Ometov A, Koucheryavy Y, Olsson T. Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *J Big Data*. 2015. <https://doi.org/10.1186/s40537-015-0031-2>.
173. Assunção MD, Calheiros RN, Bianchi S, Netto MAS, Buyya R. Big Data computing and clouds: trends and future directions. *J Parallel Distrib Comput*. 2015;79–80:3–15. <https://doi.org/10.1016/j.jpdc.2014.08.003>.
174. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in Big Data analytics. *J Big Data*. 2015. <https://doi.org/10.1186/s40537-014-0007-7>.
175. Nativi S, Mazzetti P, Santoro M, Papeschi F, Craglia M, Ochiai O. Big Data challenges in building the global earth observation system of systems. *Environ Model Softw*. 2015;68:1–26. <https://doi.org/10.1016/j.envsoft.2015.01.017>.
176. Tian X, Han R, Wang L, Lu G, Zhan J. Latency critical Big Data computing in finance. *J Finance Data Sci*. 2015;1(1):33–41. <https://doi.org/10.1016/j.jfds.2015.07.002>.
177. Perera C, Ranjan R, Wang L, Khan SU, Zomaya AY. Big Data privacy in the internet of things era. *IT Prof*. 2015;17(3):32–9. <https://doi.org/10.1109/MITP.2015.34>.
178. Jin X, Wah BW, Cheng X, Wang Y. Significance and challenges of Big Data research. *Big Data Res*. 2015;2(2):59–64. <https://doi.org/10.1016/j.bdr.2015.01.006>.
179. Mao R, Xu H, Wu W, Li J, Li Y, Lu M. Overcoming the challenge of variety: Big Data abstraction, the next evolution of data management for AAL communication systems. *IEEE Commun Mag*. 2015;53(1):42–7. <https://doi.org/10.1109/MCOM.2015.7010514>.
180. Philip Chen CL, Zhang C-Y. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. *Inf Sci*. 2014;275:314–47. <https://doi.org/10.1016/j.ins.2014.01.015>.
181. Ma Y, Wu H, Wang L, Huang B, Ranjan R, Zomaya A, Jie W. Remote sensing Big Data computing: challenges and opportunities. *Future Gener Computer Syst*. 2015;51:47–60. <https://doi.org/10.1016/j.future.2014.10.029>.

182. Jeong SR, Ghani I. Semantic computing for Big Data: approaches, tools, and emerging directions (2011–2014). *KSII Trans Internet Inf Syst.* 2014;8(6):2022–42. <https://doi.org/10.3837/tiis.2014.06.012>.
183. Sun D, Liu C, Ren D. Prospects, challenges and latest developments in designing a scalable Big Data stream computing system. *Int J Wirel Mob Comput.* 2015;9(2):155–60. <https://doi.org/10.1504/IJWMC.2015.072567>.
184. Dobre C, Xhafa F. Intelligent services for Big Data science. *Future Gener Computer Syst.* 2014;37:267–81. <https://doi.org/10.1016/j.future.2013.07.014>.
185. Qin HF, Li ZH. Research on the method of Big Data analysis. *Inf Technol J.* 2013;12(10):1974–80. <https://doi.org/10.3923/itj.2013.1974.1980>.
186. Ji C, Li Y, Qiu W, Jin Y, Xu Y, Awada U, Li K, Qu W. Big Data processing: big challenges. *J Interconnect Netw.* 2012. <https://doi.org/10.1142/S0219265912500090>.
187. Kambatla K, Kollias G, Kumar V, Grama A. Trends in Big Data analytics. *J Parallel Distrib Comput.* 2014;74(7):2561–73. <https://doi.org/10.1016/j.jpdc.2014.01.003>.
188. Dong XL, Srivastava D. Big Data integration. *Proc VLDB Endow.* 2013;6(11):1188–9. <https://doi.org/10.14778/253622.2536253>.
189. Yin H, Jiang Y, Lin C, Luo Y, Liu Y. Big Data: transforming the design philosophy of future internet. *IEEE Netw.* 2014;28(4):14–9. <https://doi.org/10.1109/MNET.2014.6863126>.
190. Nti IK, Quarcoo JA, Aning J, Fosu GK. A mini-review of machine learning in Big Data analytics: applications, challenges, and prospects. *Big Data Min Anal.* 2022;5(2):81–97. <https://doi.org/10.26599/BDMA.2021.9020028>.
191. Yu Y, Li M, Liu L, Li Y, Wang J. Clinical Big Data and deep learning: applications, challenges, and future outlooks. *Big Data Min Anal.* 2019;2(4):288–305. <https://doi.org/10.26599/BDMA.2019.9020007>.
192. Amalina F, Targio Hashem IA, Azizul ZH, Fong AT, Firdaus A, Imran M, Anuar NB. Blending Big Data analytics: review on challenges and a recent study. *IEEE Access.* 2020;8:3629–45. <https://doi.org/10.1109/ACCESS.2019.2923270>.
193. Chen X-W, Lin X. Big Data deep learning: challenges and perspectives. *IEEE Access.* 2014;2:514–25. <https://doi.org/10.1109/ACCESS.2014.2325029>.
194. Alam A, Ullah I, Lee Y-K. Video Big Data analytics in the cloud: a reference architecture, survey, opportunities, and open research issues. *IEEE Access.* 2020;8:152377–422. <https://doi.org/10.1109/ACCESS.2020.3017135>.
195. Pham Q-V, Nguyen DC, Huynh-The T, Hwang W-J, Pathirana PN. Artificial intelligence (AI) and Big Data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. *IEEE Access.* 2020;8:130820–39. <https://doi.org/10.1109/ACCESS.2020.3009328>.
196. Aydin AA. A comparative perspective on technologies of Big Data value chain. *IEEE Access.* 2023;11:112133–46. <https://doi.org/10.1109/ACCESS.2023.3323160>.
197. Kalantari A, Kamsin A, Kamaruddin HS, Ale Ebrahim N, Gani A, Ebrahimi A, Shamsheerband S. A bibliometric approach to tracking Big Data research trends. *J Big Data.* 2017. <https://doi.org/10.1186/s40537-017-0088-1>.
198. Tsai C-W, Lai C-F, Chao H-C, Vasilakos AV. Big Data analytics: a survey. *J Big Data.* 2015. <https://doi.org/10.1186/s40537-015-0030-3>.
199. Raghupathi W, Raghupathi V. Big Data analytics in healthcare: promise and potential. *Health Inf Sci Syst.* 2014;2(1):3. <https://doi.org/10.1186/2047-2501-2-3>.
200. Ram Mohan Rao P, Murali Krishna S, Siva Kumar AP. Privacy preservation techniques in Big Data analytics: a survey. *J Big Data.* 2018. <https://doi.org/10.1186/s40537-018-0141-8>.
201. Ali A, Qadir J, Rasool RU, Sathiseelan A, Zwitter A, Crowcroft J. Big Data for development: applications and techniques. *Big Data Anal.* 2016. <https://doi.org/10.1186/s41044-016-0002-4>.
202. Hasan MM, Popp J, Oláh J. Current landscape and influence of Big Data on finance. *J Big Data.* 2020. <https://doi.org/10.1186/s40537-020-00291-z>.
203. Seyedan M, Mafakheri F. Predictive Big Data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *J Big Data.* 2020. <https://doi.org/10.1186/s40537-020-00329-2>.
204. Chang V, Muñoz VM, Ramachandran M. Emerging applications of internet of things, Big Data, security, and complexity: special issue on collaboration opportunity for IoTBDs and COMPLEXIS. *Computing.* 2020;102(6):1301–4. <https://doi.org/10.1007/s00607-020-00811-y>.
205. Biswas S, Khare N, Agrawal P, Jain P. Machine learning concepts for correlated Big Data privacy. *J Big Data.* 2021. <https://doi.org/10.1186/s40537-021-00530-x>.
206. Belcastro L, Cantini R, Marozzo F, Orsino A, Talia D, Trunfio P. Programming Big Data analysis: principles and solutions. *J Big Data.* 2022. <https://doi.org/10.1186/s40537-021-00555-2>.
207. Abdalla HB. A brief survey on Big Data: technologies, terminologies and data-intensive applications. *J Big Data.* 2022. <https://doi.org/10.1186/s40537-022-00659-3>.

## Publisher's Note

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