

BIG DATA TRENDS IN DOMAINS OF APPLICATION

Sandra Lovrenčić²⁶
Martina Šestak²⁷
Kornelije Rabuzin²⁸

<https://doi.org/10.31410/itema.2018.71>

Abstract: *Big data is a well-known area of research due to many challenges, which are constantly appearing. Although specifically big data analytics has been prominently in the focus of interest, big data management as a whole should be guided properly, so that as much information and knowledge as possible can be obtained from data for their intended purpose. There are many proposed solutions that are trying to resolve various issues that arose with big data, for example weak processing capability of traditional systems, such as relational databases, to efficiently manage them. New technologies that are trying to cope with various unstructured or structured forms of data are also being researched, Hadoop and NoSQL databases being some of them. Big data challenges are also connected with requirements in the domain of their usage, such as business and commerce, healthcare, public sector, education, or scientific research. This paper examines current research of big data management, especially according to its various domains of application. It gives an insight into applications of big data in various domains of different background (business, government, etc.), followed by a discussion on issues and challenges in its application, as well as mechanisms developed to resolve them. Additionally, this paper discusses the benefits and values of using big data solutions in selected domains. The contribution of the paper is also an overview of big data associated differences and similarities among domains that are most researched in the context of this area of interest.*

Keywords: *Big data, big data management, domain of application, big data analytics, big data challenges*

1. INTRODUCTION

Big data is nowadays a common buzzword with an increasing number of applications in different domains. Big data can be referred to as a large amount of structured, semi-structured and unstructured data generated by modern applications, which requires more time and resources to be processed. Traditional relational database management systems have some difficulties in handling such data (e.g., scalability, storing complex data, fast and distributed data processing, etc.), so some new technologies and paradigms have been introduced in the last decade, for instance NoSQL and NewSQL systems.

Initially, the focus of big data technologies and solutions was meant to focus on managing three properties (dimensions) of data, i.e., 3Vs of data: data volume, velocity and variety. However, the big data concept has been intensively studied and developed over the recent years, which resulted in a number of modern big data solutions able to handle up to 10Vs of data, such as

²⁶ Faculty of organization and informatics, University of Zagreb, Varaždin, Croatia

²⁷ Faculty of organization and informatics, University of Zagreb, Varaždin, Croatia

²⁸ Faculty of organization and informatics, University of Zagreb, Varaždin, Croatia

data veracity (degree of accuracy and trustworthiness), data value (amount of benefits that can be extracted from data), data validity (degree of accuracy and correctness for the intended usage), etc.

Nevertheless, with the number of big data applications to various domains increasing, some challenges and issues related to its application have arisen. This paper gives an overview of those challenges and issues identified throughout the literature review process, as well as how the application of big data helped in dealing with requirements of different application domains. In Section 2, the process of Big Data Management is discussed in general, followed by identifying most frequent challenges of big data application and possible strategies to tackle some of the challenges, as well as the additional value generated by big data application. Section 3 contains an overview of how big data is applied to selected application domains: business and commerce, healthcare, government and public sector, education and scientific research. Throughout the analysis, a special focus has been given to challenges and issues caused by big data application in a given domain, and benefits created by its application in that domain. Finally, the paper is concluded by some final remarks on the degree of similarity and difference of challenges and benefits of big data application in selected application domains.

2. BIG DATA MANAGEMENT VALUES AND CHALLENGES

Wamba et al. [1] studied factors that contribute to improved company performance due to investments related to Big Data Analytics (BDA) owing to its high operational and strategic potential, which is beginning to realize in different industries (e.g., retail, healthcare, manufacturing, etc.). The authors focused on Big Data Analytics Capability (BDAC) as the competence to provide business insights by using data management, technology (infrastructure) and personnel capability, because they consider it to be an important organizational capability leading to a competitive advantage in the modern big data environment. Their research confirmed that BDAC has a significant positive impact on company's performance and process-oriented dynamic capabilities.

Almeida and Calistru [2] identified the following challenges related to Big Data Analysis:

- data privacy;
- access and sharing of information about clients and users;
- size of big data structures;
- analytical challenges, i.e., transparently analyze hidden information behind data.

From the business leaders' and IT practitioners' perspective, it is important to identify data pools, that generate the greatest value for the company. Due to data storage and analysis restrictions, companies need to decide which data is worth collecting and analyzing in the first place. As the size of collected data grows, a set of policy issues needs to be defined to handle data privacy, security and intellectual property as the organizational capability. Finally, to capture value from big data, companies need to deploy technologies and techniques that support innovative types of analysis.

Chen et al. [3] reviewed challenges of big data from data management perspective, and grouped them into the following categories: big data diversity, big data integration and cleaning, big data reduction, big data indexing and querying and big data analysis and mining. Big data diversity deals with different big data sources and types. Also, to obtain structured data, there is a need for unstructured data processing and extraction, which can lead to information loss, but it is not problematic if the initial data is large enough. Challenges related to big data

integration and cleaning are nowadays especially emphasized by an increased usage of user feedback and crowdsourcing in modern systems, which results in an acceptance of imperfect systems and more effort required for entity matching and resolution. To solve big data reduction challenges, companies employ machine learning and massively parallel processing techniques. To resolve big data indexing and querying challenges, nowadays new indexing methods are used, such as distributed B-tree index, as a replacement to inadequate traditional indexing methods developed for small structured data. Big data analysis and mining challenges are caused by more effort required for complex analysis of big data, and there are some approaches and solutions recommended to deal with this challenge, such as using MapReduce framework or data mining software (e.g. Weka).

Sivarajah et al. [4] performed a systematic literature review (SLR) to explore big data challenges and various BDA tools and methods. The authors grouped different big data challenges identified throughout the review process into three categories:

- data challenges – related to data dimensions (volume, velocity, variety and other);
- process challenges – related to challenges, which arise during data processing, analysis and results presentation (e.g. data acquisition); and
- management challenges – related to accessing, managing and governing the data (e.g. privacy).

Additionally, Wamba et al. [5] performed a literature review with the aim to find out the value that big data create and challenges they face. According to their results, most emphasized values are:

- creating transparency;
- enabling experimentation to discover needs, expose variability and improve performance;
- segmenting populations to customize actions;
- replacing or supporting human decision making with automated algorithms;
- innovating new business models, products and services.

On the other hand, some of the challenges they face are data policies, technology and techniques, organizational change and talent, access to data and industry structure.

Big data also represents a challenge for business intelligence systems. Examples of various proposals of new solutions in this field include: implementing data warehouses within NoSQL database management systems [6], parallel construction of OLAP cubes in big data environments [7], data warehousing architecture for big data environment [8], Octopus [9] or Ophidia [10] as BDA engines and Deep Data Warehouse as an integration of warehouse data and unstructured content [11].

3. DOMAINS OF APPLICATION

Each application domain has its characteristics and requirements when it comes to data management. For the purposes of this research, the most challenging and popular application domains have been selected to be analyzed in more detail: business and commerce, healthcare, government and public sector, education and scientific research.

In the following subsections, an overview of characteristics and challenges related to big data management in those application domains will be given, followed by an analysis of most

important challenges and issues related to BDM, as well as opportunities of BDM in those domains.

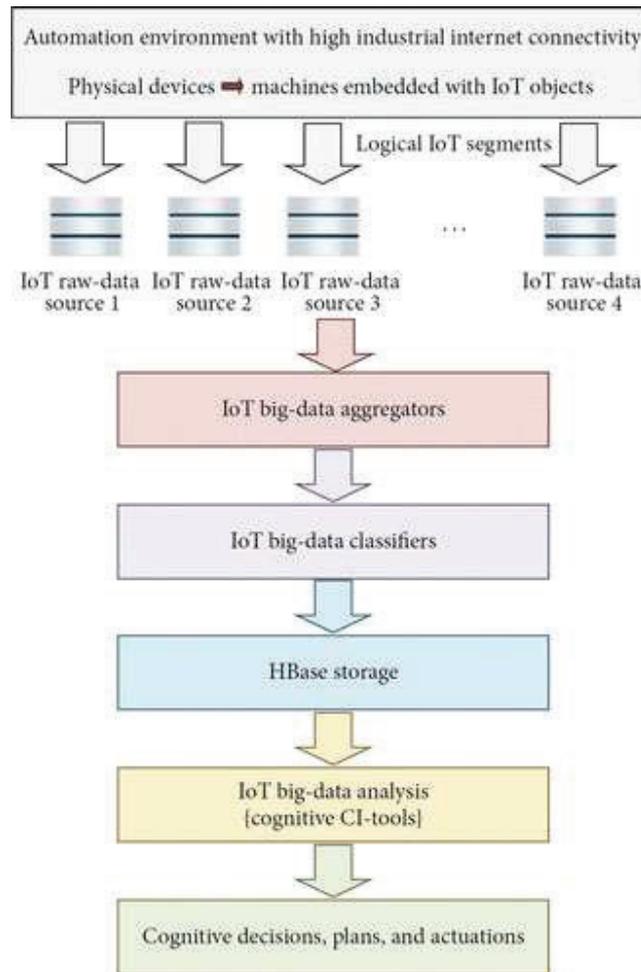
3.1. BIG DATA IN BUSINESS AND COMMERCE

In their paper, Chen and Zhang [12] discuss opportunities introduced by big data in various areas. Specifically, in business, big data creates many opportunities, such as increasing operational efficiency, improving customer service or developing new products and services. Nowadays, many companies use (or are beginning to use) big data to create valuable insights in order to remain competitive on the market. However, as Manyika et al. [13] state, such companies face several challenges, which need to be addressed to take full use of big data: shortage of analytical and managerial personnel, the need to ensure the right infrastructure, and that the economic benefits of big data are clear to users, organization and the economy in general, etc. In their report [13], the authors present several relevant insights gained from their research:

- data has become an important part of every industry and business function;
- big data creates value in several ways (e.g., by replacing/supporting human decision making, innovating new business models, products and services);
- use of big data will become a key basis of competition and growth for organizations;
- there will be a shortage of talent necessary for organizations to take advantage of big data;
- several issues will need to be addressed to capture the full potential of big data (e.g., data policies).

The authors also analyzed the transformative potential of big data in five sectors: healthcare (more effective and cost-saving care initiatives, clinical decision support systems, remote patient monitoring, etc.), public sector (creating transparency, improving operational efficiency, etc.), retail (location-based marketing, customer micro-segmentation, etc.), manufacturing (accelerating product development, product lifecycle management, etc.) and telecommunications (mobile phone location-based services, urban planning, etc.). Manyika et al. analyzed how big data will cause retailers to decrease their prices due to more powerful pricing, promotional and product information in real-time given to customers [13].

Figure 1. COIB-Framework Organization [14]



Mishra, Lin and Chang [14] proposed a Cognitive Oriental Big-data Framework (COIB-Framework) for large scale industrial automation applications that can be used for implementation of Internet of Things (IoT) big data smart industrial applications (Figure 1). Important functions of COIB-Framework are big data aggregation, classification, storage and analysis.

Lee, Kao and Yang [15] discuss the transformation of manufacturing service in big data environment considering Cyber-Physical Systems and IoT in Industry 4.0, where obtaining useful information from big data is important for sustainable innovation. Authors proposed a unified framework for self-aware and self-maintenance machines for efficient extraction of information from big data and better decision making regarding machine performance.

3.2. BIG DATA IN HEALTHCARE

Healthcare datasets consist of data on physiological and other exposure, medical imaging, disease management and similar, which is very disorganized and distributed [16]. The source data is mostly collected from administrative and clinical databases and electronic health record data. Concretely, BDA in healthcare has many potentials, such as the integration of health information to discover new patterns and helping doctors to discover diseases in early stages or accurately predict their trajectory [16]. Moreover, clinical data integration enabled by BDA

results in many benefits, such as improvement of health process efficiency and healthcare quality and reduced healthcare costs.

However, big data in healthcare also faces some challenges [16]:

- the idea of applying big data in healthcare is relatively young, so there is a lack of knowledge and experience in recognizing which data to use and how to use it;
- the lack of appropriate IT infrastructure;
- information processing without human supervision might lead to incorrect interpretations;
- technical challenges, such as integration of different kinds of data from a variety of sources, fragmented data and data inaccuracy;
- patient privacy concerns might arise due to data sharing between stakeholders.

These challenges are usually handled by implementing big data governance, developing an information sharing culture, employing security measures, and training key personnel to use BDA.

In [17], Wang and Hajli performed a study on 63 healthcare organizations to explore the relationship between BDA capabilities and business value. In a later research, Wang, Kung and Byrd grouped BDA derived benefits into five categories [18]:

1. IT infrastructure benefits;
2. operational benefits;
3. organizational benefits;
4. managerial benefits;
5. strategic benefits.

Their results show that operational benefits and IT infrastructure benefits from using BDA appear most often. In this case, operational benefits refer to improved quality and accuracy of clinical decisions, efficient processing of large number of health records, reducing the number of unnecessary treatments, etc., whereas IT infrastructure benefits include reduced system redundancy, avoiding unnecessary IT costs, quick data transfer between healthcare IT systems, etc. BDA is nowadays also used to improve Quality of Service in healthcare systems by gaining valuable insights into data quality [19].

3.3. BIG DATA IN GOVERNMENT AND PUBLIC SECTOR

Al Nuaimi et al. [20] considered big data use in smart cities, which is the concept that many governments are supporting. The use is mostly oriented towards big data storage and analytics with the idea of creation of big data communities that would collaborate and develop solutions for many areas of smart city (e.g. transport). The authors also provide several examples on national, government level: various risks assessments (e.g. terrorist attacks) based on big data, food and water availability based on climate change, government web sites with public data, etc. Authors also name challenges for smart cities, but they are applicable for country government as well:

- data sources and characteristics – various sources and data formats, unstructured data;
- data and information sharing – sharing among different departments (silos);
- data quality – various data sources are a cause of no standard storage formats, lack of structure and consistency, uncertainty and trustworthiness problems;
- security and privacy – confidential government and citizens information;
- cost – expensive implementation and testing of various systems;

- smart city (or country) population – population caused massive amount of data require intelligent applications able to quickly resolve issues that they generate.

According to authors, requirements for smart city design and implementation are [20]:

- big data management – architectures, policies, practices and procedures for management;
- big data processing platforms – capable of analysis with large processing capability;
- smart network infrastructure – connection of components and efficient transfer of collected data;
- advanced algorithms – specially developed for big data;
- open standard technology – flexibility for upgrading, maintaining and adding features;
- security and privacy – mechanisms for security and privacy as an integral part of design and implementation;
- citizen awareness – knowledge and participation in data gathering;
- government role – balance between benefits and privacy and insurance of openness, transparency, participation and collaboration.

Kim, Trimi and Chung [21] state that government datasets are in silos in departments, in various formats and forms, especially unstructured, and must deal with issues of privacy, authority and legitimacy. Many countries already initiated use of large amounts of big data and made them publicly available, but authors' research showed that:

- only small number of them can be named real big data applications, as they mostly share structured data, not real-time unstructured data;
- most of big data applications are in the citizen and business sectors, not government;
- most projects of big data applications in government sector are in early stage of development.

Authors propose several advices for better implementation of big data applications in government sector (e.g. big data project as national priority or establishing global collaboration with other governments).

Barkham, Bokhari and Saiz [22] investigate urban big data impact on real estate markets. The impact is mostly oriented towards lower property taxes, better neighborhood quality, more attractive city etc. As challenges, they name open data use and usability, privacy and technology and human capital constraints.

3.4. BIG DATA IN EDUCATION

In education, big data is used to keep track of students' grades, test scores, etc. Dayal mentions five significant impacts of big data on education [23]:

- it empowers better decision making – requires a know-how to analyze volumes of data and discover useful information;
- it improves students' performance and results – with the help of big data, it is possible to keep track of students' progress to encourage better results;
- it can produce accurate career predictions for students based on their grades and performance;
- collecting students' data can be used for mapping and creating customized study programs or encouraging creative learning;
- it can enhance the learning experience, and digital reading and course materials to make them more simple for students to understand.

West published a report in which he analyzed the potential improvements in research related to big data in education [24]. Specifically, he focuses on applying data mining to discover useful information, which can be used to gain insight into students' performance and learning approaches, data analytics to enable teachers to efficiently prepare for their classes in advance in more detail and nuance, and web dashboards and online tools, which enable the evaluation of students' actions, and how fast they can master the course material.

According to West and other researchers, the value of big data in education is in the ability to collect, store and analyze real-time data, use the collected data for predictive assessments, and visualize data to track performance.

However, big data faces some challenges in the education field as well. Some of the challenges of using big data in education are [25]:

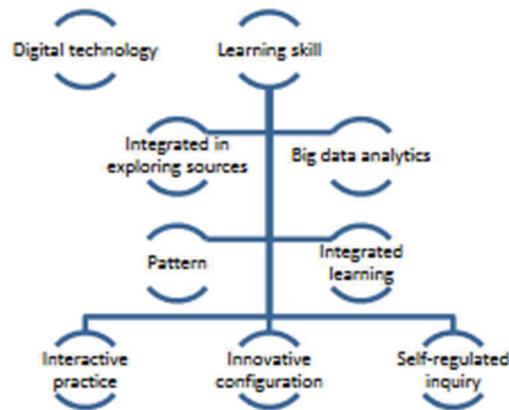
- IT infrastructure costs;
- data privacy and security issues;
- there is information of interest for educators, but difficult to measure in a quantifiable way (e.g., how well are the students learning to apply critical thinking skills);
- the discovered information may point to correlations, but not provide information about the causes (e.g., the information shows that students having classes outdoors are more creative, but that does not mean that studying outdoors automatically makes students more creative).

Poulovasillis [26] emphasizes Learning Analytics (LA) as a prominent area of big data application for improvement of learning process. Another important field is Educational Data Mining, which has the same goal as LA, but is more oriented towards automated discovery and personalization (human oriented). Regarding those two fields, authors argue that great impact can be made on successful use of Intelligent Tutoring Systems and Exploratory Learning Environments, but also state several challenges for big data in education:

- pedagogical – which information is useful to whom and in which learning context;
- technical – practically feasible and pedagogically meaningful methods for big data management process;
- socio-technical – effective use of information extracted from big data from all stakeholders;
- ethical – collection of individual data, privacy, information extraction from big data.

Huda et al. [27] performed a systematic review to find out the use of big data in Innovative Learning Environments (ILE). Authors concluded that it has potential to positively influence learning process mainly in online resources (e.g. personalized learning). They also proposed a big data approach for innovative learning design (Figure 2).

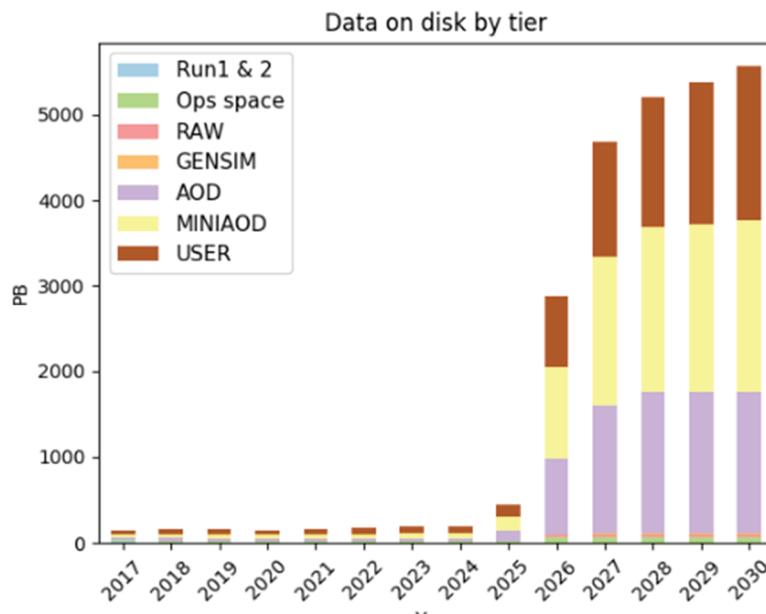
Figure 2. Design of Big Data Approach on Innovative Learning Environment (ILE) [27]



3.5. BIG DATA IN SCIENTIFIC RESEARCH

Nowadays, many scientific fields (astronomy, meteorology etc.) generate large volumes of data, thus making the process of scientific research and discovery data-intensive. Hence, during the system design process in these disciplines, much attention must be given to selecting an appropriate centralized storage and automated analysis approach.

Figure 3. The required estimated disk space for Compact Muon Solenoid (CMS) detector using the current computing model in the next 12 years [29]



Chen and Zhang [12] mention the problem of probing knowledge from data generated by scientific simulations, which is currently still unresolved. Additionally, there are many scientific fields and disciplines that require automated analysis of generated data, as well as a centralized storage repository for all researchers due to its excessive amount, which exceeds average amounts generated in, for instance, everyday Web applications. For example, through the experiments, the Large Hadron Collider (LHC) is able to generate 60 terabytes of data per day or more than 30 petabytes of data per year, and over 100 petabytes of data are archived permanently on tapes [28], whereas telescopes can record up to 30 trillion bytes of image data

per day [12]. According to [29], the total LHC dataset is already almost 1 exabyte large, and around 30 times more data than the LHC currently produces will be collected by LHC detectors in the future (Figure 3).

4. CONCLUSION

Big data paradigm is nowadays applied in many domains of different background (social, technical etc.), where the application of big data brings both some challenges and benefits depending on the characteristics of each domain. The literature review and analysis of big data application in selected application domains in this paper indicates that most domains face similar challenges related to data privacy and security issues, costs generated by implementing big data solutions, difficulties related to finding the most appropriate technology solutions capable to handle requirements of the domain, and the lack of human experts capable of properly interpreting the results of big data analysis process.

A higher degree of difference in big data application between different domains exists when it comes to opportunities and potential benefits arising from its application on solving domain-specific challenges and problems. Also, this research has pointed to a growing importance of Big Data Analytics process for knowledge discovery, which will continue to grow in the years to come with the help of modern technological solutions.

REFERENCES

- [1] S. F. Wamba, A. Gunasekaran, S. Akter, S. J. Ren, R. Dubey, and S. J. Childe, "Big data analytics and firm performance: Effects of dynamic capabilities," *J. Bus. Res.*, vol. 70, pp. 356–365, 2017.
- [2] F. Almeida and C. Calistru, "The main challenges and issues of big data management," *Int. J. Res. Stud. Comput.*, vol. 2, no. 1, pp. 11–20, 2013.
- [3] J. Chen *et al.*, "Big data challenge: a data management perspective," *Front. Comput. Sci.*, vol. 7, no. 2, pp. 157–164, 2013.
- [4] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, 2017.
- [5] S. F. Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study," *Int. J. Prod. Econ.*, vol. 165, pp. 234–246, 2015.
- [6] R. Yangui, A. Nabli, and F. Gargouri, "Automatic transformation of data warehouse schema to NoSQL data base: comparative study," *Procedia Comput. Sci.*, vol. 96, pp. 255–264, 2016.
- [7] A. Cuzzocrea, R. Moussa, and G. Xu, "OLAP*: effectively and efficiently supporting parallel OLAP over big data," in *International Conference on Model and Data Engineering*, 2013, pp. 38–49.
- [8] Q. Yang and M. Helfert, "Revisiting arguments for a three layered data warehousing architecture in the context of the Hadoop platform," 2016.
- [9] C. Xu, Y. Chen, Q. Liu, W. Rao, H. Min, and G. Su, "A unified computation engine for big data analytics," in *Big Data Computing (BDC), 2015 IEEE/ACM 2nd International Symposium on*, 2015, pp. 73–77.
- [10] S. Fiore, A. D'Anca, C. Palazzo, I. Foster, D. N. Williams, and G. Aloisio, "Ophidia: Toward big data analytics for escience," *Procedia Comput. Sci.*, vol. 18, pp. 2376–2385, 2013.
- [11] C. Gröger, H. Schwarz, and B. Mitschang, "The deep data warehouse: link-based

- integration and enrichment of warehouse data and unstructured content,” in *Enterprise Distributed Object Computing Conference (EDOC), 2014 IEEE 18th International*, 2014, pp. 210–217.
- [12] C. L. P. Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” *Inf. Sci. (Ny)*, vol. 275, pp. 314–347, 2014.
- [13] J. Manyika *et al.*, “Big data: The next frontier for innovation, competition, and productivity,” 2011.
- [14] N. Mishra, C.-C. Lin, and H.-T. Chang, “A cognitive adopted framework for IoT big-data management and knowledge discovery prospective,” *Int. J. Distrib. Sens. Networks*, vol. 11, no. 10, p. 718390, 2015.
- [15] J. Lee, H.-A. Kao, and S. Yang, “Service innovation and smart analytics for industry 4.0 and big data environment,” *Procedia Cirp*, vol. 16, pp. 3–8, 2014.
- [16] N. Mehta and A. Pandit, “Concurrence of big data analytics and healthcare: A systematic review,” *Int. J. Med. Inform.*, vol. 114, pp. 57–65, 2018.
- [17] Y. Wang and N. Hajli, “Exploring the path to big data analytics success in healthcare,” *J. Bus. Res.*, vol. 70, pp. 287–299, 2017.
- [18] Y. Wang, L. Kung, and T. A. Byrd, “Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations,” *Technol. Forecast. Soc. Change*, vol. 126, pp. 3–13, 2018.
- [19] F. A. Batarseh and E. A. Latif, “Assessing the quality of service using Big Data analytics: With application to healthcare,” *Big Data Res.*, vol. 4, pp. 13–24, 2016.
- [20] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, “Applications of big data to smart cities,” *J. Internet Serv. Appl.*, vol. 6, no. 1, p. 25, 2015.
- [21] G.-H. Kim, S. Trimi, and J.-H. Chung, “Big-data applications in the government sector,” *Commun. ACM*, vol. 57, no. 3, pp. 78–85, 2014.
- [22] R. Barkham, S. Bokhari, and A. Saiz, “Urban Big Data: City Management and Real Estate Markets,” *GovLab Digest: New York, NY, USA*, 2018.
- [23] P. Dayal, “5 Dramatic Impacts of Big Data on Education.” .
- [24] D. M. West, “Big data for education: Data mining, data analytics, and web dashboards,” *Gov. Stud. Brookings*, vol. 4, no. 1, 2012.
- [25] S. Smith, “Challenges of Big Data in Education.” .
- [26] A. Poulouvasilis, “Big Data and education,” 2016.
- [27] M. Huda, Z. Haron, M. N. Ripin, A. Hehsan, and A. B. C. Yaacob, “Exploring innovative learning environment (ILE): big data era,” *Int. J. Appl. Eng. Res.*, vol. 12, no. 17, pp. 6678–6685, 2017.
- [28] CERN, “Facts & Figures,” 2018.
- [29] G. Stewart, “HEP Software Foundation Community White Paper looks forward to the HL-LHC.”