



Review

A Survey on Big Data in Pharmacology, Toxicology and Pharmaceutics

Krithika Latha Bhaskaran ^{1,*}, Richard Sakyi Osei ², Evans Kotei ³, Eric Yaw Agbezuge ³, Carlos Ankora ⁴ and Ernest D. Ganaa ²

¹ School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India

² Information and Communication Technology Department, Faculty of Applied Science and Technology, Dr. Hilla Limann Technical University, Wa P.O. Box 553, Ghana

³ Computer Science Department, Faculty of Applied Science, Kumasi Campus, Kumasi Technical University, Kumasi 00233, Ghana

⁴ Department of Computer Science, Faculty of Applied Sciences and Technology, Ho Technical University, Ho P.O. Box HP 217, Ghana

* Correspondence: krithika.lb@vit.ac.in

Abstract: Patients, hospitals, sensors, researchers, providers, phones, and healthcare organisations are producing enormous amounts of data in both the healthcare and drug detection sectors. The real challenge in these sectors is to find, investigate, manage, and collect information from patients in order to make their lives easier and healthier, not only in terms of formulating new therapies and understanding diseases, but also to predict the results at earlier stages and make effective decisions. The volumes of data available in the fields of pharmacology, toxicology, and pharmaceutics are constantly increasing. These increases are driven by advances in technology, which allow for the analysis of ever-larger data sets. Big Data (BD) has the potential to transform drug development and safety testing by providing new insights into the effects of drugs on human health. However, harnessing this potential involves several challenges, including the need for specialised skills and infrastructure. In this survey, we explore how BD approaches are currently being used in the pharmacology, toxicology, and pharmaceutics fields; in particular, we highlight how researchers have applied BD in pharmacology, toxicology, and pharmaceutics to address various challenges and establish solutions. A comparative analysis helps to trace the implementation of big data in the fields of pharmacology, toxicology, and pharmaceutics. Certain relevant limitations and directions for future research are emphasised. The pharmacology, toxicology, and pharmaceutics fields are still at an early stage of BD adoption, and there are many research challenges to be overcome, in order to effectively employ BD to address specific issues.

Keywords: big data; drug development; healthcare pharmacology; pharmaceutics; toxicology



Citation: Latha Bhaskaran, K.; Osei, R.S.; Kotei, E.; Agbezuge, E.Y.; Ankora, C.; Ganaa, E.D. A Survey on Big Data in Pharmacology, Toxicology and Pharmaceutics. *Big Data Cogn. Comput.* **2022**, *6*, 161. <https://doi.org/10.3390/bdcc6040161>

Academic Editors: Domenico Talia and Fabrizio Marozzo

Received: 13 October 2022

Accepted: 2 December 2022

Published: 19 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The field of research involving analysis of the functioning of drugs in the human body is referred to as pharmacology. It details a class of medical science intended to address the employment of drugs in preventing, analysing, and treating diseases. As such, pharmacologists research the chemical properties of drugs and their impacts on the human body, in order to develop novel drugs and test them for efficacy and safety. Toxicology is the study of poisons, and relevant research helps to identify the effects of toxins and toxicants on the human body. It is a branch of pharmacology that aids in addressing the characterisation, identification, and quantification of the biological agents and chemicals that are responsible for creating adverse effects of the chemical agents on the human body. They develop novel methods to test for toxicity, and identify and quantify the toxicants in provided samples [1,2]

Pharmaceutics is a branch of pharmacy that manages the process of turning new chemical agents or old drugs into medications that can be used safely and efficiently by patients. It is also referred to as the science of dosage, which informs the design. Many chemicals with pharmacological properties require special measures to assist in attaining therapeutically related amounts at their sites of action. Pharmaceutics also helps in relating the formulations of drugs to their disposition and delivery in the body. Pharmacology, pharmaceutics, and toxicology are all categorised as interdisciplinary sciences [3,4]. They involve the extraction of knowledge from several disciplines, such as chemistry, biology, mathematics, and physics. BD has been one of the most researched topics in recent decades. BD denotes large or complex data sets that render the use of existing data-processing methods inadequate. Challenges associated with BD analysis include data storage, data capturing, sharing, searching, visualising, transferring, information privacy, and data source updating.

BD in health and medical care has attracted the attention of researchers owing to its various benefits. BD-based solutions for clinical decision support systems have revealed promising results in treating diseases such as Alzheimer's disease [5]. BD approaches have also been used for optimising electronic medical records [6]. The data generated in pharmacology, toxicology, and pharmaceutics include outcomes regarding the clinical behaviours, pharmacokinetics, chemical properties, associated toxicity of drugs. Data from cell culture experiments and animal studies are also included in this category. Thus, the volume of data generated in pharmaceutics, pharmacology, and toxicology has been increasing rapidly. This is primarily due to the employment of next-generation sequencing (NGS) and high-throughput screening (HTS). HTS is a technology that enables the simultaneous testing of millions of chemical elements. Additionally, NGS enables the efficient sequencing of RNA or DNA molecules.

It has been estimated that the global data volume generated by the toxicology, pharmaceutics, and pharmacology fields is approximately 2.5 zettabytes, where one zettabyte is equal to one sextillion bytes. BD is generally useful in pharmacology, pharmaceutics, and toxicology, as it can help to improve the prediction rate and accuracy of the effects of the researched drugs. It has helped to enhance the accuracy and speed of drug discovery and development. Although there are certain associated advantages of using BD in pharmacology, toxicology, and pharmaceutics, there are also certain drawbacks. BD can be challenging and overwhelming to navigate, and it may be difficult to trace useful data from the vast database. In addition, the implementation of BD is generally expensive, due to the overheads related to data collation, transformation, and analyses/modelling. The challenges and the overwhelming nature of BD can be resolved through the use of a data visualisation tool. Such an approach allows users to see the data and filter the necessary data, making it easier to understand, navigate, and extract the specific necessary information. Figure 1 shows a generic workflow representation of BD in the Hospital and Pharmacology ecosystem, based on the studies [7–9].

Hospital and clinics are the touchpoints for patients, that provide the symptom analysis, pre-diagnosis, physician advice, case history, and preliminary investigations from both the patient and physician perspectives. The data repository is the data store that collates data from all entities in the ecosystem, which are stored for further analytics, and as data sets used for prediction. In the scenarios presented in Figure 1, hospitals, clinics, and pharmacies provide the data that will eventually be stored in the big data repository. The repository possesses a big data nature, due to the data velocity and the heterogeneity of the data that are collated. The data from the BD repository are used for analytics such as identifying illness patterns, trends, and pre-emptive prediction of pandemics. Furthermore, BD analytics can be used to develop predictive models, which may be used to identify high-risk patients. Predictive model and pattern analytics can be used to model and develop targeted interventions. The use of a monitoring system coupled with BD models can form a proactive monitoring platform that can provide feedback to physicians. Based

on Figures 1 and 2, a graph-based representation of Scenario 1—Drug efficacy monitoring system using BD—was derived.

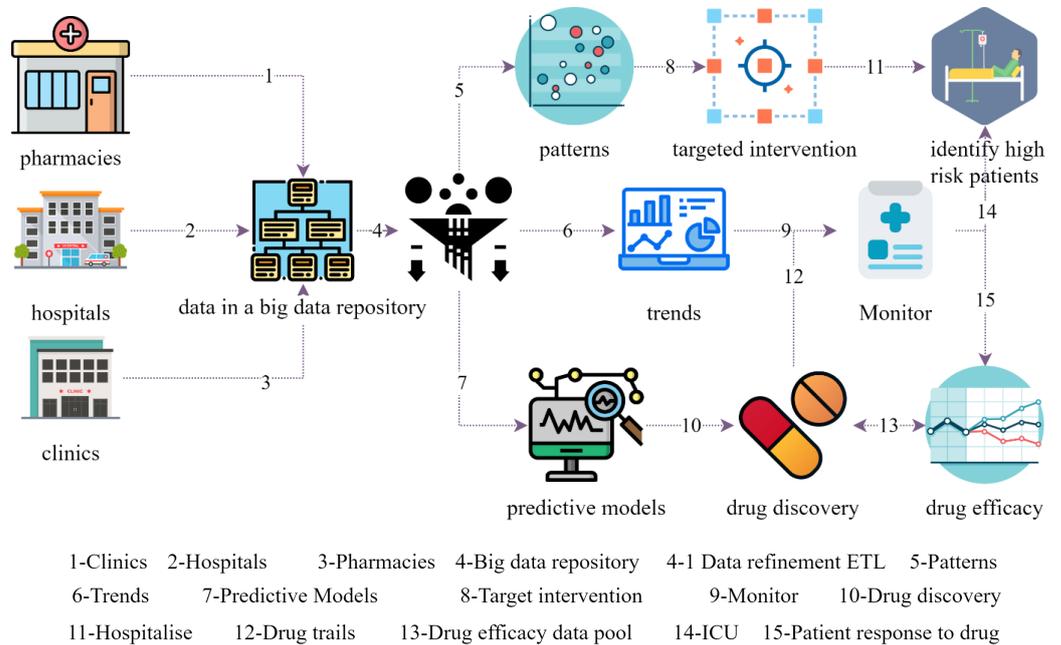


Figure 1. Generic representational workflow of big data in the Hospital and Pharmacology ecosystem.

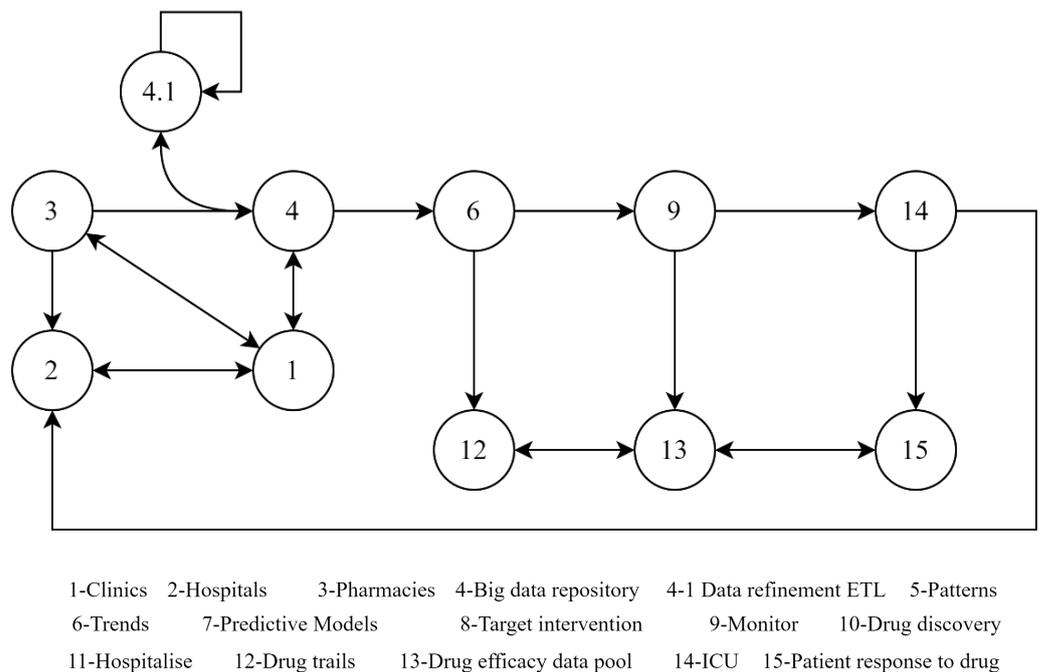


Figure 2. Flowchart for drug efficacy monitoring system using big data.

Scenario 1 considers a dataflow graph path for the drug efficacy monitoring system, one of the applications of BD in the healthcare domain. 1. The patient is attended at the clinic for pre-investigation; 2. the patients referred for specialist consultation are received at the hospital; 3. the required prescribed drug is collected from the pharmacy; 4. data of all patient activities are collected in the BD repository; 4.1. data cleaning and transformation are carried out as a continuous process; 6. the trend of the patient is analysed; 9. the patient is monitored in real-time during the drug course; 14. check whether hospitalisation

is required; 15. drug efficacy is improved based on all the data that is collected; 13. an efficacy report is shared with the drug discovery team, in order to improve the drugs; 14. the patient is recoursed with an improved drug; and 2. feedback is sent to the physician for recommendation of a new prescription and dosage. This is one example scenario of the use of BD in the healthcare and pharmaceutical ecosystem. BD has extensive uses, and its potential has still not been fully discovered and practically implemented in real scenarios.

Another scenario based on Figure 1 is shown in Figure 3, which depicts the representation of Scenario 2: Drug discovery for a pandemic situation using BD.

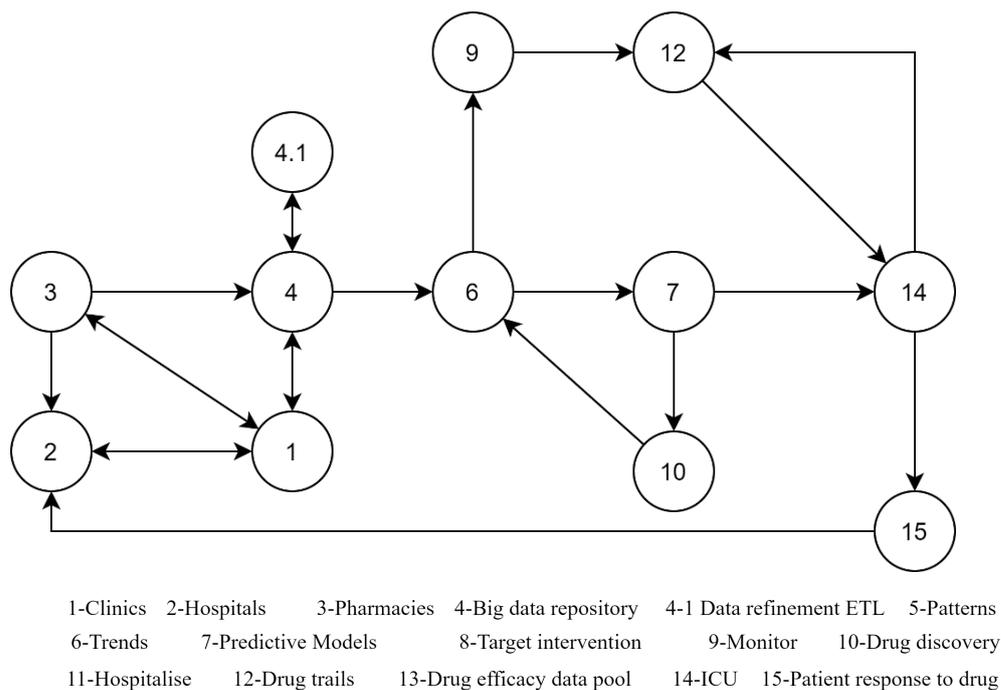


Figure 3. Flowchart for drug discovery in pandemic situation using big data.

Scenario 2—Drug discovery for a pandemic situation—uses BD, following the path 3–2–4–6–7–10–6–9–12–14–15–2. First, when a sudden surge of patients with unknown illness is reported at (3.) clinics and (2.) hospitals, (4.) a new set of big data is rapidly created for diagnosis. Prescriptions and pathological reports are generated, (6.) the trends are matched with existing pool to find similar of any historic pandemic; and (7.) the data are quickly modelled to present a pandemic containment prediction predicated on the speed for the contagion vector, as well as its time to mutate, time to outgrow the available medical facilities, and similar data useful for prediction of pandemic containment. Then, the data from 9, 12, and 10 are used recursively used for prediction and drug discovery and provided to (14.) and monitored at (15.) These scenarios represent the application of BD in healthcare and related domains. As data plays the dominant role in the era of data analytics, BD has great scope in health science and related fields.

Through extensive research, it was confirmed that very few studies have attempted to review existing works to analyse the implementation of BD in the pharmacology, toxicology, and pharmaceuticals fields. Thus, the present study is expected to greatly contribute to upcoming research, by providing clear knowledge regarding the implementation of BD in the field of drug development, and will also encourage the development of different big data-based models to serve the medical community. Therefore, the main goal of the present study is to review the literature regarding the employment of BD in pharmaceuticals, pharmacology, and toxicology.

1.1. Objectives of the Study

The present study has the following objectives:

- To review the implementation of BD in the toxicology, pharmacology, and pharmaceuticals fields;
- To offer a comparative analysis regarding the implementation of BD in pharmacology, toxicology, and pharmaceuticals;
- To provide directions for future researchers and end-users to facilitate the usage of BD.

1.2. Paper Organization

The introduction section describes the importance of the topic, and the remainder of the article is organised as follows: Section 2 investigates the literature related to the work. Section 3 discusses the benefits of BD. Sections 4 and 5 present the survey analyses. Section 6 presents the limitations of the paper. Finally, Section 7 concludes our survey.

1.3. Definitions and Concepts of BD

BD can be defined as a huge and complex database that stores data that are heterogeneous and featuring multiple scales of information retrieved from many sources. BD can be related to healthcare in terms of electronic healthcare reports, clinical trial data, and administrative claims. The processing of BD includes feeding the data to the system, data preservation for storage, data analysis, and visualization of outcomes. Grouping the BD and smaller machine resource, offers many advantages, such as high availability, pooling of resources, and scalability [10]. As healthcare information is becoming digitalized, there has been widespread development in healthcare BD, and value-based care has motivated the healthcare sector to use BD analytics to make strategic professional decisions. The healthcare sector is handling many challenges based on the variety, volume, and veracity of data in this sector. Thus, BD plays a huge role by motivating healthcare innovation, influenced by many financial systems, considering factors such as the requirements of patients, motivations of the developers, and technical development. It has been observed that BD can offer optimal treatment decisions for patients and healthcare providers, based on population statistics [11].

1.3.1. Why BD?

Large volumes of data are generally available in two formats: Structured or unstructured. Structured data can be generated by machines and humans using a particular model or schema, and are usually accumulated in a database. Structured data are ordered around outlines with the data types that are clearly defined. Certain examples that characterise structured data include data, time, numbers, and strings stored in the database's columns. On the other hand, unstructured data do not possess any pre-defined model or schema. Associated examples of unstructured data are log files, mobile data, social media posts, text files, and other media. These media do not have any pre-defined schema set, so they are categorised as unstructured data [12].

The amount of data generated by popular corporations, small-scale industries, and scientific projects has been growing at an extraordinary level. These high volumes of data produced present incredible processing, storage, and analytical challenges that must be carefully dealt with and considered. Furthermore, traditional relational database management systems (RDBMSs) and the associated data processing tools are inadequate in dealing with huge data effectively, where the data size is typically measured in petabytes or terabytes. These existing tools lack the ability to deal with a large amount of data effectively when the size is enormous. Fortunately, paradigms and BD tools such as Hadoop and MapReduce are available to solve these BD challenges.

Certain features of BD are explained in the following.

Value

This allows for the extraction of beneficial information from large data sets [13]. The mechanism is designed based on how the data are added to create the knowledge. It has been stated that having access to the data is good, but it may be useless if it cannot be

converted into values. Thus, it is the most important feature of BD that many institutions and enterprises invest the most into, in order to generate income and knowledge [14].

Variability

This particular characteristic represents the consistency of the data over time [15].

Variety

This represents the kinds of information associated with structured, unstructured, or semi-structured data. Key values include web clicks, relational tables, articles, e-mail messages, streamed audio and videos, and other media. This information may be gathered through sensors, smartphones, and social networks. In the healthcare sector, the information is gathered from unstructured and semi-structured sources, such as clinical trial outcomes, electronic health records, and physiological and chemical sensors [16].

Volume

This refers to the amount of data available from the various related resources. The data volume in BD typically starts at the Terabyte level. An example of this characteristic associated with the biomedical field is the proteomics database generated by the University of Munich, which has collected a data volume of approximately 5.17 TB, possessing 92% of unknown genes. In medical imaging, the visible human project has archived 5189 female anatomical images and 1871 male cryosections [16].

Velocity

This is the speed of data processing, transfer, generation, and collection. The content of the data is constantly changing with the addition of new data and various forms of streamed data from many sources at varying speed levels. Sequencing technologies have recently allowed for the production of billions of DNA sequences at low cost; this information is stored on desktop computers and shared within institutions worldwide.

Veracity

This denotes the accuracy and reliability of the information, the correctness of the data, quality, and data governance. This particular attribute is completely dependent on the data source.

1.4. Growth of BD

There has recently been a huge shift in the volume and speed of the data, beyond the comprehension of human minds. In 2013, the total volume of data in the world was estimated at 4.4 Zettabytes. This volume experienced an enormous increase, up to 44 Zettabytes by 2020. Although there has been a steady rise in technology, at present, it is not easy to analyse such enormous data. The demand for analysis of large data sets has paved the way for the rise of BD over the past decade. Data analytics, data analysis, and BD originate from the long-standing database management domain, which is completely dependent on the extraction, storage, and optimisation methods that are usually used for data stored in RDBMSs. Since the early 20th century, the internet has offered unique data analysis and data collection opportunities [17]. With the expansion of web traffic and online stores, companies such as Amazon, Yahoo, and eBay began analysing customer behaviours by investigating their clicking rates and tracking customer locations through their IP addresses, revealing a new world of possibilities. In addition, HTTP-based web traffic has increased the volume of unstructured and semi-structured data. To analyse these data, organisations require new approaches and solutions for storage issues, in order to investigate these new data efficiently.

1.5. Development of BD in the Medical Sector

There has been a constant increase in the demand for solutions regarding efficient analytical tools. This trend has also been noticed with regard to analysing large volumes of data. Organisations and institutions are searching for approaches to make use of the power of BD to enhance their competitive advantage, decision making, or business performance. BD provides potential solutions for private and public organisations; however, regarding outcomes, the practical employment of BD in various kinds of organisations requires domain-level adoption and re-structuring. Specifically, the healthcare industry has started shifting from a disease-centred to a patient-centred model, which is applicable in value-based healthcare delivery systems. To meet the demand and provide efficient patient-centred care, it is important to address and investigate a large amount of data from the healthcare sector. Many issues arise when healthcare data are considered. Healthcare has always produced large amounts of data. In addition, the introduction of electronic medical records and the large amount of data collected by different sensors or the data that the patients generate through social media has created many data streams. The appropriate use of such data can enable healthcare organisations to support clinical decision-making, public health management, and disease surveillance [18].

Classification provides a significant approach for bringing intelligence to medical data. Due to the simplicity of the k NN classification algorithm, it has been widely employed in several sectors. However, when the sample size is large and the features of the attributes are large, the effectiveness of the k NN algorithm will be reduced. A study [19] has proposed a novel k NN algorithm and compared it with the other existing k NN algorithms. In particular, the classification was made in the query instance neighbourhood of the existing k NN classifiers, and weights were allocated to each class. The recommended algorithm considered the class distribution around the query instance, in order to ensure that the assigned weights do not impact the outliers. The results of the considered study revealed that the recommended algorithm could efficiently enhance the effectiveness of the classification of the k NN algorithm when processing large data sets while maintaining the classification accuracy of the k NN algorithm, as well as providing better performance in terms of classification. However, the considered study only researched single-class classification while, in terms of application, multi-class classification is more popular and necessary. Additionally, healthcare data typically have a high missing rate, where these missing fields have been shown to greatly impact the classification results in existing works.

Similarly, another study [20] has developed a BD analytics-enabled transformation system based on the practice-based view. This revealed the causal relationships among BD capabilities, benefit dimensions, IT-enabled business values, and transformation practices. This model was then validated in a medical setting, offering a strategic view of BD analysis. Three vital paths for value chains were detected for medical organisations, through implementing a model that offers practical insights for managers. This study revealed the important elements and links for understanding the transformation of BD. One major limitation of the considered study was the data source. Additionally, better validation could have been performed by collecting and investigating primary data.

To date, the healthcare sector has not completely utilised the potential of BD. While the constantly developing academic research on the concepts of BD analytics has been technically oriented, there is an increasing demand for understanding the strategic implications of BD. Intending to address this lack, the study [21] has attempted to investigate the historical development, component functionalities, and architectural design of BD analytics. They identified 5 BD analytical capabilities from 26 BD implementations, including unstructured data, analytical capability for pattern, decision support capability, traceability, and predictive capability. The main limitation of the considered study was that IT adoption usually lags, when compared with other sectors, which is one of the main reasons why such cases are difficult to find. Although many cases have been found from various sources, the majority of cases were detected from vendors.

Wearable medical tools with sensors continuously generate a large amount of data, which can be considered as BD, in the form of unstructured and semi-structured data. Due to the complexity of the data, it is not easy to investigate valuable information that could help in decision making. Alternately, data security is another major requirement of BD in the healthcare sector. To address this issue, traditional research [22] has attempted to recommend novel architectures for implementing IoT, in order to accumulate and process scalable sensor data for healthcare applications. The recommended architecture consists of two main frameworks: Grouping and choosing (MC) and metafog redirection (MF-R) frameworks. MF-R frameworks employ BD technologies such as Apache HBase and Apache Pig to collect and store the sensor data produced from various sensor devices.

On the other hand, various security frameworks have been studied in the attempt to build models that combine multi-variate and non-stationary data. The obtained models utilize a log-normal distribution for the margins with linear trends and peak series [23].

2. Applications of BD

This section briefly reviews the body of related work that is available and indexed by reliable databases such as SCOPUS and WoS. The keywords used were under the subject categories of 'BD', 'pharmacology', 'toxicology' and 'pharmaceutics'.

The employment of BD for safety management in various areas, such as traffic safety [24], public safety [25], food safety [26], and patient safety [27], has recently been extensively studied. In addition, the influence of BD on drug discovery and design has been explored, in terms of future developments of medicine [6]. The core points in the discussion were the challenges that arise while implementing BD technologies, preserving the quality and privacy of data sets, and how the industry should adapt to welcome the BD era. It was concluded that, while BD has a significant impact on the advancement of pharmaceutical science, there are still many challenges to overcome.

The perspective of BD analytics in adapted medicine, focusing on how it could improve patient care, has been discussed in [25]. It was emphasised that the advancements in information technology have made this possible, but challenges remain to be addressed. BD analytics provide the potential to improve patient care, but more research is required to make this a reality.

The author in [26] has noted that current in vitro toxicity data could be used to develop models and tools to help in chemical toxicity research. The core points of the discussion were that the data are rich in information that can be used to evaluate complex bioactivities, and that a BD approach is necessary for relevant processing. It was found that the data are valuable for chemical toxicity research, but more tools need to be developed to help researchers use it. The pharmaceutical industry is facing a challenge in terms of productivity, in light of which BD initiatives may provide the insights needed to turn the industry around [28].

BD and translational medicine have evolved, and disruptive technology is bringing them together. The evolution of BD and translational medicine has been discussed, as well as the hindrances in applying BD techniques to translational medicine and the future of translational medicine. The author concluded that the future of translational medicine is bright and that the "Complete Health Record" concept will revolutionise the way in which translational medicine is practised [29].

BD is essential in safety sciences [27], and can be used to find similar substances and clusters of properties. Moreover, the need for safety BD [30] has been rapidly growing with constant development, and integration with science and technology has added more life to safety science research [31].

The author in [27] has sought to better understand the interactions between BD and Dynamic Simulation Modelling (DSM), as well as how incorporating them could be useful to healthcare decision-makers. The core points in the discussion were the benefits of BD and DSM, and how they can be used together to improve healthcare delivery. Integrated

BD and DSM offer complementary value in healthcare, in terms of addressing complex, systemic health economics and outcomes questions.

2.1. *BD in Toxicology*

The rate of data generation associated with toxicology continues to multiply, and the volume of data that is generated has been growing drastically. This is due to advancements in software solutions and the chemical-informatics method, which increase the accessibility of open resources such as biological, chemical, and toxicology data. Thus, the significant necessity for BD analytics to store and access the data associated with the toxicology domain has surged. Concerning this aspect, a conventional study [32] has proposed a machine learning method for raw HRMS-DIA (High-resolution mass spectrometry-Data independent acquisition) data. They evaluated the machine learning model by training, validating, and testing on sets of solvents and blood samples containing drugs considered to be usual in forensic toxicology, with the aim of categorical prediction using a feed-forward neural network framework. With the application of the employed machine learning approach, the specificity and sensitivity of the validation process and the test set for the prediction sample classes were observed to be in a suitable range for routine use in the laboratory. The study clearly emphasised the efficacy of employment of BD along with machine learning algorithms.

Probabilistic topic modelling has been used to analyse large-scale genomic data to uncover hidden patterns; in particular, this method was used to analyse a toxic genomic data set, and it was found that patterns related to the impact of doses and time points of treatment could be identified. The authors concluded that this method can reduce animal use in research [33].

A better understanding of how BD helps to delineate personalised approaches in severe mental illness and the provision of a quantitative synthesis of BD approaches for metabolomics in severe mental illness is necessary [34]. Notably, BD has the potential to improve our understanding of the developmental trajectories of mental disorders.

The considered existing research has used broad data utilized in clinical studies conducted from the perspectives of neurology, tumours, cardiovascular disease, psychiatric diseases, and other implementations [35]. Traditional research has emphasized the advantages of BD, in that it enables the study of diseases at the genetic level, thus offering more valuable treatments than traditional or usual treatments, as well as providing the ability to discover the evolution trajectory of humans. BD has an optimistic impact on medical studies, and its growth continues.

2.2. *BD in Pharmacology*

The implementation of BD in precision medicine has been welcomed. Pharmacogenomics—that is, the study of the effect of genes on a person's reaction to certain drugs—is within the realm of precision medicine. This new area combines pharmacology and genomics to improve valid and safe drugs and doses that respond to variations in individual genes. Precision medicine has a relatively limited role in daily care; however, researchers expect that this approach will encompass many healthcare sectors in the coming years. In addition, BD has the potential to facilitate personalised precision medicine [27].

Various analysis techniques [29] and tools are being implemented for genetic/genomic discovery in pharmacogenomics. However, the BD-related issues faced by pharmacogenomics need to be addressed, in order to maximise the potential in the field. Compared with applications in IT fields, such as social network analysis, the data sets used for drug discovery research are relatively small. However, with the development of combinatorial chemistry synthesis, HTS techniques, and genomics/genetics knowledge, the databases for drugs and drug candidates are growing rapidly.

New modelling approaches are needed to handle these larger data sets [36]. The existing research [37] has attempted to investigate the feasibility of BD analysis on 3290 approved drugs and formulations, for which 1,637,499 adverse events have been recorded in both

human and animal species for approximately 70 years. A BD technique was utilized in this study, which is known to be a powerful analytic approach. However, it was revealed that the principle feasibility of a combined text mining and statistical method also led to numerous pitfalls, such as inadequate arrangement of pre-clinical ontologies and insufficiency of controlled vocabulary.

A conventional study [27] has attempted to identify the factors associated with the success of hypertension drug treatment, using BD approaches along with machine learning methods. As a result, it was disclosed that proton-pump inhibitors (PPIs) and hydroxymethylglutaryl coenzyme (HMGCoA) reductase inhibitors could significantly enhance the success rate of hypertension. In addition, new machine learning methodologies with BD have helped in identifying the prominent anti-hypertension therapy by re-generating medications available for new symptoms.

In a previous study [38], the author has attempted to determine standard methods that would likely help to increase the usability of (publicly available or privately produced) biological data. It was identified that data integrity is significant during pre-clinical drug development, and that investigators should use consistent methods to exploit the functions of privately and publicly created biological data. The author also emphasised that the increasing interest in and the interpretation of cross-platform approaches is significant.

BD can be used in paediatric drug [31] development. The use of BD for clinical trial design, efficiency, and safety of data has been attained in clinical trials. Therefore, exploring the current opportunities and challenges of BD in future paediatric drug development must be enriched. Although BD has the potential to play a significant role in paediatric drug development, and there are still many challenges that need to be addressed.

BD can be used in drug research to determine efficacy and safety signals [39]. The steps involve data acquisition, extraction, aggregation, analysis, modelling, and interpretation. BD can leverage and improve clinical decisions at the point of care, uncovering or validating drug efficacy and safety.

The steps of pharmacogenomics studies, [38] has considered data collection for interpretation and highlighted the bioinformatics aspects that can pose problems. The major challenges of data processing and analysis can lead to inaccurate results. Therefore, paying careful attention to these steps is important, in order to avoid mistakes and produce accurate pharmacogenomics studies.

The author in [40] have discussed the discovery of novel bromodomain BRD4 binders. It was inferred that public databases are useful for predictive model building, and that machine learning can allow for the extraction of real knowledge, despite the noise present in structure-activity data. Therefore, public databases are key assets in drug discovery, and machine learning plays a significant role in mining real data.

BD has changed the field of drug development [41]. Novel methods for therapeutic drug discovery, inference of clinical toxicity, candidate drug prioritisation, and machine learning techniques for drug discovery are becoming familiar. Experts from various platforms should conduct closer collaborations to translate the analysis results for treatment and prognosis in medical practice [42].

BD is significant for medical use and requires re-thinking, regarding the data storage infrastructure, the analysis growth, and the associated tools to drive advancements in the considered field [43]. In addition, BD is undoubtedly important for clinical practice, and physicians are responsible for developing and using BD to enhance patient care.

Machine learning has been utilized to predict psychiatric outcomes [44] in humans, where these techniques are more powerful than traditional statistical approaches. The author also discussed ways to optimise machine-learning techniques in the context of psychiatric research. BD has transformed natural product research and helped researchers both ask and answer new questions [45]. The author also highlighted the limitations regarding our current engagement with large data sets.

2.3. BD in Pharmaceuticals

Clinical behaviours are important in pharmaceuticals and life science, as they can be employed to evaluate whether a particular treatment is efficient and to check whether it is safe for human beings. In addition, clinical behaviours are costly and time-consuming to assess, and many clinical traits may fail to be observed during testing; furthermore, recruiting the right patients is also crucial. The entire trial process is also difficult. With the assistance of BD analytics, pharmaceutical industries can recruit the right patients for clinical trials, employing data such as genetic information, the status of the disease, and personality traits to increase the drug's success rate. This also helps in precisely determining the appropriate medicine(s) for treatment and the diagnosis of the considered disorders, performed using the most related and relevant data along with analysis of certain characteristics, such as behavioural patterns and genetic makeup. Using this BD, pharmaceutical companies can design personalized medicines in line with a particular patient's genetics and lifestyle.

The construction of medical BD involves not just a simple application and collection of medical data but, instead, is a complex systematic model. An existing research study [46] has discussed China's experience in constructing a regional medical BD ecosystem. The construction of the medical BD includes several institutions and high-level management, and cooperation was observed to enhance innovation and effectiveness. Compared with the construction of infrastructure, it is more time-consuming and challenging to develop proper data standards, data mining tools, and data integration. Similarly, another traditional study [47] has attempted to construct a proof-of-concept illustrating that BD approaches possess the capability to enhance the safety of drug monitoring in hospitals and, as such, can highly aid pharmaco-vigilance professionals to determine adverse drug events through data-driven targeted analysis of Drug-Drug Interactions (DDI). They also designed an automatic DDI detection model based on the treatment of the data and the laboratory analysis from electronic health records accumulated in a clinical data warehouse. The research results revealed that the developed DDI model worked effectively and that the time required for computation was manageable. This developed model can be used for regular monitoring processes.

Likewise, another traditional paper [48] has attempted to address data quality issues in electronic patient records using a computerized electronic patient report system with the abstraction of Map reduce and Apache HIVE of BD technology. The existing research also attempted to analyse which patients are spending more money, compared to patients with reduced maps. The data were obtained through a traditional system of Hadoop, through the functions of extract, transform, and load (ETL). The considered model was observed to resolve issues related to the use of conventional manual models. Security was also observed to be improved, as the system demands appropriate authentication for access. However, the developed model does not seem to send any alert regarding the expiration dates of drugs. In addition, factors such as assets and security were not included in the existing system.

The body of work presented in the literature survey indicates the growing importance and trend towards adapting BD in the fields of pharmacology, toxicology, and pharmaceuticals. The existing literature has demonstrated that BD can solve various problems and, so, the application of BD in these fields needs to be reinforced.

3. Benefits of Big Data

The data associated with the healthcare sector are enormous. They are stored in and withdrawn from clinics, hospitals, and insurance companies, resulting in the under-use of resources, data redundancy, and inadequacy. However, stakeholders have increased their voices and requirements to improve the exploitation and exploration of traditional data. With the employment of BD:

- Healthcare organizations can construct networks to bring about extensive changes in the educational field of medicine, practice, and research;

- the sorting, storage and analysis of patient information can be supported;
- the identification of therapeutic assessments, disease identification and prevention, surgical planning, and outcome predictions can be enhanced.

With clinical semiology, computer science, advanced imaging, radiology, biochemistry, and genomics, BD has emerged as a promising tool that can assist in developing a wide range of technical devices, surgical approaches, pharmacological therapies, and others.

3.1. Economic Benefits

Although well-developed countries spend a large amount of their gross domestic products on healthcare systems, the expenditures spent may not help to boost health outcomes. Furthermore, the increases in medical care expenses and health benefits indicate that additional strategies should be incorporated to enhance the efficiency of healthcare systems and protect public investments [49]. Despite having the highest healthcare expenditures, the U.S. has reported that it has not increased its life expectancy or last-day quality of life [50]. There is a rapid increment in relative economic demands, due to the sharp elevation in the age of the population and increasing chronic diseases. It has been estimated that the senior population will increase from 14 million to 19 million in 2020 and, by 2050, it will increase to 40 million. The impact of these demands has been observed in European regions. By employing BD as a tool for forecasting models, the economy of the healthcare sector can be improved, in order to develop expenditure projections for potential medicinal hazards, discover various policies that can assist in resolving intricate circumstances, and implement the most efficient approaches against various threats. It has been reported that BD might produce more than 300 billion in savings per year for the U.S. healthcare sector [50].

3.2. Technological Benefits

Social media and portable devices have become a common source of data collection and feedback results, due to their familiarity in the present population, closeness, and portability. In the medicinal sector, physiological sensors and electronic health records (EHRs) are the main aspects for the follow-up and monitoring of patients. All of these tools have generated a movement towards the revamping of resources. Thus, the mechanism to analyse and manage the information should also evolve. This hectic demand may be met through the use of BD analytics. For example, in hospitals, BD on smartphones can create an effective tool to forward inspirational and medicinal messages to patients, in order to enhance their lifestyles and accomplish a prominent treatment that will enhance their health and welfare [51].

BD has had a significant impact on the advancement of pharmaceutical science, but there are still many challenges to overcome. Our survey indicated the impact of BD on drug discovery and design, and allowed for speculation on future developments in the field. The challenges are related to implementing and maintaining the quality and privacy of data sets, and how the industry must adapt to welcome the BD era. Especially in the pharmaceutical industry, the challenges related to BD initiatives could provide the insights needed to turn the industry around. The anticipated future of BD applications demonstrates their potential for use in drug discovery to speed up drug manufacturing, improving productivity, drug performance, drug safety, clinical analysis, clinical study, and drug personalisation in drug efficacy, as represented in Figure 4.

Drug discovery is expected to be the most benefited field. By using BD, the following advantages can be achieved: 1. BD can help to identify new drug targets; 2. identifying new drug candidates; 3. improving the efficiency of drug discovery processes; 4. improving the accuracy of drug discovery predictions; 5. identifying potential drug interactions; and 6. identifying adverse drug reactions.

As for drug efficacy and performance, there are many benefits of BD in drug efficacy, including detecting patterns in large data sets, identifying new drug targets, and improving the accuracy of predictions regarding how a drug will behave in the body. There are

also many benefits of BD relating to drug performance. The prime benefit is that BD can help researchers and doctors to better understand how drugs work and how they can be improved. BD can also help doctors and patients make more informed decisions about which drugs to use and how to use them. Additionally, BD can help researchers to identify new drug targets and develop new drugs quickly and efficiently.

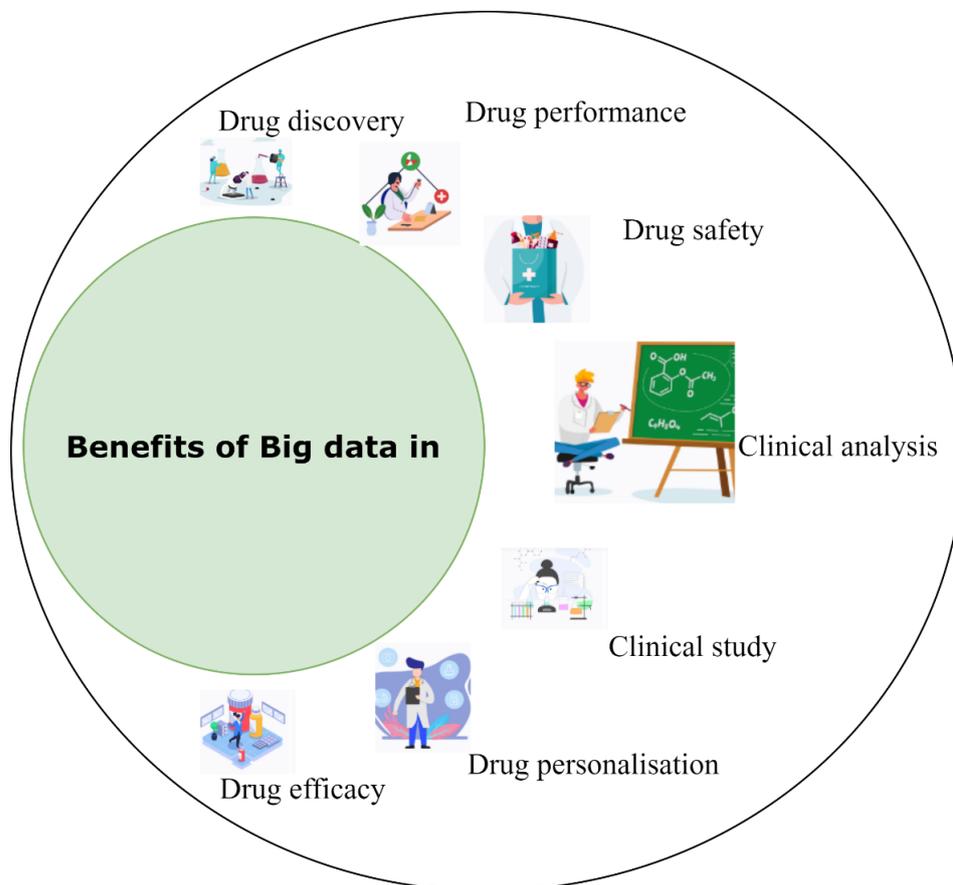


Figure 4. Benefits of BD in Pharmacology, Toxicology and Pharmaceutics.

The benefits of BD in clinical study analysis are as follows: 1. Increased data accuracy, due to the volume and variety of data sources; 2. faster identification of trends and correlations; 3. more efficient and effective clinical trial design and execution; 4. better understanding of patient populations and their needs; and 5. improved decision making, including more timely and effective interventions.

The benefits of BD in drug safety analysis are as follows: 1. Increased accuracy and precision in identifying potential safety concerns; 2. earlier detection of potential safety concerns; 3. more efficient and effective safety surveillance; and 4. better understanding the safety profiles of drug.

There are many benefits of BD in drug personalisation. The key benefit is that it can help improve the accuracy of predictions regarding how a particular drug will work in a particular person. This is because BD can help to improve our understanding of the complex interactions between drugs and the human body. It can also help identify which patients are most likely to respond to a particular drug and which patients may experience adverse effects.

4. Comparative Analysis

Based on the literature survey, Table 1 below presents the domains, the purposes, the broad areas of application, and the core advantages related to BD.

Table 1. Summary of BD in Pharmacology, Toxicology and Pharmaceutics.

	Purpose	Broad Area	Core Area
Pharmaceutical science [41]	Highlight the benefit by capturing the multi-dimensional contributions	Quality and privacy of data sets	Advancement in pharmaceutical
Biomedicine [43]	Improve patient care through personalised medicine programs.	Information technology and its impact on biomedicine.	Biomedicine to improve patient care.
Chemical toxicity [34]	Build a modern chemical toxicity research using historical data repositories.	Synthesis of bioassays and biological proteins/receptors.	Complex bioactivity evaluation
Toxicogenomics [33]	Store and channelise biological data in biomedical science.	High-throughput screening technologies	Explore hidden patterns of genomic data.
Pharmaceutics [28]	Productivity in the pharmaceutical industry by providing insights into improving output while reducing costs.	Pharmaceutical industry	Productivity
Translational Medicine [29]	Understanding the evolution of Translational Medicine.	Alternative methods	Translational medicine.
Toxicology [35]	Reveal patterns, trends, and associations of substances.	Substances exploration	Predict toxic properties.
Health Economics [27]	Informed patient-centred care and with healthcare decision-making and improving healthcare.	Dynamic simulation modelling (DSM)	Synergies
Drug development [39]	Identify novel drugs or repurposed existing drugs.	Drug development	Personalised precision medicine
Pharmacogenomics [44]	Identify genetic and genomic factors for drug treatment response.	Pharmacogenomics	Drug efficacy
Drug efficacy [45]	Machine language-based drug efficacy on patient BD sets.	Machine learning techniques	Drug performance
Drug discovery [38]	Improve pre-clinical drug discovery and development.	Life sciences	Data integrity.
Drug efficacy [52]	Improve the clinical trial design and improve safety.	Paediatric drug development	Clinical trials
Drug Safety and Efficacy [53]	Enhance clinical decisions and uncover the efficacy and safety of drugs.	Healthcare	Medication safety & effectiveness.
Pharmacogenomics/Bioinformatics [54]	Bioinformatics aspects of pharmacogenomic.	Pharmacogenomics	BD bioinformatics.
In Silico Mining/Inhibitor-Predictive modelling [40]	Find new Bromodomain BRD4 binders and build predictive BRD activity models.	Drug discovery	Predictive model building.
Drug development [39]	Therapeutic target discovery, candidate drug prioritisation, and inference accuracy improvement.	Drug development	Precision medicine

Table 1. *Cont.*

	Purpose	Broad Area	Core Area
Neurocritical Care [55]	Improve patient care	Healthcare industry	Neurocritical care unit
Mental illness [56]	Mental illness data analysis and inference.	Early detection and intervention strategies	Personalised approaches
Biomedicine [57]	Study human origins and migration and develop better treatments for diseases.	Medical industry	BD in medical research.
Psychiatry [58]	Uncover psychiatric disorders based on pathophysiological data	Machine learning	Deep learning
Natural products [59]	To enable discoveries that were impossible before.	Natural product sciences.	Natural product sciences.

5. Critical Analysis

The critical analysis was conducted with the aim of tracing which particular field has implemented BD. The results of the critical analysis are graphically represented in Figure 5.

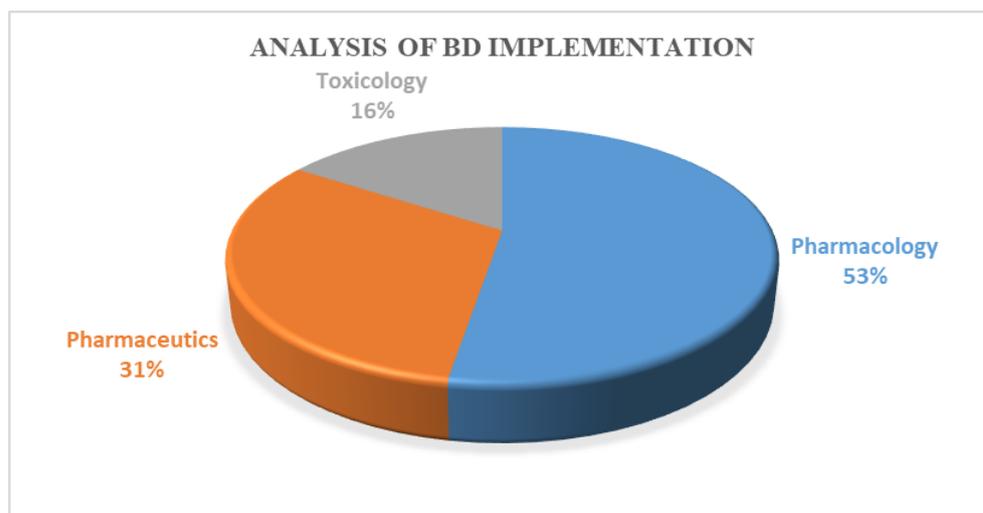


Figure 5. Analysis of BD implementation in Pharmacology, Toxicology and Pharmaceuticals.

From Figure 5, it is clear that the pharmacology sector is observed to employ BD more than the other two fields. This is because this sector produces more data while developing drugs for the identified chronic diseases. Additionally, through extensive research, it has been found that very few researchers have attempted to conduct research on the implementations of BD in the considered fields. Therefore, it might be more beneficial for the medicinal community to have investigators attempt to execute research regarding the technical implementation of BD in the drug development sector.

6. Limitations of BD

The limitations of BD include cases where the quality of the data is compromised and when the users fail to utilise BD analytics approaches correctly. In addition, certain other limitations are discussed in this section.

The main limitation of BD is mostly associated with the quality of data. The main determinant of this aspect is completely dependent on the source of the data. In addition, the management and storage of data also play a vital role in the characteristics of the data. Some significant features regarding data quality are data completeness, accuracy, and adequacy. The modification of these features may pose a threat to users, especially

when using BD, as it can endanger the integrity of the associated results of the analysis [51]. Thus, BD is said to return inaccurate results due to corruption of the quality of the data, resulting from the generation of false assumptions. These false assumptions might produce weak knowledge with large errors. This does not result only from the data quality, but also the data selection approach and the sample size. Large data sets with a high number of attributes can enable statistically significant results to be obtained; however, on the other hand, due to the huge data size, users may choose data arbitrarily while neglecting information regarding the data representativeness, resulting in selection bias. In both cases, the results and the accuracy of a BD-based approach might be compromised [51].

These aforementioned limitations also involve the user. It is significant to understand when to implement BD to resolve certain issues. Users should be well-aware of the features of the data and perceive compatibility among the data sets to obtain an appropriate and beneficial analysis, and should be well-aware of the underlying difficulties that arise when data are compared within domains without similar features and attributes [60]. To correctly utilise BD analytics, the main objectives and implementation results should be understood a priori. Additionally, users should educate themselves, in order to understand the possible mechanism(s) behind the researched phenomenon or object, in order to hypothesise the possible results. Finally, they should estimate the outcomes with the presumed objective at the initial step. If the users neglect or commit any mistakes in the steps above, the BD mechanism will likely produce erroneous results [60].

Table 2. Challenges and future directions.

BD Impacting Domain	Benefits	Challenges
Drug discovery	<ul style="list-style-type: none"> • Identify new drug targets • Identify new drug candidates • Efficiency of drug discovery processes • Drug discovery predictions • Potential drug interactions • Adverse drug reactions 	<ul style="list-style-type: none"> • BD initiative process • Privacy protection of data sets • Data governance • Regional data compliance law • Data storage • Data security • Data sharing • Data collaboration
Drug efficacy & performance	<ul style="list-style-type: none"> • Detect patterns in large datasets • Identify new drug targets • Accuracy of predictions • How drugs work • Informed decisions • Faster development of new drugs 	<ul style="list-style-type: none"> • General Data Protection Regulation • Data capturing process • Data protection • Data analytics • Data digitisation process • Data exploration and navigation
Clinical study analysis	<ul style="list-style-type: none"> • Increased accuracy • Identification of safety concerns • Effective safety surveillance • The safety profile of a drug 	<ul style="list-style-type: none"> • Data authenticity • Identity governance • Human errors • Data maintenance • Data interpretation techniques
Drug safety analysis	<ul style="list-style-type: none"> • Increased accuracy and precision • Detection of safety concerns • Efficient safety surveillance • Understanding the profile of a drug 	<ul style="list-style-type: none"> • Data transformation • Data benchmark • Data trend & pattern development
Drug personalisation	<ul style="list-style-type: none"> • Understanding human & drug interactions • Identify the patient’s response to drug • Equivalent risk of adverse effects 	<ul style="list-style-type: none"> • Real-time data inference • Real-time Data collection infrastructure • Real-time pre-processing algorithms • Real-time mining and extraction

Based on the summary presented in Table 2, future research lines to address the existing challenges may be gleaned to take the domain in new directions, helping to advance the considered fields and, thereby, indirectly helping to better humankind.

7. Conclusions and Discussion

The pharmaceutical industry is facing a challenge in productivity, and BD initiatives may provide the insights necessary to turn the industry around. Considering this, the present study detailed a substantial attempt to review the existing literature regarding the implementation of BD in the pharmacology, pharmaceuticals, and toxicology sectors. The pharmacology, toxicology, and pharmaceuticals fields are still in the early stage of BD adoption. Additionally, according to the critical analysis, the pharmacology sector has employed BD more than the other two sectors. Based on our survey, the key inferences were as follows: first, BD can help researchers better understand the effects of drugs and other chemicals on the human body, which can help to improve the safety and efficacy of drugs and other chemicals; second, BD can help to improve the accuracy of predictions regarding the effects of drugs and chemicals, which can improve safety in drug development and help to avoid potential adverse drug interactions; finally, BD can help improve our understanding of how the body metabolises drugs and other chemicals, which can improve the safety and efficacy of drugs and other chemicals. The domains considered in our survey are ultimately necessary for humanity, and BD may significantly impact the betterment of these domains. BD has revolutionary potential, providing new ways to understand and predict the effects of drugs. However, BD in this domain also poses new challenges, which should be taken up as key research problems. Despite these challenges, in the future, BD will likely play an important role in pharmacology, toxicology, and pharmaceuticals, critically helping to improve drug safety and efficacy.

Author Contributions: Review design, Ideation, initial draft, figure illustration and conceptualization, K.L.B.; literature curation, review and editing, R.S.O., E.D.G.; Additional illustration, additional data curation E.Y.A.; Improve the presentation, validation and restructuring of the manuscript, C.A.; Proofreading the manuscript and improvements, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Laurence, D. What is pharmacology? A discussion. *Trends Pharmacol. Sci.* **1997**, *18*, 1051–1052. [[CrossRef](#)] [[PubMed](#)]
2. Mückter, H. What is toxicology and how does toxicity occur? *Best Pract. Res. Clin. Anaesthesiol.* **2003**, *17*, 5–27. [[CrossRef](#)] [[PubMed](#)]
3. Hassan, W.; Zafar, M. Pharmacology, Toxicology, and Pharmaceuticals Research Output in One Hundred and Fifty Countries for the Year 2019–2020. *Can. J. Med.* **2021**, *3*, 56–60. [[CrossRef](#)]
4. Biglu, M.; Hossein, O. Scientific Profile of Pharmacology, Toxicology and Pharmaceuticals Fields in Middle East Countries: Impacts of Iranian Scientists. *Int. J. Adv. Pharm. Sci.* **2010**, *1*, 122–127.
5. Pc, S.; Sherimon, V.; Sp, P.; Nair, R.V.; Mathew, R. A Systematic Review of Clinical Decision Support Systems in Alzheimer's Disease Domain. *Int. J. Online Biomed. Eng.* **2021**, *17*, 74–90.
6. Gattan, A.M. A Knowledge Based Analysis on Big Data Analytics in Optimizing Electronic Medical Records in Private Hospitals. *Int. J. Online Biomed. Eng.*, **2021**, *17*, 119–134. [[CrossRef](#)]
7. Chan, C.L.; Chang, C.C. Big Data, Decision Models, and Public Health. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6723. [[CrossRef](#)]
8. Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci* **2019**, *20*, 2783. [[CrossRef](#)]
9. Bernetti, M.; Bertazzo, M.; Masetti, M. Data-Driven Molecular Dynamics: A Multifaceted Challenge. *Pharmaceuticals* **2020**, *13*, 253. [[CrossRef](#)]
10. Price, W.N.; Cohen, I.G. Privacy in the age of medical big data. *Nat. Med.* **2019**, *25*, 37–43. [[CrossRef](#)]
11. Beam, A.L.; Kohane, I.S. Big Data and Machine Learning in Health Care. *JAMA* **2018**, *319*, 1317–1318. [[CrossRef](#)] [[PubMed](#)]
12. Li, Y.; Zhang, Y.; Li, W.; Jiang, T. Marine Wireless Big Data: Efficient Transmission, Related Applications, and Challenges. *IEEE Wirel. Commun.* **2018**, *25*, 19–25. [[CrossRef](#)]

13. Bello-Orgaz, G.; Jung, J.J.; Camacho, D. Social big data: Recent achievements and new challenges. *Inf. Fusion* **2016**, *28*, 45–59. [[CrossRef](#)] [[PubMed](#)]
14. Basha, S.A.K.; Basha, S.M.; Vincent, D.R.; Rajput, D.S. Chapter 11—Challenges in Storing and Processing Big Data Using Hadoop and Spark. In *Deep Learning and Parallel Computing Environment for Bioengineering Systems*; Academic Press: Cambridge, MA, USA, 2019; pp. 179–187.
15. Vesoulis, Z.A.; Husain, A.N.; Cole, F.S. Improving child health through Big Data and data science. *Pediatr. Res.* **2022**, 1–8. [[CrossRef](#)]
16. Venkatesh, R.; Balasubramanian, C.; Kaliappan, M. Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique. *J. Med. Syst.* **2019**, *43*, 272. [[CrossRef](#)] [[PubMed](#)]
17. Singh, P.; Singh, A. Growth Trend in Global Big Data Research Publications as Seen From SCOPUS Database. *Prof. J. Libr. Inf. Technol.* **2018**, *8*, 49–61.
18. Elhoseny, M.; Abdelaziz, A.; Salama, A.S.; Riad, A.M.; Muhammad, K.; Sangaiyah, A.K. A hybrid model of Internet of Things and cloud computing to manage big data in health services applications. *Future Gener. Comput. Syst.* **2018**, *86*, 1383–1394. [[CrossRef](#)]
19. Xing, W.; Bei, Y. Medical Health Big Data Classification Based on KNN Classification Algorithm. *IEEE Access* **2020**, *8*, 28808–28819. [[CrossRef](#)]
20. Wang, Y.; Kung, L.; Wang, W.Y.C.; Cegielski, C.G. An integrated big data analytics-enabled transformation model: Application to health care. *Inf. Manag.* **2018**, *55*, 64–79. [[CrossRef](#)]
21. Wang, Y.; Kung, L.; Byrd, T.A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Chang.* **2018**, *126*, 3–13. [[CrossRef](#)]
22. Manogaran, G.; Varatharajan, R.; Lopez, D.; Kumar, P.M.; Sundarasekar, R.; Thota, C. A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. *Future Gener. Comput. Syst.* **2018**, *82*, 375–387. [[CrossRef](#)]
23. Chebana, F.; Ouarda, T.B.M.J. Multivariate non-stationary hydrological frequency analysis. *J. Hydrol.* **2021**, *593*, 125907. [[CrossRef](#)]
24. Shi, Q.; Abdel-Aty, M. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 380–394. [[CrossRef](#)]
25. Cejun, C.A.O.; Congdong, L.I.; Yu, W.; Ting, Q.U.; Wei, Z. Evolution and governance mechanism of urban public safety risk in big data era. *China Saf. Sci. J.* **2020**, *27*, 151.
26. Marvin, H.J.P.; Janssen, E.M.; Bouzembrak, Y.; Hendriksen, P.J.M.; Staats, M. Big data in food safety: An overview. *Crit. Rev. Food Sci. Nutr.* **2017**, *57*, 2286–2295. [[CrossRef](#)] [[PubMed](#)]
27. Marshall, D.A. Transforming Healthcare Delivery: Integrating Dynamic Simulation Modelling and Big Data in Health Economics and Outcomes Research. *Pharmacoeconomics* **2016**, *34*, 115–126. [[CrossRef](#)] [[PubMed](#)]
28. Tormay, P. Big Data in Pharmaceutical R&D: Creating a Sustainable R&D Engine. *Pharmaceut. Med.* **2015**, *29*, 87–92.
29. Jordan, L. The problem with Big Data in Translational Medicine. A review of where we have been and the possibilities ahead. *Appl. Transl. Genom.* **2015**, *6*, 3–6.
30. Wang, B. Safety intelligence as an essential perspective for safety management in the era of Safety 4.0: From a theoretical to a practical framework. *Process Saf. Environ. Prot.* **2021**, *148*, 189–199. [[CrossRef](#)]
31. Wang, B.; Wu, C. Safety informatics as a new, promising and sustainable area of safety science in the information age. *J. Clean. Prod.* **2020**, *252*, 119852. [[CrossRef](#)]
32. Streun, G.L.; Elmiger, M.P.; Dobay, A.; Ebert, L.; Kraemer, T. A machine learning approach for handling big data produced by high resolution mass spectrometry after data independent acquisition of small molecules—Proof of concept study using an artificial neural network for sample classification. *Drug Test. Anal.* **2020**, *12*, 836–845. [[CrossRef](#)] [[PubMed](#)]
33. Chung, M.H. Asymmetric author-topic model for knowledge discovering of big data in toxicogenomics. *Front. Pharmacol.* **2015**, *6*, 81. [[CrossRef](#)] [[PubMed](#)]
34. Zhu, H.; Zhang, J.; Kim, M.T.; Boison, A.; Sedykh, A.; Moran, K. Big data in chemical toxicity research: The use of high-throughput screening assays to identify potential toxicants. *Chem. Res. Toxicol.* **2014**, *27*, 1643–1651. [[CrossRef](#)] [[PubMed](#)]
35. Hartung, T. Making big sense from big data in toxicology by read-across. *ALTEX* **2016**, *33*, 83–93. [[CrossRef](#)] [[PubMed](#)]
36. Yan, X.; Sedykh, A.; Wang, W.; Yan, B.; Zhu, H. Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations. *Nat. Commun.* **2020**, *11*, 2519. [[CrossRef](#)] [[PubMed](#)]
37. Clark, M.; Steger-Hartmann, T. A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. *Regul. Toxicol. Pharmacol.* **2018**, *96*, 94–105. [[CrossRef](#)] [[PubMed](#)]
38. Brothers, J.F., 2nd; Ung, M.; Escalante-Chong, R.; Ross, J.; Zhang, J.; Cha, Y.; Lysaght, A.; Funt, J.; Kusko, R.L. Integrity, standards, and QC-related issues with big data in pre-clinical drug discovery. *Biochem. Pharmacol.* **2018**, *152*, 84–93. [[CrossRef](#)] [[PubMed](#)]
39. Qian, T.; Zhu, S.; Hoshida, Y. Use of big data in drug development for precision medicine: An update. *Expert Rev. Precis. Med. Drug Dev.* **2019**, *4*, 189–200. [[CrossRef](#)]
40. Casciuc, I. Pros and cons of virtual screening based on public ‘Big Data’: In silico mining for new bromodomain inhibitors. *Eur. J. Med. Chem.* **2019**, *165*, 258–272. [[CrossRef](#)]
41. Dossetter, A.G.; Ecker, G.; Laverty, H.; Overington, J. ‘Big data’ in pharmaceutical science: Challenges and opportunities. *Future Med. Chem.* **2014**, *6*, 857–864.
42. Lv, Z.; Qiao, L. Analysis of Healthcare Big Data. *Future Gener. Comput. Syst.* **2020**, *109*, 103–110. [[CrossRef](#)]

43. Costa, F.F. Big data in biomedicine. *Drug Discov. Today* **2014**, *19*, 433–440. [[CrossRef](#)] [[PubMed](#)]
44. Li, R.; Kim, D.; Ritchie, M.D. Methods to analyze big data in pharmacogenomics research. *Pharmacogenomics* **2017**, *18*, 807–820. [[CrossRef](#)] [[PubMed](#)]
45. Koren, G.; Nordon, G.; Radinsky, K.; Shalev, V. Machine learning of big data in gaining insight into successful treatment of hypertension. *Pharmacol. Res. Perspect.* **2018**, *6*, 396. [[CrossRef](#)] [[PubMed](#)]
46. Li, B.; Li, J.; Jiang, Y.; Lan, X. Experience and reflection from China's Xiangya medical big data project. *J. Biomed. Inform.* **2019**, *93*, 103149. [[CrossRef](#)]
47. Bouzillé, G. An Automated Detection System of Drug-Drug Interactions from Electronic Patient Records Using Big Data Analytics. *Health Technol. Inf.* **2019**, *264*, 45–49.
48. Zhang, C.; Ma, R.; Sun, S.; Li, Y.; Wang, Y.; Yan, Z. Optimizing the Electronic Health Records Through Big Data Analytics: A Knowledge-Based View. *IEEE Access* **2019**, *7*, 136223–136231. [[CrossRef](#)]
49. Papanicolaou, I.; Woskie, L.R.; Jha, A.K. Health Care Spending in the United States and Other High-Income Countries. *JAMA* **2018**, *319*, 1024. [[CrossRef](#)]
50. Bodas-Sagi, D.; Labeaga, J. Big Data and Health Economics: Opportunities, Challenges and Risks. *Int. J. Interact. Multimed. Artif. Intell.* **2018**, *4*, 47. [[CrossRef](#)]
51. Wang, W.; Krishnan, E. Big data and clinicians: A review on the state of the science. *JMIR Med. Inf.* **2014**, *2*, e2913. [[CrossRef](#)]
52. Mulugeta, L.Y.; Yao, L.; Mould, D.; Jacobs, B.; Florian, J.; Smith, B.; Sinha, V.; Barrett, J.S. Leveraging Big Data in Pediatric Development Programs: Proceedings From the 2016 American College of Clinical Pharmacology Annual Meeting Symposium. *Clin. Pharmacol. Ther.* **2018**, *104*, 81–87. [[CrossRef](#)] [[PubMed](#)]
53. Christensen, M.L.; Davis, R.L. Identifying the 'Blip on the Radar Screen': Leveraging Big Data in Defining Drug Safety and Efficacy in Pediatric Practice. *J. Clin. Pharmacol.* **2018**, *58*, 86–93. [[CrossRef](#)] [[PubMed](#)]
54. Barrot, C.C.; Woillard, J.B.; Picard, N. Big data in pharmacogenomics: Current applications, perspectives and pitfalls. *Pharmacogenomics* **2019**, *20*, 609–620. [[CrossRef](#)] [[PubMed](#)]
55. Foreman, B. Neurocritical Care: Bench to Bedside (Eds. Claude Hemphill, Michael James) Integrating and Using Big Data in Neurocritical Care. *Neurotherapeutics* **2020**, *17*, 593–605. [[CrossRef](#)] [[PubMed](#)]
56. Rajula, H.S.R.; Manchia, M.; Carpiniello, B.; Fanos, V. Big data in severe mental illness: The role of electronic monitoring tools and metabolomics. *Per. Med.* **2021**, *18*, 75–90. [[CrossRef](#)]
57. Liu, Y.; Li, N.; Zhu, X.; Qi, Y. How wide is the application of genetic big data in biomedicine. *Biomed. Pharmacother* **2021**, *133*, 111074. [[CrossRef](#)] [[PubMed](#)]
58. Koppe, G.; Meyer-Lindenberg, A.; Durstewitz, D. Deep learning for small and big data in psychiatry. *Neuropsychopharmacology* **2021**, *46*, 176–190. [[CrossRef](#)]
59. Cech, N.B.; Medema, M.H.; Clardy, J. Benefiting from big data in natural products: Importance of preserving foundational skills and prioritizing data quality. *Nat. Prod. Rep.* **2021**, *38*, 1947–1953. [[CrossRef](#)]
60. Ellaway, R.H.; Pusic, M.V.; Galbraith, R.M.; Cameron, T. Developing the role of big data and analytics in health professional education. *Med. Teach.* **2014**, *36*, 216–222. [[CrossRef](#)]