

## Review

# Enhancing Digital Health Services with Big Data Analytics

Nisrine Berros <sup>1,\*</sup> , Fatna El Mendili <sup>2</sup>, Youness Filaly <sup>1</sup> and Younes El Bouzekri El Idrissi <sup>1</sup><sup>1</sup> Engineering Sciences Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra 14000, Morocco<sup>2</sup> Image Laboratory, School of Technology, Moulay Ismail University, Meknes 50050, Morocco

\* Correspondence: nisrine.berros@uit.ac.ma

**Abstract:** Medicine is constantly generating new imaging data, including data from basic research, clinical research, and epidemiology, from health administration and insurance organizations, public health services, and non-conventional data sources such as social media, Internet applications, etc. Healthcare professionals have gained from the integration of big data in many ways, including new tools for decision support, improved clinical research methodologies, treatment efficacy, and personalized care. Finally, there are significant advantages in saving resources and reallocating them to increase productivity and rationalization. In this paper, we will explore how big data can be applied to the field of digital health. We will explain the features of health data, its particularities, and the tools available to use it. In addition, a particular focus is placed on the latest research work that addresses big data analysis in the health domain, as well as the technical and organizational challenges that have been discussed. Finally, we propose a general strategy for medical organizations looking to adopt or leverage big data analytics. Through this study, healthcare organizations and institutions considering the use of big data analytics technology, as well as those already using it, can gain a thorough and comprehensive understanding of the potential use, effective targeting, and expected impact.

**Keywords:** big health data; electronic health records; analytics; machine learning; NoSQL database



**Citation:** Berros, N.; El Mendili, F.; Filaly, Y.; El Bouzekri El Idrissi, Y. Enhancing Digital Health Services with Big Data Analytics. *Big Data Cogn. Comput.* **2023**, *7*, 64. <https://doi.org/10.3390/bdcc7020064>

Academic Editors: Domenico Talia and Fabrizio Marozzo

Received: 13 February 2023

Revised: 9 March 2023

Accepted: 16 March 2023

Published: 30 March 2023



**Copyright:** © 2023 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 health sector has always generated a large amount of data due to the increased record-keeping needs in the context of patient care [1]. Much of this available and particularly valuable data are in a semi-structured or unstructured form. Further, its diverse and dynamic nature makes it challenging to extract valuable insights through the use of traditional analytical methods [2]. Thus, big data in the field of health is an important issue, not only because of its enormous volume but also because of its diversity and how quickly it can be managed [3]. The human capacity to process this data is limited, making effective decision support necessary. Due to this, big data analytics must be integrated into the health industry. Big data analytics has the capability to examine a diverse set of intricate data and generate valuable information that would otherwise be unobtainable. In the healthcare field, it can not only detect emerging trends but also enhance the quality of healthcare, decrease costs, and facilitate prompt decision-making [4]. As stated in the McKinsey International Institute report, if big data are harnessed and used effectively, the U.S. healthcare system value will be saved more than \$300 billion annually, with approximately two-thirds of that amount coming from a reduction in healthcare costs of around 8%. By making use of big data technology and the automated analysis of the results, it is possible for useful information to emerge that until recently has remained in obscurity. The ability of big data analytics to recognize the heterogeneity of diseases allows not only a timely diagnosis but also for the evaluation of existing treatments [5,6]. Big data analytics can turn large amounts of continuous data into actionable insights by analyzing and connecting

information from multiple sources. This capability to provide this kind of insight is especially crucial, particularly in emergency medical situations, as it can greatly determine the outcome of a patient's life or death [7]. We have seen during the coronavirus pandemic the usefulness of medical data and how such information can be helpful in the management of health crises during a pandemic. Health organizations must seriously consider integrating the technological tools required to treat this massive amount of data that has the potential to save lives. The digitization of clinical examinations and medical records in healthcare systems has become a widespread and accepted norm since the development of computer systems and their potential [8].

### *1.1. Motivation*

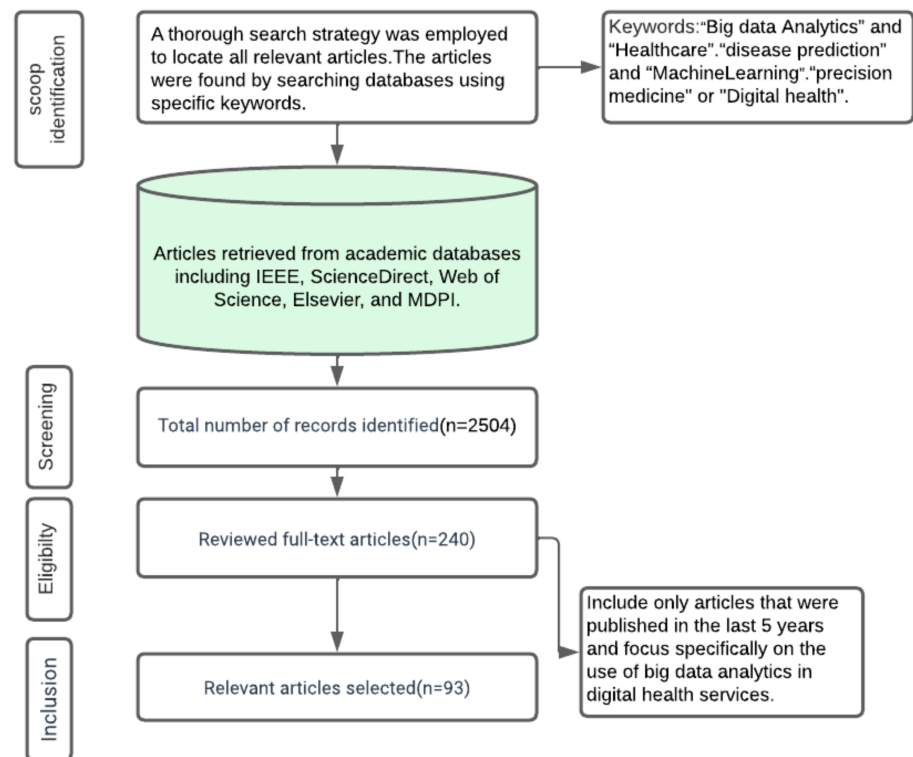
The main contribution of this article is to give an analytical insight into the use of big data analytics in medical institutions. This paper aims to understand how big data are applied in the field of digital health services, to present the available tools and applications, describe the most important actions and research work, as well as the technical and organizational challenges that arise. Healthcare organizations and institutions considering implementing big data analytics technology, as well as those already using it, have the opportunity through this study to gain a comprehensive and detailed understanding of its potential for use, effective targeting, expected impact, and the challenges they will face.

### *1.2. Research Methodology*

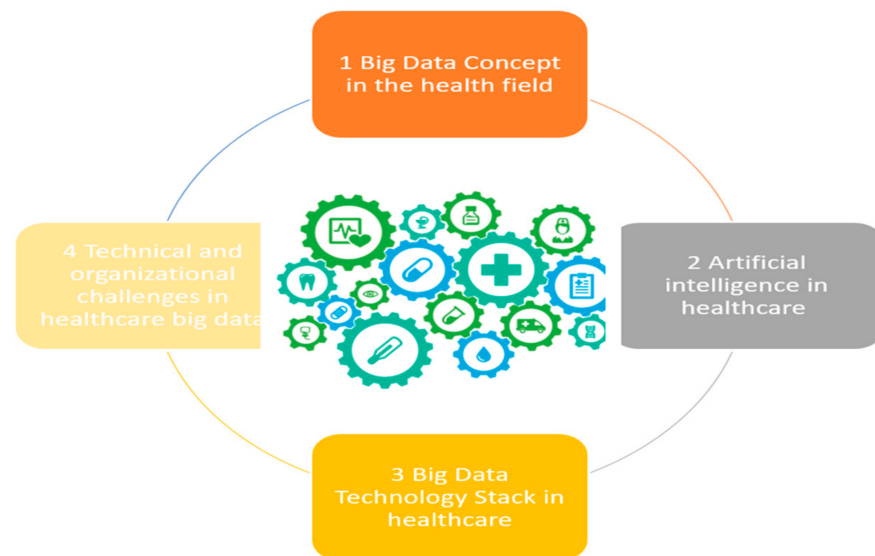
For this review, we focused on exploring major questions related to big data analytics in healthcare. We conducted a thorough search of the literature articles indexed in Scopus, Web of Science, Science Direct, and other reputable databases. We used a combination of keywords and Boolean operators to refine our search, including terms related to big data analytics, healthcare, and data analysis. We selected articles based on their relevance to our research question, as well as their citation count and impact factor. Using Zotero, a reference management tool, we organized the selected articles and made notes on their content and findings. We also reviewed relevant conference proceedings to ensure the comprehensive coverage of the topic. By conducting a thorough review of the literature, we aimed to provide a comprehensive overview of the current state of knowledge on big data analytics in healthcare. Figure 1 presents the research methodology used.

### *1.3. Paper Organization*

This review paper will be organized as follows: first, in the introduction, we present our motivations and work related to the topic. Then, the concept of using big data in health will be discussed. The second part focuses specifically on the features and sources most commonly used for big data analysis in healthcare. Additionally, instances of the classification of analytics in medicine are provided. Then, in Part 3, an overview of machine learning techniques and their uses in medicine are presented. The big data technology stack in healthcare is presented in Part 4. In Part 5, different technical and organizational challenges in healthcare are discussed and analyzed. In part 6 a Proposed Strategy for Implementing Big Data Analytics in Healthcare is presented. The final part of the paper is the conclusion, where we will summarize and draw final insights (Figure 2).



**Figure 1.** Research methodology followed.



**Figure 2.** Topics covered in this article.

#### 1.4. Existing Surveys

There are many studies in the literature that show the potential big data analytics can offer to medical organizations and what type of data can be analyzed. However, very few studies have shown how data analysis technology is performed in the healthcare sector and what the major organizational challenges are that an organization willing to integrate big data into their system may face. Table 1 presents a summary of the key related reviews, including a description of each review's contribution and the topic covered. A comparison of our work to the others is provided at the bottom of the table.

**Table 1.** An overview of the related work.

ID	Reference	Year	Overview
1	[9]	2015	This survey examines the utilization of big data in healthcare and explores the benefits it can bring to the healthcare industry. It delves into the various data sources that should be utilized and brought together for analysis. Numerous difficulties with big healthcare data are also addressed.
2	[10]	2016	This paper addresses both the difficulties and potential of big data in the medical industry, including the pipeline for processing it. It also presents a variety of machine-learning techniques for mining and analyzing data.
3	[11]	2018	This paper gives an in-depth look at the technologies and techniques used to create analytical applications in healthcare. It examines the progression of healthcare big data and the data mining algorithms used, including their general usage and specific applications in healthcare. Additionally, it covers the essential platforms and technologies needed for a successful health analytics solution.
4	[12]	2019	This paper discusses the significant impacts of big data on various medical actors and healthcare providers, as well as the difficulties in utilizing all of this big data and the applications that are already accessible.
5	[13]	2019	In this article, the authors investigate the various technologies, tools, and applications used for data integration in healthcare. They also address the current difficulties encountered when integrating large amounts of healthcare data and explore potential future research opportunities in this field.
6	[14]	2020	This paper covers technological advances and advancements in big data analytics in healthcare as well as infrastructure, artificial intelligence (AI), and cloud computing. In addition, it also explores the primary techniques, frameworks, and resources for big healthcare data analytics in medical engineering.
7	[15]	2020	This paper provides an overview of big data analytics systems used in healthcare and highlights the various algorithms, techniques, and tools that can be implemented in cloud, wireless, and internet of things environments. The authors propose the concept of SmartHealth as a way to bring all these platforms together and have a unified standard learning healthcare system for the future.
8	[16]	2021	This study places particular emphasis on applications of big data analytics for the healthcare field, especially NoSQL databases. The authors also propose a BDA architecture dubbed Med-BDA for the healthcare industry to address BDA's issues in this field. They also present strategies to make their proposals successful, and the authors in this article also make a comparison with the literature to justify the importance of their work.
9	[17]	2021	This article discusses the use of ML, big data, and blockchain technology in medicine, healthcare, public health surveillance, and case prediction during the COVID-19 pandemic and other epidemics. It also covers potential challenges for medical professionals and health technologists in creating future-oriented models to enhance human life.
10	[5]	2022	This paper examines the fundamental concepts of big data, its management, analysis, and potential applications, specifically in the field of health.
11	[8]	2022	The primary objective of this paper is to gather and categorize the utilization of big data from various perspectives and to provide a comprehensive analysis of the application of big data analytics within medical institutions in Poland.
12	[18]	2022	This study examines the literature on big data applications in the context of the COVID-19 pandemic, specifically focusing on their use in four key industries, including healthcare. By comparing the utilization of big data applications before and during the pandemic, the paper provides an overview of the current significance of big data in the COVID-19 era and how these applications align with relevant big data analytics models.
13	[4]	2022	This survey explores the utilization of big data (BD) in the fields of pharmacy, pharmacology, and toxicology. It examines how researchers have employed BD to address issues and discover solutions. The survey uses a comparative analysis to examine the application of big data in these three domains.

### 1.5. Current Survey

Our study provides a comprehensive and in-depth examination of the utilization of big data analytics in medical institutions. Unlike other surveys on the subject, we not only present a summary of the available tools and applications but also delve deeper into the key actions and research efforts being undertaken in this field. Additionally, we address the technical and organizational challenges that arise when implementing big data analytics in digital health services. In the end, we offer a simple strategy that can be adopted by organizations that want to integrate big data analytics based on the best practice in the field of healthcare. The goal of this study was to provide healthcare organizations and institutions with a clear understanding of the potential use, effective targeting, and expected impact of big data analytics technology, thus helping them make informed decisions about its implementation.

## 2. Big Data Concepts in the Health Field

Big data are generally viewed as a set of data that are too large or too heterogeneous and complex in structure to be handled by traditional data processing software. Big data challenges include collecting, storing, analyzing, transferring, sharing, and visualizing the information it contains. Scientists, entrepreneurs, and medical professionals are often required to use data from a range of sources, including big data from the international literature, the Internet, medical records, patient registries, and even ‘smart’ devices.

### 2.1. Features of Big Data in Healthcare

- Volume

In digital health, the increase in the amount of data is a result of both the digitization of already available data and the creation of new data formats. The volume of data available consists of personal medical records, radiology and fluoroscopy images, clinical trials, surveys, demographic data, human genomes, genetic sequences, etc. The exponential rise in data in the healthcare industry is due to the integration of new types of big data, including three-dimensional images, biological data, and data from sensor technologies. To handle the large volumes of healthcare data, for example, authors in [19] have used natural language processing (NLP) techniques to extract meaningful information from clinical notes in EHRs for complementary and integrative health (CIH). By automating the extraction of CIH information, this research can address the challenge of dealing with the volume of unstructured data in EHRs.

- Variety

Traditionally, the vast majority of data available in healthcare have been unstructured data, such as medical records and handwritten notes from medical and nursing staff describing symptoms, indications, behavior, medical images, etc. Of course, there has been an upsurge in structured data in recent years, such as electronic drug prescribing information, quantitative data on an instrument and test measurements, and general data that are attempted to be recorded in a single structure so that they can be used as a basis for data analysis. In addition to the data that are obviously recorded, data from new sources, such as wellness devices that record patients’ pulse or sleep time, social networks, and genomic research, the use of different data sources allows for the obtaining of faster and more reliable results. In the study [20], the authors demonstrate how monitoring social media conversations related to vaccines can address the various problems of big data by providing a way to organize and make sense of a large amount of unstructured social media data.

- Velocity

In healthcare, most data traditionally come from static sources, such as X-rays, hospital documents, patient records, health logs, etc. In some applications, however, it is necessary to process and use the data in real-time, for example, to monitor blood pressure and heart

function during surgery [21]. There are also cases where data processing is necessary at a relatively slower pace, such as the daily determination of glucose levels in diabetics [22]. Another example is information about a known disease, which develops at a much slower rate in terms of percentage compared to a new epidemic that is developing. In the latter case, the data arrive at a high rate and are “new” information. It is imperative to quickly process this information in order to resolve the matter in a timely manner. To analyze healthcare data in real-time or near real-time, researchers have proposed the use of big data analytics to develop predictive models that can detect and respond to health emergencies. For example, using machine learning algorithms to predict the outbreak of infectious disease and monitor the spread of the disease in real-time [23].

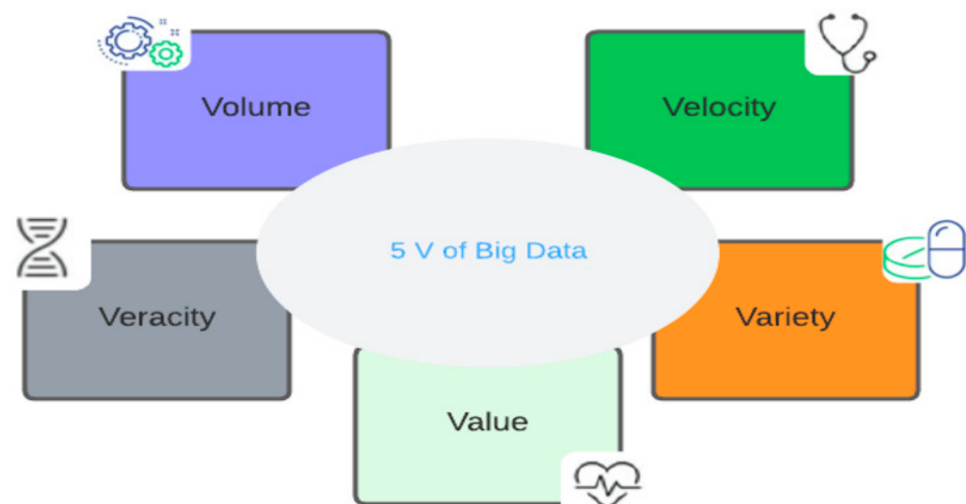
- Veracity

There are several similarities between the study of data reliability in financial transactions and healthcare: the accuracy of patient data, correctly filling in hospital or clinic fields, patient insurance, linkage to bank accounts, the recording of payment amounts, etc. [3]. Of course, in the health sector, there are data that are not observed in other sectors, such as information about a diagnosis, treatment, administration of medication, care, and any other information deemed necessary to be recorded. The validity of these data is, in any case, as important as the data mentioned above. Ensuring the accuracy of big data is critical in healthcare to prevent medical errors, incorrect diagnoses, and treatment decisions. To address this issue, various techniques such as data cleaning, data validation, data integration, and normalization are used to ensure that the data are reliable and consistent.

- Value

The cost of healthcare is unsustainable and constantly rising. However, the multiple benefits offered by the use and exploitation of big data in healthcare are far more numerous. For example, in the study [24], the authors developed machine learning algorithms to predict hospital readmissions and reduce healthcare costs. The algorithms were able to accurately predict readmissions, and healthcare providers were able to intervene early and provide targeted interventions to reduce the risk of readmission.

The following figure, Figure 3, illustrates the 5 Vs of big data in the healthcare sector.



**Figure 3.** Big data characteristics in the healthcare sector.

## 2.2. Data Sources

For the healthcare sector, relevant data are needed to build systems that have a positive impact on the health and well-being of individuals. In this section, we introduce three data sources and analyze how each can be leveraged through concrete examples.

- Electronic Health Records



Electronic patient records are a source of an enormous amount of data containing information about the social, demographic, medical, and health aspects of the patient's health. However, without reliable decision support, the human brain can only process a certain amount of information. In order to develop real-time knowledge and support systems that are preventive, predictive, and diagnostic in the healthcare industry, it is important to have an infrastructure that is constantly updated. Computational models are required to assist medical professionals in data organization, pattern recognition, and result interpretation [25]. the following table shows some of the possible data that an electronic medical record may contain, as well as their data type (Table 2).

**Table 2.** Possible content of an electronic medical record.

Data	Format of Representation
First and Last Name	Text
Gender	Code
Date of birth	Date
Clinical notes	Text
Laboratory tests, X-ray tests	Code/Number
Radiological examinations	Image/Signal
Medications	Number/code/Text
Vaccines	Code

- **Social networks**

The resurgence of communication via social networks is one of the most important factors in the dramatic evolution of healthcare. According to a recent estimate, approximately one billion tweets have been exchanged, illustrating the depth of communication between organizations, patients, and providers. Social networks now offer researchers new ways to reach out to patients and include them in their research. One such project is TuAnalyze, a collaboration between TuDiabetes1 and Boston Children's Hospital that allows diabetics to track, assess, and share their findings while actively participating in diabetes research [26]. Without a doubt, one of the most intriguing applications of data analytics is its ability to predict and monitor significant epidemics for the benefit of public health. Predictions of major health outcomes, such as an exacerbation of asthma attacks, can be improved by combining social network analysis with environmental data. Specifically, Google searches, Twitter activity, and air quality data can be used to estimate the number of daily emergency room admissions for an asthma event [27]. According to a study published in [28], there was a rise in tweets discussing the situation in Nigeria at least three days before the Ebola outbreak was brought to public attention and seven days before the Centers for Disease Control issued an official alert. As a result, many researchers are now harnessing social media's potential to advance global awareness and improve health.

- **Internet of Things**

Millions of people use devices to monitor various aspects of their health behavior. These devices can monitor things such as heart rate, mobility, sleep quality, and blood sugar quality. The recorded data can be used to detect any danger and alert a physician, depending on the service offered by the device, all in real-time [29]. Due to advances in technology, particularly sensor technology, there is a growing interest in wearable and implantable sensors. These technological advances have made continuous and multimodal sensing possible. Simultaneously, advances in sensor miniaturization, noise reduction, and microelectronics development have increased the flexibility and reliability of implantable sensors [30].

### 2.3. Healthcare Big Data Analytics Classification

Several types of big data analytics are used in the healthcare industry (Figure 4), including descriptive analytics, diagnostic analytics, and predictive and prescriptive an-

alytics [18]. In this section, we discuss the specifics of each type of analysis and how it manifests itself in the healthcare field.



**Figure 4.** Classification of big data analytics in healthcare.

(a) Descriptive Analytics

Descriptive analytics consists of the description of the existing situation and helps to outline the picture of past performance on the basis of historical data and through the use of business intelligence and data mining. To perform this level of analysis, various techniques are used [31]. Descriptive analysis, known as unsupervised learning, among other things, summarizes what happens in the management of health services and what effect does a parameter have on the system? Descriptive analysis is the simplest level of understanding and use. It is a simple description of the data, with no further analysis, exploration, or analysis. The descriptive analysis defines, characterizes, aggregates, and classifies data in order to provide health practitioners with useful information for understanding and analyzing decisions, performance, and consequences. For example, this includes discharge rates, the average length of stay, and other relevant metrics for hospitals.

(b) Diagnostic Analytics

Diagnostic analytics seeks to explain why certain events occurred and what factors contributed to them. For example, diagnostic analytics attempts to understand the reasons behind the frequent readmissions of some patients [32] using various methods such as clustering and decision trees. To find the source of an issue and help people understand its nature and impact, an extensive examination and guided analysis of the existing data utilizing tools such as imaging techniques are required [33]. This may include the ability to understand the effects of system inputs and processes on performance. For instance, there are a number of significant factors, such as patient, provider, or organization-related issues, that may contribute to longer wait times for the provision of some healthcare services [34,35].

(c) Predictive Analytics

Predictive analytics reflect the ability to predict future events while assisting in the identification of trends and identifying potential uncertain outcomes; for example, it may be asked to predict whether or not a patient will develop complications. Predictive models are often constructed using machine learning techniques. Predictive analytics use massive data sets to improve customer experience, improving results compared to conventional business strategies [7]. They are used to analyze large volumes of data, as well as unstructured data, which produce the results to predict future developments. From an information science perspective, predicting future developments based on current data sets is a difficult



issue. Business intelligence programs of this kind help to calculate data streams on a larger scale, including social media content, shopping experiences, users' daily activities, and surveys [36]. For example, a pharmacist may need to know how much of a medicinal preparation to keep in stock in the inventory in anticipation of an outbreak of an epidemic. A doctor may also need to predict certain clinical events, such as the length of a patient's stay, the possibility that a patient will choose to undergo surgery, or the possibility that a patient will have complications or even die [4].

(d) Prescriptive Analytics

Decisions in the prescriptive analysis must be based on a wide range of practical alternatives, which can enable decision-makers in an organization to diagnose emerging opportunities or problems and recommend the best course of action to capitalize on the analysis provided in time while also taking into account the consequences and expected outcomes of decisions [37]. This analysis method automatically synthesizes Big Data and provides insights into a large number of possible outcomes before an analysis is performed. This information can be used by the decision-maker to support their actions. Prescriptive analytics give advice on what should be performed, what the best outcome will be, and how they can obtain it.

### 3. Artificial Intelligence in Medical Field

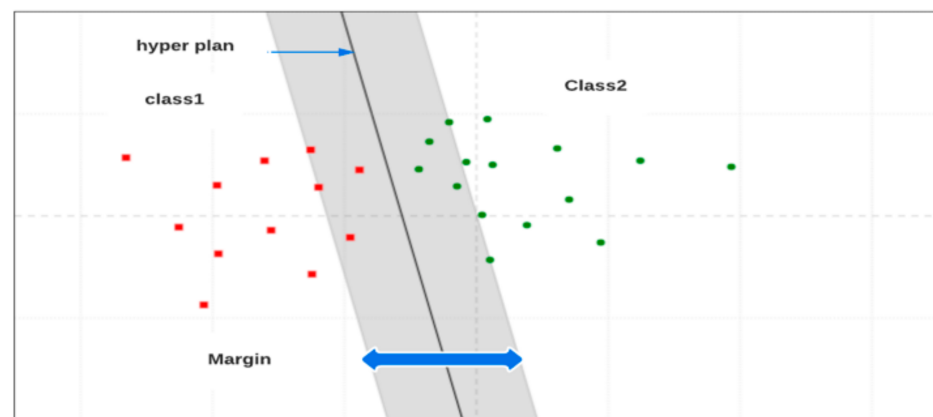
The use of artificial intelligence in medical research has the potential to lead to extremely sophisticated e-Health [38]. Machine learning (ML) is recognized as one of the most important scientific fields that can be integrated into the processes of diagnosis, prognosis, and even the treatment of diseases with the help of clinical decision support systems [39]. Another point about using machine learning techniques in healthcare is the elimination of human involvement to some degree, which reduces the likelihood of human error. This is particularly relevant when processing automation tasks; tedious routine work is where humans make the most errors [17]. In contrast, deep learning is a subfield of machine learning, which is a more sophisticated method that enables computers to automatically extract, analyze, and grasp relevant information from unstructured data by mimicking human thinking and learning [40]. Due to the volume of data generated for each patient, machine learning techniques have enormous potential in the healthcare field. The algorithms listed below are commonly used in health informatics.

- K-Nearest Neighbor Algorithm

We can define the k-nearest neighbor (k-NN) technique as a non-parametric algorithm, which means that the data set determines the model's structure. This is the reason why it is widely used; it does not rely on theoretical mathematical assumptions [41]. It also belongs to so-called "lazy" algorithms, which means that it does not need to learn or train all the data used in the prediction phase, and all the data can be used for the "test" phase. As a result, data learning is faster, and prediction is slower and more expensive and is thus more time and memory-consuming.

- Support Vector Machines (SVM)

Support vector machines, or SVMs, are a group of techniques used in classification and regression. They belong to a family of generalized linear classifiers. SVM is a practical method for classifying data. Typically, training and testing data for a classification task comprise certain data instances. Each instance in the training set includes a goal value and a number of other attributes. SVM classification is an example of fully supervised learning. Known labels aid in determining whether or not a system is on the right track [42]. According to [43], the SVM classifier has superior performance compared to other classifiers based on machine learning. Arrhythmic beat classification is used for anomaly detection in the electrocardiogram. The following figure (Figure 5) depicts a Support Vector Machine (SVM) model in two dimensions.



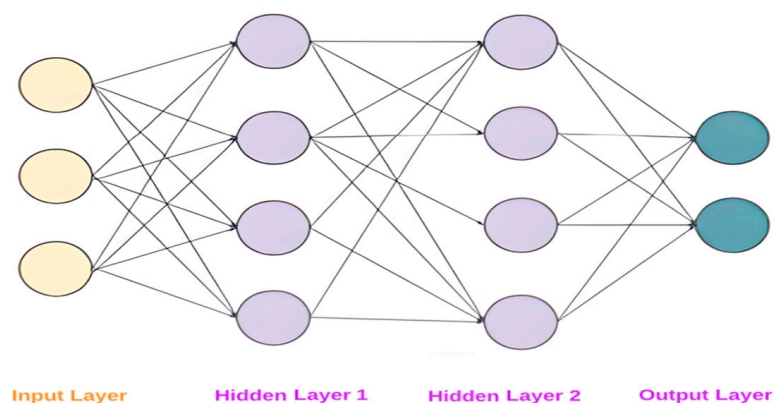
**Figure 5.** SVM model in two dimensions.

- **K-Means Clustering Techniques**

It has been demonstrated that data clustering is a useful technique for identifying structures in medical datasets. The k-means partitioning algorithm is one of the most popular and widely used clustering algorithms, and it belongs to a larger class of learning techniques that do not require unsupervised learning [44]. Clustering a dataset using k-means is simple. The fundamental idea is to find  $k$  centroids, one for each cluster, and link each element to the closest centroid, as long as the number ( $k$ ) of clusters (groups) to be formed is predetermined.

- **Artificial Neural Networks**

Artificial neural networks streamline representations of the brains of living things, particularly humans. Their functions and the structure of biological neural networks are similar to those of biological neurons in the brain. They attempt to combine the function of the human brain with a strictly abstract mathematical way of thinking, thus distinguishing artificial intelligence from biology and the classical function of computers [45]. Figure 6 depicts the fundamental structure of the algorithm.



**Figure 6.** A neural network's basic structure.

However, scientists, with the source inspired by the structure of the biological neuron, have managed to create an equivalent model of the so-called artificial neuron. A biological neuron receives input signals in the form of electrical impulses in its dendrites, processes them, and then transmits them to neighboring neurons via the axis and synapses. The primary goal of using artificial neural networks is to solve specific problems or to work autonomously in certain processes, such as image recognition. The issue of opacity in artificial neural networks is of critical concern, especially in safety-critical applications where the ability to comprehend and interpret decisions is paramount. Due to the black-box nature of neural networks, it can be challenging to identify potential sources of error

or bias, hindering our understanding of the underlying mechanisms behind decisions. While generating explanations or using more interpretable models have been proposed to address this issue, they may reduce accuracy or increase complexity. Therefore, it is essential for researchers and practitioners to weigh the trade-offs that are involved in using neural networks in safety-critical contexts and ensure that their use is justifiable and appropriately evaluated.

- Application of Machine Learning in Healthcare

There has been a considerable amount of research in recent publications to diagnose, predict or identify diseases. Nowadays, a variety of diseases are extensively diagnosed using different machine learning (ML) algorithms because of improvements in processing power and substantial studies on the subject [46]. The authors in [47] proposed a computational approach that relies on the SVM algorithm to predict Alzheimer's disease by utilizing gene and protein sequencing information. According to the obtained results in their research, the accuracy of their technique for Alzheimer's disease detection was 85.7%. U. Ahmed et al. [48] designed a framework consisting of two types of models: an SVM model and an ANN model. In order to predict if a patient has diabetes or not, these models examine the dataset to identify if a diabetes diagnosis is positive or negative. The prediction accuracy of their suggested fused technique was 94.87%. S. Thapa et al. [46] suggested a method for detecting Parkinson's disease patients based on feature selection and support vector machines. Based on the experiment's findings, TSVM can be a better classifier for a problem involving binary classifications such as Parkinson's disease delin-eation. To track the characteristics of brain tumors and improve detection efficiency, the authors in [49] developed a convolutional neural network-based model and MRI detection technology. This research model's main function is to segment and recognize MRIs: it employs a convolutional layer to improve recognition efficiency. Zheng et al. [50] used fusion k-means and SVMs to identify breast cancer. K-means were used in the experiment to identify the different hidden patterns of cancerous and benign tumors. H. K. van der Burgh et al. [51] merged clinical information from individuals with amyotrophic lateral sclerosis (a condition that results in the loss of neurons that regulate voluntary muscles) with MRI pictures. By using deep neural networks and this data, scientists were able to predict survivorship. M. Ghiasi et al. [52] designed a model dubbed the classification and regression tree (CART) model to detect coronary heart disease based on a decision tree learning algorithm. When compared to the reported targets, the results of the CART models showed the highest possible accuracy for coronary heart disease diagnosis (100%). D. Brinati et al. [53] created an interactive decision tree model to help clinicians identify COVID-19-positive patients using blood test analysis and machine learning instead of a PCR test. Their research demonstrated the feasibility and utility of using the latter two tools as an alternative to polymerase chain reaction (PCR) testing. While authors in [54] built a system based on the electronic medical record to help doctors categorize and prioritize patients in the emergency department, their system uses image data transformation as an input and a convolutional neural network algorithm as a classifier, to select patients who should go to the emergency department. The model presents a good performance of 0.86%.

In summary, Table 3 depicts in detail the applications of different machine learning techniques in healthcare analytics.

As shown in Table 3, machine learning methods can be used for a variety of applications, such as disease diagnosis, patient risk stratification, drug discovery, and resource optimization. The choice of algorithm depends on the specific use case and the type of data being analyzed. Some algorithms, such as logistic regression and decision trees, are well-suited for binary classification tasks, while others, such as clustering and neural networks, can be used for unsupervised learning and more complex tasks. While machine learning algorithms can be powerful tools for healthcare analysis, it is important to consider their limitations and potential biases. Machine learning algorithms should be validated and tested to ensure their accuracy and reliability in real-world healthcare settings.

**Table 3.** Recent uses of machine learning in the healthcare field.

Year	Application of Machine Learning in Medicine Context of Research	Technique Used
2020	Using various data balancing techniques to improve liver disease prediction [55]	Random Forest
2019	Predicting Opioid Use Disorder [56]	
2018	Prediction of osteoarthritis disease [57]	
2022	Prediction of Coronary Artery Disease and Acute Coronary Syndrome [58]	Ensemble Learning
2020	Neurological disease prediction [59]	
2020	Prediction of heart disease risk [60]	
2022	A Gene Prediction Function for Type 2 Diabetes Mellitus [61]	Logistic regression
2020	Prediction of graft dysfunction in pediatric liver transplantation [62]	
2020	Diabetes Progression Index Score Prediction [63]	
2022	Clinical Diagnosis of Alzheimer’s Disease [64]	Support vector Machines
2021	Classification of MRI images of brain tumors [65]	
2021	Early Alzheimer’s Disease Detection Using Blood Plasma Proteins [66]	
2019	Breast Cancer Detection Using the Decision Tree [67]	Decision trees
2019	Predicting breast cancer survivability [68]	
2020	Liver Diseases Prediction using KNN [69]	
2018	Lower Back Pain Classification [70]	K nearest Neighbors
2018	Heart disease diagnosis [71]	
2021	Prediction of cardiovascular disease via anomaly detection based on grouping [72]	
2021	Estimated number of confirmed COVID-19 cases [73]	K-means
2017	Dengue fever forecast [74]	
2022	Recognizing human activity from sensor data [75]	
2019	Classify lesions in optical tomographic images of breast masses [76]	Neural Networks
2017	Automatic identification of nasopharyngeal carcinoma [77]	

The capacity of researchers is greatly facilitated by open access to epidemiological, management, and clinical data in the health sector, which should help increase the volume of data and improve the quality of scientific research, as well as the scientific reach of institutions and the research community. In fact, the dominant trend in healthcare, which promises the most significant innovations, is that of data-driven patient care. Recording and collating all a patient’s information provides a more accurate picture of the care being performed and, in general, of population health management. It can also reduce inappropriate drug prescriptions and, in many cases, save lives.

#### 4. Big Data Technology Stack in Healthcare

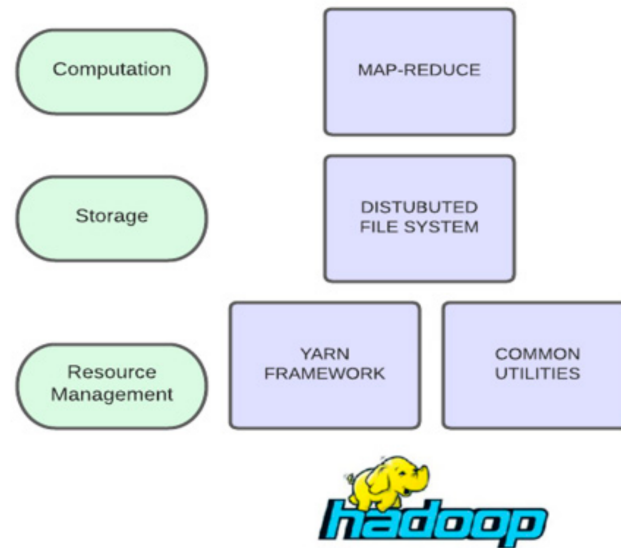
Once the fundamental issues regarding the use, collection, and management of big data in healthcare have been understood, it is appropriate to explore the tools provided by technology for data use. As is almost always the case in areas of software use, there is also in the use of big data the possibility of choosing between the use of open-source software and commercial solutions, which require the use of financial resources. The chosen platform must, in any case, manage data entry, processing, storage, and retrieval, as well as provide data analysis capabilities. This section presents the main options available.

##### 4.1. Infrastructure and Virtualization

To be able to store and process huge amounts of health data efficiently, hardware resources ranging from highly scalable storage systems to computing resources for data centers, and HPC systems are required. For this purpose, there are three subareas: cloud and grid solutions, data centers and HPC systems [78]. Cloud solutions provide the user with the illusion of virtually infinite computing and storage resources and thus allow companies and researchers to easily acquire them. Cloud solutions hide the details of the proposed hardware and rely on technologies for implementing large data centers. Data centers are needed for building cloud infrastructures as well as for in-house companies to provide computing and storage resources. For data centers, commodity hardware is primarily used to scale horizontally in a cost-effective manner.

#### 4.1.1. Apache Hadoop

Hadoop [79] is an open framework based on a distributed system that stores and processes very large computational clusters on core architecture and is reinforced by three primary elements, as shown in Figure 7.



**Figure 7.** The structure of apache Hadoop.

The core of Hadoop consists of a storage component called HDFS, a distributed file system, a processing component based on the map-reduce model, and a resource manager called YARN, “Yet Another Resource Negotiator” [80]. Hadoop splits the data into large blocks and distributes them among the diverse compute nodes that make up the computing system. It then transfers the code for execution to the nodes so that parallels, i.e., the simultaneous processing of data, can take place on those nodes [81]. In essence, the location property of the data is exploited, and nodes manage the individual data to which they have access. Finally, it should be noted that while the basic structure of Hadoop consists of the elements already mentioned, Apache extensions are often used to enrich Hadoop’s capabilities, depending on the situation, with the most important ones being Apache Spark, Apache Storm, Apache Flink, Apache Hive, Apache HBase, Apache Flume, Apache Sqoop, and Apache Pig. Hadoop could potentially be employed to develop medical analytics solutions. However, as previously stated, it is a batch big data platform that fails to fully capitalize on the potential of real-time emergencies [11].

#### 4.1.2. Apache Spark

Spark was created after Hadoop and provides the developer with an interface focused on a data structure known as a “Resilient Distributed Dataset” (RDD) that is intended to be a collection of objects distributed across a set of compute nodes, which provide efficient hardware failure management [82]. On the other hand, Spark provides the capacity to perform computations in shared memory, where the speed is significantly higher than on a disk. In this way, it becomes possible to implement iterative algorithms that access data multiple times on each iteration, without this coming at the expense of computational time, since the data access time in memory is faster and “closer” to the processor of the compute nodes. Apache Spark, in addition to the way it handles big data, offers the following key extensions [83]:

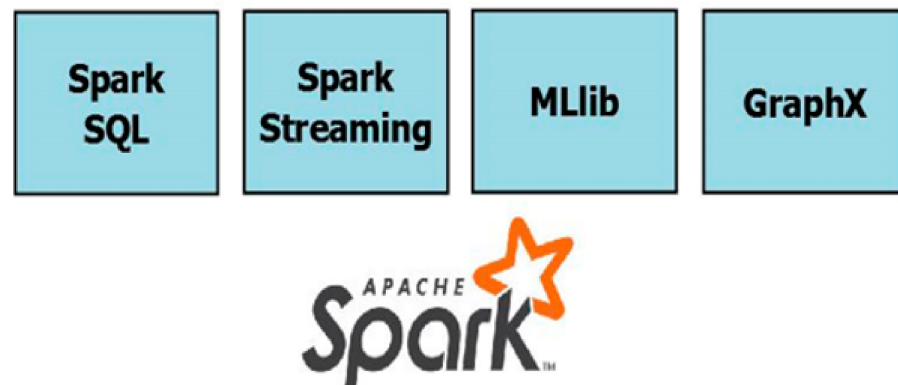
- Spark SQL: allows queries to be performed on data using SQL in conjunction with the Java, Scala, Python, and R programming languages.
- Spark Streaming: allows for the processing of streaming data, i.e., data that enters the system while calculations are already underway on the previous data. This



feature is very important because in Hadoop, new data cannot be preprocessed during processing, but the entire data set must be available when a MapReduce process is started. Java, Scala, and Python programming languages are supported.

- **MLlib:** This is a machine learning library that allows this type of algorithm to run up to 100 times faster than Hadoop.
- **GraphX:** Provides an API (application programming interface) for graphical data, allowing for productive computations using iterative algorithms.

Figure 8 shows the components of Apache Spark.



**Figure 8.** Apache Spark components.

Apache Spark can collect data from a variety of health data sources. It can handle large amounts of structured, semi-structured, and unstructured healthcare data, such as electronic health records (EHRs), diagnostic images, and genetic information. It can also preprocess, clean, and transform the data into a format suitable for analysis. Spark streaming components such as MLlib can be employed to analyze healthcare data in real time, which is produced by wearable health devices [84]. The data consists of crucial health metrics such as weight, blood pressure, respiratory rate, ECG, and blood glucose levels. By utilizing machine learning algorithms, the analysis can detect any potential critical health conditions before they manifest. For example, the authors in [85] created a health status prediction system in real-time for breast cancer by utilizing Spark streaming and machine learning. The system was designed to predict health status using machine learning models applied to streaming data. In the same context, the authors in [86] have suggested a heart disease monitoring system that utilizes the Spark framework for continuous and real-time monitoring. The system also employs the random forest algorithm with MLlib to build a prediction model for heart disease.

In summary, Apache Spark's ability to process, analyze, and integrate large amounts of healthcare data, combined with its machine learning and real-time capabilities, make it a valuable tool for addressing big data healthcare problems.

#### 4.2. The Use of NoSQL Databases

Relational databases, structured on the basis of the SQL language, have been the most popular data management method for many years among organizations and technology professionals. With the advent of big medical data, which are characterized by both its large size and diverse structure, there is a need to be able to process data on a large scale in order to draw consistent conclusions [87]. SQL-based systems cannot provide a stand-alone solution to the problem of managing these data. This problem can be solved by using NoSQL databases, which offer dynamic data management, flexibility, and scalability over relational databases. Their characteristics make them ideal for managing large, non-homogeneous data that are frequently updated and have frequently changed data field formats, in addition to the data itself [82]. The main NoSQL database options are MongoDB, Neo4j, CouchBase, Dynamo DB, HBase, and Cassandra. For healthcare



companies, the use of MongoDB has dominated over others. MongoDB is provided by 10Gen and can be effectively combined with the use of JSON (JavaScript Object Notation), XML, etc. According to the company itself, MongoDB is flexible, easy to use, and offers high performance, availability, and automatic scaling. Among other important features, it has the ability to perform text searches and connect to Hadoop. According to the official solution website, some indicative examples of solutions provided by MongoDB in the healthcare industry include:

- The creation of a complete patient profile that includes all tests performed on the patient and extracts useful relationships between them using data mining techniques. It is easy to modify and add new test data to the profile and compare old and new data.
- Early detection and containment of epidemics: big data has the potential to save human lives in situations where no other method can. Collecting data on emerging diseases, which have the potential to spread, widely, could be leveraged by applications that could serve as a tool for medical personnel and the extraction of risk indicators, such as the speed of spread, the number of people affected, symptoms, comparisons with data from previous epidemics, and suggesting the possibility of rapidly implementing population containment measures if necessary.
- The early diagnosis of rare diseases: it is possible to identify rare diseases that may have a common set of symptoms, but each of them or a subset of them is not a formidable indication. This observation is especially important because medical practitioners make their diagnoses primarily based on the experience and history of the patients they have examined in the past, which makes the process of early diagnosis of rare diseases extremely difficult, given the nature of human reasoning. Applications that have a large statistical dataset make it very easy to extract indicators to identify a disease and are an extremely useful tool for medical staff.
- Immediate consultation in real-time: in the case of laboratory data from patient tests, it is possible for medical personnel to draw immediate quantitative conclusions. As measurements from all types and sources of data can be visually displayed in single tables via graphs, there is no need for an independent review of individual tests by attending physicians.

#### 4.3. Commercial Platforms for Healthcare Data Analytics

A large number of databases that are available in the field of drug development create the need to identify priorities and methods for selecting appropriate information from a vast universe of big data. In this context, the Open PHACTS initiative implements the semantic weight-based search of research questions conducted in the context of pharmaceutical research [88]. The Open PHACTS program has a clear impact in several ways. The most important contribution is the use of the system in scientific research. Several scientific publications have resulted from the extensive use of this system, which allows for data analysis that has been very difficult to achieve in the past. Many pharmaceutical companies have integrated their internal data into Open PHACTS so that they can easily query all the information they have, whether public or private. Another contribution comes from the realization that large amounts of diverse semantic pharmaceutical data can be analyzed efficiently, thus improving data quality. The success of the Open PHACTS project has demonstrated the practicality of using data in biomedical research. Indeed, the fact that providers have chosen to offer their data reinforces the value of the action and helps sustain the Open PHACTS system.

Another very interesting project is the Artemis project. This project uses mining techniques, patented by McGregor, that are designed for non-trivial and possibly meaningful abstract information from huge datasets, where the digital data are generated by monitoring devices [89]. The analysis system employs abstraction techniques on the input data to identify recurrent patterns. It subsequently evaluates whether individuals with various health conditions, including infections, respiratory distress syndrome, and different forms of sleep apnea, exhibit similar data patterns in their normal state. The Artemis project

leverages three medical connectivity systems provided by Capsule Tech, ExcelMedical, and True Process clinical centers to continuously feed real-time data into a cloud-based database and analytics platform powered by IBM's InfoSphere and DB2 relational database [90].

IBM Watson is a complex computer system capable of answering questions in natural language. Medical personnel express in natural language the problem they are facing, describing symptoms and other relevant factors. Watson then performs an analysis and compiles a list of possible causes [91]. The sources of big data that Watson refers to can be physician and nurse notes and records, the electronic medical records of patients, clinical trials and research, scientific articles, as well as the information provided by patients themselves. Although it was developed and advertised as a diagnostic and treatment consultant, in reality, Watson was primarily used to treat patients who had already been diagnosed with a disease by suggesting ways to treat it [92].

The subsequent table presents a comparative analysis of various big data technologies applied in the context of healthcare.

The right technology for healthcare data analytics is determined by several factors, including the complexity and volume of the data, the system's required speed and scalability, the resources available, expert knowledge, and the defined targets and use cases. In general, open-source tools such as Hadoop and Spark provide a cost-effective and flexible solution for handling huge and varied healthcare datasets, as well as supporting various machine learning algorithms and techniques. They may, however, necessitate more technical skills and maintenance efforts than commercial tools. Commercial tools such as IBM Watson, Artemis, and Open PHACTS, on the other hand, often come with pre-built models and features that can accelerate the development and deployment of healthcare analytics applications, as well as provide more user-friendly interfaces and support services. They may, however, be more expensive and have fewer customization options. When selecting a technology for healthcare data analytics, healthcare professionals should carefully evaluate their specific needs and constraints, as well as factors such as data security, regulatory compliance, interoperability, and ethical considerations. It is also important to remember that technology selection is a continuous process that may necessitate continuous evaluation and optimization in response to changing needs and advances in the field. Table 4 represents a comparative analysis of various big data technologies in healthcare.

**Table 4.** Comparative analysis of various big data technologies in healthcare.

	<b>Hadoop</b>	<b>Spark</b>	<b>IBM Watson</b>	<b>Artemis</b>	<b>Open PHACTS</b>
Data Storage	HDFS (Hadoop Distributed File System)	RDDs (Resilient Distributed Datasets)	IBM Cloud Object Storage	MySQL, PostgreSQL, Oracle	Virtuoso Universal Server
Data Processing	Map Reduce, Hadoop Ecosystem (Pig, Hive, HBase, etc.)	Spark Core, Spark SQL, Spark Streaming, GraphX, MLlib	Watson Studio, SPSS Modeler	Cypher Query Language, RDF triplestores	SPARQL, RDF triplestores
Data Integration	Apache Nifi, Talend, Pentaho	Apache Nifi, Talend, Pentaho	IBM InfoSphere DataStage, Talend	ETL tools, REST APIs	ETL tools, REST APIs
Healthcare Applications	Clinical Decision Support, Drug Discovery, Genomics, Imaging Analytics	Predictive Analytics, Electronic Health Records Analysis, Clinical Decision Support	Drug Discovery, Genomics, Precision Medicine	Clinical Trials, Patient Data Management	Drug Discovery, Pharmacovigilance, Disease Networks

## 5. Technical and Organizational Challenges in Healthcare Big Data

The challenges that arise when using big data analytics technology are numerous and are particularly important to make the effort effective. The challenges are heterogeneous and diverse. The key points that a healthcare provider must consider in each case in this context are as follows:

(a) Data repositories

Although it has already been reported that the available health data are growing exponentially, the majority of it is in individual repositories: a phenomenon that has been called “data silos” [93]. These are essentially data repositories that are kept within an organization or even individual parts of organizations and are not accessible to the outside world. The lack of a common spirit of collaboration between organizations and internally between different departments inevitably hinder data sharing. It is, therefore, up to the institution concerned to ensure that this risk is avoided by developing the right spirit among employees, which is usually not a standard procedure.

(b) Data quality

Data quality refers to all the key characteristics that describe big data, as presented in detail in the previous section. In particular, there are four categories (4 vs.), and the following challenges are encountered:

- Volume: For efficient exploitation of data, the ability to manage and store the volume of data as well as determine their size must be ensured. Scalability is almost always required, as needs are constantly increasing, as is the volume to be exploited.
- Velocity: Any organization must consider the speed with which it can store, process, and use available data and continuously improve its performance, especially when the rate of data arrival is fast.
- Validity: Ensuring the validity of the data is critical to the project’s needs and is a demanding process.
- Variety: Identifying all data sources, the technical challenges that each source imposes, and managing them effectively is an integral part of any big data analysis effort and is a major challenge.

(c) Periodic data refresh

This is a purely technical issue, but one that can create difficulties if it is not respected. It is essentially about data management. In some cases, it is necessary to periodically delete or update data, and the systems available have specific capabilities. Therefore, there is a need to ensure that dynamic data management can be performed.

(d) Analytics challenges

Beyond the technical requirements, it is extremely important to “decrypt” the data, to understand it, and to develop analytical thinking and methods to create value. It is common for data to be misinterpreted by humans, leading to results that are not desired.

(e) Application of expertise

The needs of the health sector are constantly increasing through new research, observations, scientific articles, etc. However, at the same time, the technological capabilities that can help meet the needs are also increasing. Therefore, it is essential to be aware of technological developments and to intervene, if necessary, to overcome the inherent difficulties and extend the system’s functionality.

(f) Prediction models

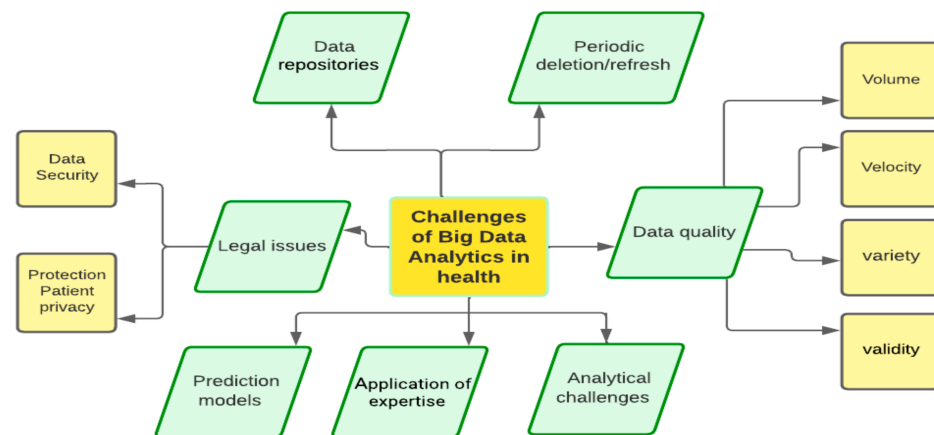
A key area of big data analytics is the generation of models that estimate and predict various situations. Specifically, in the healthcare industry, there is a need for the continuous study of data and the estimation of expected events to maximize the benefits and value of the data.

(g) Legal issues

There is a wide range of legal issues that need to be addressed, and it is necessary to keep abreast of developments in this area. System security must be ensured against unauthorized access to data by unauthorized persons. In this context, the challenges are

the same as in the area of security systems, which require a lot of time and effort. On the other hand, health information is extremely sensitive and should never be used to directly or indirectly identify individuals: a concept that is known as patient privacy. The challenge is even greater, especially in the absence of a stable and universally accepted framework.

The following figure, Figure 9, illustrates the key challenges in the healthcare sector regarding the use of big data.



**Figure 9.** Key challenges in the healthcare sector regarding the use of big data.

The need to develop tools and methods to meet all the issues raised by the use of big data in healthcare organizations requires a collective, organized, and rigorously defined effort.

## 6. Proposed Strategies for Implementing Big Data Analytics in Healthcare for Smart City

In the context of the Smart City concept, the integration of big data analytics in healthcare can play a critical role in improving the overall quality of life. Healthcare providers can gain a more comprehensive understanding of the community's health needs by leveraging the vast amounts of data that are generated by various sources such as wearable devices, electronic health records, and social media platforms. This can lead to more effective and targeted interventions in addressing health issues, as well as the development of proactive healthcare strategies to avoid illnesses in the first place. Furthermore, the use of big data analytics can aid in the optimization of healthcare resource allocation, lowering costs and increasing efficiency. As such, it is crucial for healthcare organizations and institutions to consider the Smart City context when developing and implementing big data analytics strategies in healthcare.

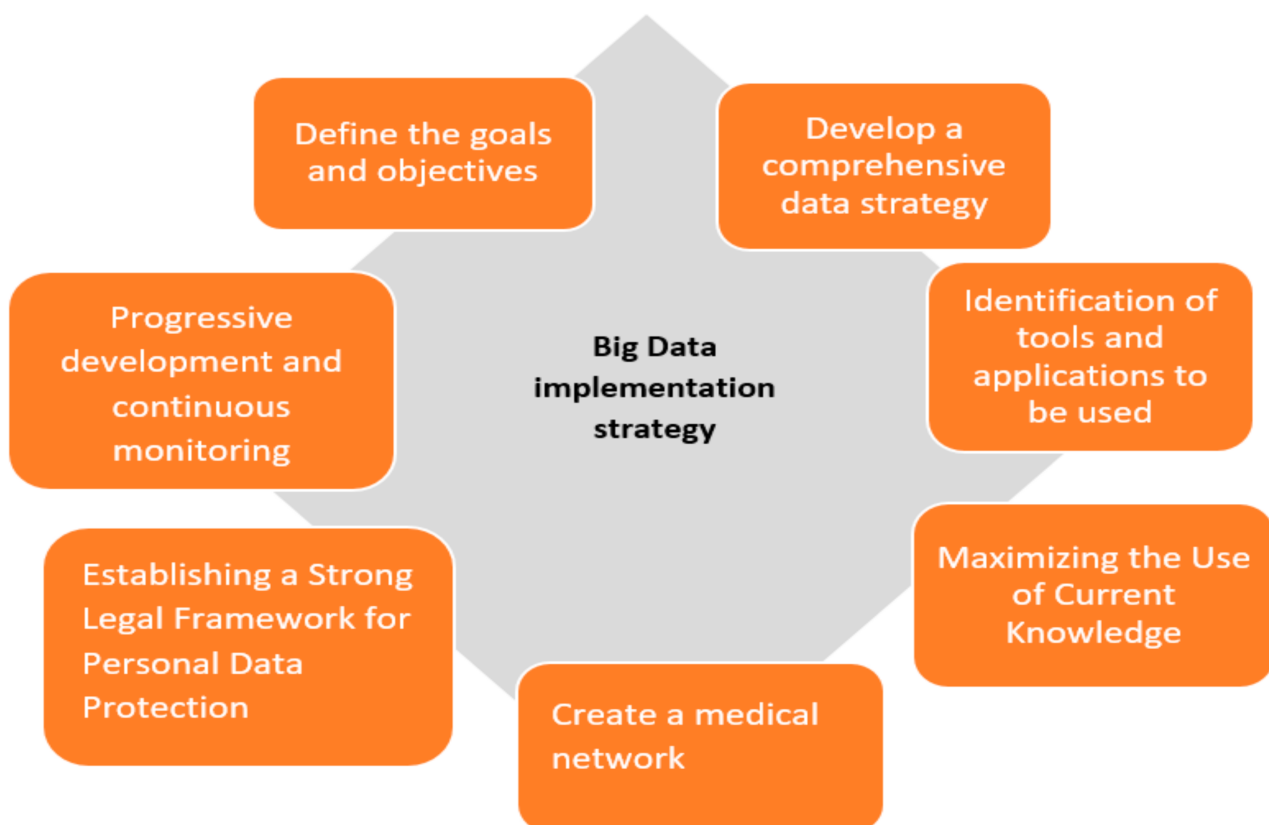
Based on the best practices in the field of big data analytics in healthcare, we provide a general framework for healthcare organizations to follow; by applying this simple strategy, health professionals can effectively leverage the potential of big data analytics to improve patient outcomes in their medical institutions.

- ✓ Define the goals and objectives: Clearly define the goals and objectives of the big data analytics initiative, such as improving patient outcomes, reducing healthcare costs, or enhancing the quality of care.
- ✓ Develop a comprehensive data strategy: Develop a comprehensive data strategy that outlines how the data will be collected, stored, processed, and analyzed to support the big data analytics initiative.
- ✓ The identification of tools and applications to be used: Invest in the right technology and infrastructure to support big data analytics, such as cloud computing, data warehouses, and data analytics tools. The effective use of big data technologies has many benefits, including the ability to measure the effectiveness and efficiency of interventions in real clinical practice. At the same time, it offers the possibility

of aggregating epidemiological, clinical, economic, and management data that can contribute to the generation of correlation information between the health of humans, economic resources, and health outcomes.

- ✓ **Maximizing the Use of Current Knowledge:** It is imperative to adopt a perspective that integrates and uses existing knowledge. This approach will enhance data comprehension, facilitate the systematic generation of novel insights, and foster a data-driven culture within the medical institution.
- ✓ **Create a medical network:** Collaborate with patients, healthcare providers, and researchers to ensure that the big data analytics project aligns with their needs and goals.
- ✓ **Establishing a Strong Legal Framework for Personal Data Protection:** Data protection, in particular, plays a key role in the successful implementation of big data. Particular attention must be paid to the processing of personal data, and it is important to take into account the legal framework conditions and technological possibilities for its implementation.
- ✓ **Progressive development and continuous monitoring:** A progressive integration can help better monitor and continually evaluate the big data analytics initiative to ensure that it is delivering value and positively impacting patient care.

In the following figure (Figure 10), the suggested strategy for implementing big data in the healthcare industry is summarized.



**Figure 10.** Suggested strategy for implementing big data in the healthcare industry.

## 7. Conclusions

There is no doubt that financial and human resources will be invested in the near future to improve health services through big data analytics. The number of problems solved through their use is enormous, and at present, there does not seem to be an alternative technology with comparable potential. For this reason, it is certain that the use of data on a large scale will concern not only “large” institutions and organizations in the future but that each clinic and doctor will have to use the technological tools available to them



in order to provide health services. This is optimal because large sums of money are wasted unnecessarily, either due to inefficient management resulting from poor handling or incorrect treatment and diagnosis. More importantly, the human factor, i.e., the radical upgrade of health services that can usher in a new era, is the most important reason to dispel any doubts about the proliferation of big data analysis in the future. This paper demonstrates the abundance of opportunities to deliver more targeted, large-scale, and cost-effective healthcare by leveraging the available data and big data analytics. However, the healthcare sector has been shown to have specific characteristics and challenges that require additional research efforts in order to fully benefit from the opportunities. In our next work, we will propose a methodology to develop big data analysis in the health field and design a new flexible architecture that meets the challenges mentioned in this review.

**Author Contributions:** Methodology and Conceptualization, N.B.; validation and review, F.E.M. and Y.E.B.E.I.; writing—original draft preparation, N.B. and Y.F. 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. Manogaran, G.; Thota, C.; Lopez, D.; Vijayakumar, V.; Abbas, K.M.; Sundarsekar, R. Big Data Knowledge System in Healthcare. In *Internet of Things and Big Data Technologies for Next Generation Healthcare*; Bhatt, C., Dey, N., Ashour, A.S., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 133–157. [\[CrossRef\]](#)
2. Munawar, H.S.; Qayyum, S.; Ullah, F.; Sepasgozar, S. Big Data and Its Applications in Smart Real Estate and the Disaster Management Life Cycle: A Systematic Analysis. *Big Data Cogn. Comput.* **2020**, *4*, 4. [\[CrossRef\]](#)
3. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: Promise and potential. *Health Inf. Sci. Syst.* **2014**, *2*, 3. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Bhaskaran, K.L.; Osei, R.S.; Kotei, E.; Agbezuge, E.Y.; Ankora, C.; Ganaa, E.D. A Survey on Big Data in Pharmacology, Toxicology and Pharmaceuticals. *Big Data Cogn. Comput.* **2022**, *6*, 161. [\[CrossRef\]](#)
5. Dash, S.; Shakyawar, S.K.; Sharma, M.; Kaushik, S. Big data in healthcare: Management, analysis and future prospects. *J. Big Data* **2019**, *6*, 54. [\[CrossRef\]](#)
6. Strickland, N.H. PACS (picture archiving and communication systems): Filmless radiology. *Arch. Dis. Child.* **2000**, *83*, 82–86. [\[CrossRef\]](#)
7. Janke, A.T.; Overbeek, D.L.; Kocher, K.E.; Levy, P.D. Exploring the Potential of Predictive Analytics and Big Data in Emergency Care. *Ann. Emerg. Med.* **2016**, *67*, 227–236. [\[CrossRef\]](#)
8. Batko, K.; Ślęzak, A. The use of Big Data Analytics in healthcare. *J. Big Data* **2022**, *9*, 3. [\[CrossRef\]](#)
9. Wang, L.; Alexander, C.A. Big Data in Medical Applications and Health Care. *Curr. Res. Med.* **2015**, *6*, 1–8. [\[CrossRef\]](#)
10. Tresp, V.; Overhage, J.M.; Bundschuh, M.; Rabizadeh, S.; Fasching, P.A.; Yu, S. Going Digital: A Survey on Digitalization and Large-Scale Data Analytics in Healthcare. *Proc. IEEE* **2016**, *104*, 2180–2206. [\[CrossRef\]](#)
11. Harerimana, G.; Jang, B.; Kim, J.W.; Park, H.K. Health Big Data Analytics: A Technology Survey. *IEEE Access* **2018**, *6*, 65661–65678. [\[CrossRef\]](#)
12. Bahri, S.; Zoghalmi, N.; Abed, M.; Tavares, J.M.R.S. BIG DATA for Healthcare: A Survey. *IEEE Access* **2019**, *7*, 7397–7408. [\[CrossRef\]](#)
13. Dhayne, H.; Haque, R.; Kilany, R.; Taher, Y. In Search of Big Medical Data Integration Solutions—A Comprehensive Survey. *IEEE Access* **2019**, *7*, 91265–91290. [\[CrossRef\]](#)
14. Wang, L.; Alexander, C.A. Big data analytics in medical engineering and healthcare: Methods, advances and challenges. *J. Med. Eng. Technol.* **2020**, *44*, 267–283. [\[CrossRef\]](#)
15. Shafqat, S.; Kishwer, S.; Rasool, R.U.; Qadir, J.; Amjad, T.; Ahmad, H.F. Big data analytics enhanced healthcare systems: A review. *J. Supercomput.* **2020**, *76*, 1754–1799. [\[CrossRef\]](#)
16. Imran, S.; Mahmood, T.; Morshed, A.; Sellis, T. Big data analytics in healthcare A systematic literature review and roadmap for practical implementation. *IEEE/CAA J. Autom. Sinica* **2021**, *8*, 1–22. [\[CrossRef\]](#)
17. Chattu, V.K. A Review of Artificial Intelligence, Big Data, and Blockchain Technology Applications in Medicine and Global Health. *Big Data Cogn. Comput.* **2021**, *5*, 41. [\[CrossRef\]](#)
18. Al-Sai, Z.A.; Husin, M.H.; Syed-Mohamad, S.M.; Abidin, R.M.S.; Damer, N.; Abualigah, L.; Gandomi, A.H. Explore Big Data Analytics Applications and Opportunities: A Review. *Big Data Cogn. Comput.* **2022**, *6*, 157. [\[CrossRef\]](#)



19. Zhou, H. Developing Natural Language Processing to Extract Complementary and Integrative Health Information from Electronic Health Record Data. In Proceedings of the 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), Rochester, MN, USA, 11–14 June 2022; pp. 474–475. [\[CrossRef\]](#)
20. Piedrahita-Valdés, H.; Piedrahita-Castillo, D.; Bermejo-Higuera, J.; Guillem-Saiz, P.; Bermejo-Higuera, J.R.; Guillem-Saiz, J.; Sicilia-Montalvo, J.A.; Machío-Regidor, F. Vaccine Hesitancy on Social Media: Sentiment Analysis from June 2011 to April 2019. *Vaccines* **2021**, *9*, 28. [\[CrossRef\]](#)
21. Khaloufi, H.; Abouelmehdi, K.; Beni-hssane, A.; Saadi, M. Security model for Big Healthcare Data Lifecycle. *Procedia Comput. Sci.* **2018**, *141*, 294–301. [\[CrossRef\]](#)
22. Kumar, D.R.; Rajkumar, K.; Lalitha, K.; Dhanakoti, V. Bigdata in the Management of Diabetes Mellitus Treatment. In *Internet of Things for Healthcare Technologies 73*; Chakraborty, C., Banerjee, A., Kolekar, M.H., Garg, L., Chakraborty, B., Eds.; Springer Singapore: Singapore, 2021; pp. 293–324. [\[CrossRef\]](#)
23. Alfred, R.; Obit, J.H. The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review. *Heliyon* **2021**, *7*, e07371. [\[CrossRef\]](#)
24. Wang, H.; Cui, Z.; Chen, Y.; Avidan, M.; Abdallah, A.B.; Kronzer, A. Predicting Hospital Readmission via Cost-Sensitive Deep Learning. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2018**, *15*, 1968–1978. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Leff, D.R.; Yang, G.-Z. Big Data for Precision Medicine. *Engineering* **2015**, *1*, 277–279. [\[CrossRef\]](#)
26. Weitzman, E.R.; Kelemen, S.; Mandl, K.D. Surveillance of an Online Social Network to Assess Population-level Diabetes Health Status and Healthcare Quality. *Online J. Public Health Inform.* **2011**, *3*, ojphi.v3i3.3797. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Ram, S.; Zhang, W.; Williams, M.; Pengetnze, Y. Predicting Asthma-Related Emergency Department Visits Using Big Data. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1216–1223. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Odum, M.; Yoon, S. What can we learn about the Ebola outbreak from tweets? *Am. J. Infect. Control* **2015**, *43*, 563–571. [\[CrossRef\]](#)
29. Ma, Y.; Wang, Y.; Yang, J.; Miao, Y.; Li, W. Big Health Application System based on Health Internet of Things and Big Data. *IEEE Access* **2017**, *5*, 7885–7897. [\[CrossRef\]](#)
30. Darwish, A.; Sayed, G.I.; Hassanien, A.E. The Impact of Implantable Sensors in Biomedical Technology on the Future of Healthcare Systems. In *Intelligent Pervasive Computing Systems for Smarter Healthcare*, 1st ed.; Sangaiah, A.K., Shantharajah, S., Theagarajan, P., Eds.; Wiley: Hoboken, NJ, USA, 2019; pp. 67–89. [\[CrossRef\]](#)
31. Islam, M.S.; Hasan, M.M.; Wang, X.; Germack, H.D.; Noor-E-Alam, M. A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. *Healthcare* **2018**, *6*, 54. [\[CrossRef\]](#)
32. Fisahn, C.; Sanders, F.H.; Moisi, M.; Page, J.; Oakes, P.C.; Wingerson, M.; Dettori, J.; Tubbs, R.S.; Chamiraju, P.; Nora, P.; et al. Descriptive analysis of unplanned readmission and reoperation rates after intradural spinal tumor resection. *J. Clin. Neurosci.* **2017**, *38*, 32–36. [\[CrossRef\]](#)
33. Yu, Y.; Li, M.; Liu, L.; Li, Y.; Wang, J. Clinical big data and deep learning: Applications, challenges, and future outlooks. *Big Data Min. Anal.* **2019**, *2*, 288–305. [\[CrossRef\]](#)
34. Simpao, A.F.; Ahumada, L.M.; Gálvez, J.A.; Rehman, M.A. A review of analytics and clinical informatics in health care. *J. Med. Syst.* **2014**, *38*, 45. [\[CrossRef\]](#)
35. Khalifa, M. Health Analytics Types, Functions and Levels: A Review of Literature. *Stud. Health Technol. Inform.* **2018**, *251*, 137–140.
36. Alharthi, H. Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *J. Infect. Public Health* **2018**, *11*, 749–756. [\[CrossRef\]](#)
37. Mosavi, N.S.; Santos, M.F. How Prescriptive Analytics Influences Decision Making in Precision Medicine. *Procedia Comput. Sci.* **2020**, *177*, 528–533. [\[CrossRef\]](#)
38. Dicuonzo, G.; Galeone, G.; Shini, M.; Massari, A. Towards the Use of Big Data in Healthcare: A Literature Review. *Healthcare* **2022**, *10*, 1232. [\[CrossRef\]](#)
39. Khan, P.; Kader, F.; Islam, S.M.R.; Rahman, A.B.; Kamal, S.; Toha, M.U.; Kwak, K.-S. Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances. *IEEE Access* **2021**, *9*, 37622–37655. [\[CrossRef\]](#)
40. Chauhan, N.K.; Singh, K. A Review on Conventional Machine Learning vs. Deep Learning. In Proceedings of the 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 28–29 September 2018; pp. 347–352. [\[CrossRef\]](#)
41. Xing, W.; Bei, Y. Medical Health Big Data Classification Based on KNN Classification Algorithm. *IEEE Access* **2020**, *8*, 28808–28819. [\[CrossRef\]](#)
42. Wang, L.; Liu, J.; Zhu, S.; Gao, Y. Prediction of Linear B-Cell Epitopes Using AAT Scale. In Proceedings of the 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, Beijing, China, 11–13 June 2009; pp. 1–4. [\[CrossRef\]](#)
43. Venkatesan, C.; Karthigaikumar, P.; Paul, A.; Satheeskumaran, S.; Kumar, R. ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications. *IEEE Access* **2018**, *6*, 9767–9773. [\[CrossRef\]](#)
44. Li, Y.; Wu, H. A Clustering Method Based on K-Means Algorithm. *Phys. Procedia* **2012**, *25*, 1104–1109. [\[CrossRef\]](#)
45. Hasson, U.; Nastase, S.A.; Goldstein, A. Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron* **2020**, *105*, 416–434. [\[CrossRef\]](#)
46. Thapa, S.; Adhikari, S.; Ghimire, A.; Aditya, A. Feature Selection Based Twin-Support Vector Machine for the Diagnosis of Parkinson's Disease. In Proceedings of the 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC), Kuching, Malaysia, 1–3 December 2020; pp. 1–6. [\[CrossRef\]](#)

47. Xu, L.; Liang, G.; Liao, C.; Chen, G.-D.; Chang, C.-C. An Efficient Classifier for Alzheimer's Disease Genes Identification. *Molecules* **2018**, *23*, 3140. [\[CrossRef\]](#)
48. Ahmed, U.; Issa, G.F.; Khan, M.A.; Aftab, S.; Said, R.A.T.; Ghazal, T.M.; Ahmad, M. Prediction of Diabetes Empowered with Fused Machine Learning. *IEEE Access* **2022**, *10*, 8529–8538. [\[CrossRef\]](#)
49. Wang, W.; Bu, F.; Lin, Z.; Zhai, S. Learning Methods of Convolutional Neural Network Combined with Image Feature Extraction in Brain Tumor Detection. *IEEE Access* **2020**, *8*, 152659–152668. [\[CrossRef\]](#)
50. Zheng, B.; Yoon, S.W.; Lam, S.S. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst. Appl.* **2014**, *41*, 1476–1482. [\[CrossRef\]](#)
51. van der Burgh, H.K.; Schmidt, R.; Westeneng, H.-J.; de Reus, M.A.; van den Berg, L.H.; van den Heuvel, M.P. Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *Neuroimage Clin.* **2016**, *13*, 361–369. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Ghiasi, M.M.; Zendehboudi, S.; Mohsenipour, A.A. Decision tree-based diagnosis of coronary artery disease: CART model. *Comput. Methods Programs Biomed.* **2020**, *192*, 105400. [\[CrossRef\]](#)
53. Brinati, D.; Campagner, A.; Ferrari, D.; Locatelli, M.; Banfi, G.; Cabitza, F. Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J. Med. Syst.* **2020**, *44*, 135. [\[CrossRef\]](#)
54. Yao, L.-H.; Leung, K.-C.; Hong, J.-H.; Tsai, C.-L.; Fu, L.-C. A System for Predicting Hospital Admission at Emergency Department Based on Electronic Health Record Using Convolution Neural Network. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 11–14 October 2020; pp. 546–551. [\[CrossRef\]](#)
55. Ambesange, S.; Vijayalaxmi, A.; Uppin, R.; Patil, S.; Patil, V. Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques. In Proceedings of the 2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bengaluru, India, 6–7 November 2020; pp. 98–102. [\[CrossRef\]](#)
56. Wadekar, A. Predicting Opioid Use Disorder (OUD) Using A Random Forest. In Proceedings of the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), Milwaukee, WI, USA, 15–19 July 2019; Volume 1, pp. 960–961. [\[CrossRef\]](#)
57. Aprilliani, U.; Rustam, Z. Osteoarthritis Disease Prediction Based on Random Forest. In Proceedings of the 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Yogyakarta, Indonesia, 27–28 October 2018; pp. 237–240. [\[CrossRef\]](#)
58. Jamthikar, A.D.; Gupta, D.; Mantella, L.E.; Saba, L.; Johri, A.M.; Suri, J.S. Ensemble Machine Learning and Its Validation for Prediction of Coronary Artery Disease and Acute Coronary Syndrome Using Focused Carotid Ultrasound. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2503810. [\[CrossRef\]](#)
59. Kaur, S.; Aggarwal, H.; Rani, R. Neurological disease prediction using ensembled Machine Learning Model. In Proceedings of the 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Wanknaghat, India, 6–8 November 2020; pp. 410–414. [\[CrossRef\]](#)
60. Mienye, I.D.; Sun, Y.; Wang, Z. An improved ensemble learning approach for the prediction of heart disease risk. *Inform. Med. Unlocked* **2020**, *20*, 100402. [\[CrossRef\]](#)
61. Alshamlan, H.; Taleb, H.B.; Al Sahow, A. A Gene Prediction Function for Type 2 Diabetes Mellitus using Logistic Regression. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020; pp. 1–4. [\[CrossRef\]](#)
62. Prodanova, K.; Uzunova, Y. Prediction of Graft Dysfunction in Pediatric Liver Transplantation by Logistic Regression. In Proceedings of the 2020 International Conference on Mathematics and Computers in Science and Engineering (MACISE), Madrid, Spain, 14–16 January 2020; pp. 260–263. [\[CrossRef\]](#)
63. Lei, L. Prediction of Score of Diabetes Progression Index Based on Logistic Regression Algorithm. In Proceedings of the 2020 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Zhangjiajie, China, 18–19 July 2020; pp. 954–956. [\[CrossRef\]](#)
64. Bhagyashree, S.R. Clinical Diagnosis of Alzheimer's Disease Employing Support Vector Machine. In Proceedings of the 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 23–24 April 2022; pp. 1–5. [\[CrossRef\]](#)
65. Shahajad, M.; Gambhir, D.; Gandhi, R. Features extraction for classification of brain tumor MRI images using support vector machine. In Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 28–29 January 2021; pp. 767–772. [\[CrossRef\]](#)
66. Eke, C.S.; Jammeh, E.; Li, X.; Carroll, C.; Pearson, S.; Ifeakor, E. Early Detection of Alzheimer's Disease with Blood Plasma Proteins Using Support Vector Machines. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 218–226. [\[CrossRef\]](#)
67. Sathiyarayanan, P.; Pavithra, S.; Saranya, M.S.A.; Makeswari, M. Identification of Breast Cancer Using The Decision Tree Algorithm. In Proceedings of the 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 29–30 March 2019; pp. 1–6. [\[CrossRef\]](#)
68. Fu, B.; Liu, P.; Lin, J.; Deng, L.; Hu, K.; Zheng, H. Predicting Invasive Disease-Free Survival for Early Stage Breast Cancer Patients Using Follow-Up Clinical Data. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 2053–2064. [\[CrossRef\]](#)
69. Ambesange, S.; Nadagoudar, R.; Uppin, R.; Patil, V.; Patil, S.; Patil, S. Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques. In Proceedings of the 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), Vijayapur, India, 8–10 October 2020; pp. 1–6. [\[CrossRef\]](#)

70. Sandag, G.A.; Tedry, N.E.; Lolong, S. Classification of Lower Back Pain Using K-Nearest Neighbor Algorithm. In Proceedings of the 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 7–9 August 2018; pp. 1–5. [\[CrossRef\]](#)
71. Pawlovsky, A.P. An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis. In Proceedings of the 2018 International Conference on Electronics, Information, and Communication (ICEIC), Honolulu, HI, USA, 24–27 January 2018; pp. 1–4. [\[CrossRef\]](#)
72. Ripan, R.C.; Sarker, I.H.; Hossain, S.M.M.; Anwar, M.; Nowrozy, R.; Hoque, M.M.; Furhad, H. A Data-Driven Heart Disease Prediction Model Through K-Means Clustering-Based Anomaly Detection. *SN Comput. Sci.* **2021**, *2*, 112. [\[CrossRef\]](#)
73. Vadyala, S.R.; Betgeri, S.N.; Sherer, E.A.; Amritphale, A. Prediction of the number of COVID-19 confirmed cases based on K-means-LSTM. *Array* **2021**, *11*, 100085. [\[CrossRef\]](#)
74. Manivannan, P.; Devi, P.I. Dengue Fever Prediction Using K-Means Clustering Algorithm. In Proceedings of the 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputtur, India, 23–25 March 2017; pp. 1–5. [\[CrossRef\]](#)
75. Sarkar, A.; Hossain, S.K.S.; Sarkar, R. Human activity recognition from sensor data using spatial attention-aided CNN with genetic algorithm. *Neural Comput. Appl.* **2022**, *35*, 5165–5191. [\[CrossRef\]](#)
76. Xu, Q.; Wang, X.; Jiang, H. Convolutional neural network for breast cancer diagnosis using diffuse optical tomography. *Visual Computing for Industry, Biomed. Art* **2019**, *2*, 6. [\[CrossRef\]](#)
77. Mohammed, M.A.; Ghani, M.K.A.; Hamed, R.I.; Ibrahim, D.A.; Abdullah, M.K. Artificial neural networks for automatic segmentation and identification of nasopharyngeal carcinoma. *J. Comput. Sci.* **2017**, *21*, 263–274. [\[CrossRef\]](#)
78. Costa, F.F. Big data in biomedicine. *Drug Discov. Today* **2014**, *19*, 433–440. [\[CrossRef\]](#)
79. Landset, S.; Khoshgoftaar, T.M.; Richter, A.N.; Hasanin, T. A survey of open source tools for machine learning with big data in the Hadoop ecosystem. *J. Big Data* **2015**, *2*, 24. [\[CrossRef\]](#)
80. Shvachko, K.; Kuang, H.; Radia, S.; Chansler, R. The Hadoop Distributed File System. In Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 3–7 May 2010; pp. 1–10. [\[CrossRef\]](#)
81. Azeroual, O.; Fabre, R. Processing Big Data with Apache Hadoop in the Current Challenging Era of COVID-19. *Big Data Cogn. Comput.* **2021**, *5*, 12. [\[CrossRef\]](#)
82. Fu, J.; Sun, J.; Wang, K. SPARK—A Big Data Processing Platform for Machine Learning. In Proceedings of the 2016 International Conference on Industrial Informatics—Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan, China, 3–4 December 2016; pp. 48–51. [\[CrossRef\]](#)
83. Salloom, S.; Dautov, R.; Chen, X.; Peng, P.X.; Huang, J.Z. Big data analytics on Apache Spark. *Int. J. Data Sci. Anal.* **2016**, *1*, 145–164. [\[CrossRef\]](#)
84. Liu, W.; Li, Q.; Cai, Y.; Li, Y.; Li, X. A Prototype of Healthcare Big Data Processing System Based on Spark. In Proceedings of the 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), Shenyang, China, 14–16 October 2015; pp. 516–520. [\[CrossRef\]](#)
85. Ed-daoudy, A.; Maalmi, K. Application of Machine Learning Model on Streaming Health Data Event in Real-Time to Predict Health Status Using Spark. In Proceedings of the 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), Rabat, Morocco, 21–23 November 2018; pp. 1–4. [\[CrossRef\]](#)
86. Ed-Daoudy, A.; Maalmi, K. Real-Time Machine Learning for Early Detection of Heart Disease Using Big Data Approach. In Proceedings of the 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 3–4 April 2019; pp. 1–5. [\[CrossRef\]](#)
87. Chandra, D.G. BASE analysis of NoSQL database. *Future Gener. Comput. Syst.* **2015**, *52*, 13–21. [\[CrossRef\]](#)
88. Williams, A.J.; Harland, L.; Groth, P.; Pettifer, S.; Chichester, C.; Willighagen, E.L.; Evelo, C.T.; Blomberg, N.; Ecker, G.; Goble, C.; et al. Open PHACTS: Semantic interoperability for drug discovery. *Drug Discov. Today* **2012**, *17*, 1188–1198. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Khazaei, H.; Mench-Bressan, N.; McGregor, C.; Pugh, J.E. Health Informatics for Neonatal Intensive Care Units: An Analytical Modeling Perspective. *IEEE J. Transl. Eng. Health Med.* **2015**, *3*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
90. McGregor, C.; Inibhunu, C.; Glass, J.; Doyle, I.; Gates, A.; Madill, J.; Pugh, J.E. Health Analytics as a Service with Artemis Cloud: Service Availability. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 5644–5648. [\[CrossRef\]](#)
91. Salvi, E.; Parimbelli, E.; Basadonne, A.; Viani, N.; Cavallini, A.; Micieli, G.; Quaglini, S.; Sacchi, L. Exploring IBM Watson to Extract Meaningful Information from the List of References of a Clinical Practice Guideline. In *Artificial Intelligence in Medicine*; Springer: Cham, Switzerland, 2017; pp. 193–197. [\[CrossRef\]](#)
92. Contractor, D.; Telang, A. (Eds.) *Applications of Cognitive Computing Systems and IBM Watson*; Springer: Singapore, 2017. [\[CrossRef\]](#)
93. Koutkias, V. From Data Silos to Standardized, Linked, and FAIR Data for Pharmacovigilance: Current Advances and Challenges with Observational Healthcare Data. *Drug Saf.* **2019**, *42*, 583–586. [\[CrossRef\]](#) [\[PubMed\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.