STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, January 2025, pp 378–395.
Published online in International Academic Press (www.IAPress.org)

Big Data in the Revolution of Medical Data: A Review

Amal Azeroual [1,*], Benayad Nsiri [1], Rachid Oulad Haj Thami [2], Taoufiq Belhoussine Drissi [3]

[1] *STIS, M2CS, National Graduate School of Arts and Crafts of Rabat (ENSAM), Biomedical Engineering, Mohammed V University in Rabat, Rabat, Morocco*
[2] *National School of Computer Science and Systems Analysis (ENSIAS), Mohammed V University in Rabat, Rabat, Morocco*
[3] *GEITIIL, Faculty of Science Ain Chock, University Hassan II, Casablanca, Morocco*

Abstract Big Data plays a crucial role in the medical sector, fundamentally transforming medical data collection, organization, and interpretation. This shift significantly enhances healthcare quality, propels medical research, and improves healthcare system effectiveness. Medical Big Data comprises a vast and diverse array of health-related information, generated at an unprecedented scale and speed, including electronic health records, medical imaging, genomic data, clinical trials, and data from wearable devices. Analyzing this data can reveal vital insights into disease patterns, treatment effectiveness, and population health trends, thereby aiding in creating personalized medicine, predictive analysis, and innovative healthcare solutions. Effective utilization of Medical Big Data requires advanced computational and analytical methods to extract meaningful insights, thereby fueling progress in healthcare and medical research. This review aims to provide specialists with a comprehensive overview of Big Data's application in diagnostic and medical domains, including its current usage in healthcare. We particularly focus on how integrating Big Data with artificial intelligence has led to more accurate predictive models for disease outbreaks and patient health risks, enhancing preventive care strategies. Furthermore, our analysis indicates that Big Data-driven personalization of treatment has significantly improved adherence to therapies and health outcomes in chronic disease management.

Keywords Big Data, Healthcare, Medical data, Artificial intelligence, Medical image

DOI: 10.19139/soic-2310-5070-2054

1. Introduction

Big Data initially emerged in sectors like business, healthcare, and marketing, but its impact has since expanded significantly. Now, it aids in predicting disease outbreaks, managing retail inventory, forecasting equipment failures in manufacturing, improving network performance, optimizing urban planning, personalizing media content, and enhancing education [1]. The concept of Big Data, which first emerged in the 1990s, refers to datasets too complex to be managed and analyzed with traditional computer software [2]. In the realm of healthcare, this concept has gained substantial importance due to the significant increase in medical data generated in recent decades. Big Data in healthcare encompasses a diverse array of data sources, technologies, and methodologies, forming datasets that often exceed the processing capabilities of standard algorithms [3]. The application of Big Data has become increasingly prevalent across various health domains, greatly benefiting biomedical research. This is evident in areas such as medical research, disease prediction, pathology diagnosis, and patient monitoring, where the utilization of medical Big Data has become an invaluable asset. The critical role of this innovation was further underscored during the COVID-19 pandemic, where it facilitated enhanced patient care through

*Correspondence to: Amal Azeroual (Email: amal.azeroual@um5r.ac.ma). Department Research Center STIS, M2CS, National Graduate School of Arts and Crafts of Rabat (ENSAM), Biomedical Engineering, Mohammed V University in Rabat, Rabat, Morocco ENSAM RABAT, B.P. : 6207, Avenue des Forces Armées Royales, Rabat 10100.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

more streamlined and reliable diagnostic processes. The integration of Big Data into medical specialties such as cardiology, oncology, and neurology enables significant advancements in the diagnosis and treatment of diseases. For example, in cardiology, massive data analyses from cardiac monitoring devices can predict heart attacks with increased accuracy. In oncology, Big Data helps better understand the genetic profiles of cancers, leading to personalized treatments and targeted therapies. Similarly, in neurology, the analysis of brain imaging data and clinical studies can improve the early detection of neurodegenerative diseases like Alzheimer's disease.
Recent studies have shown that the implementation of Big Data analytics in healthcare settings has significantly improved patient outcomes, particularly in terms of achieving more accurate diagnoses and creating customized treatment plans [4]. To provide a clearer picture of the current landscape, Table 1 presents a comprehensive summary of recent reviews in this field, underscoring the breadth and depth of Big Data's impact on healthcare.

Table 1. Summary of recent reviews relevant to medical Big Data

Review paper	**Year**	**Objective**
Adehge et al. [5]	2024	The review highlights new trends and technologies set to influence the future of big data in healthcare. It points out potential research areas and advocates for interdisciplinary partnerships to drive further innovation in this evolving sector.
Arowoogun et al. [6]	2024	The paper explores foundational concepts, methodologies, and emerging trends such as the integration of artificial intelligence, real-time analytics, and the influence of wearable technologies.
Ogundipe et al. [7]	2024	The review highlights the significant influence of big data on the development of healthcare products. By adopting data-driven strategies, stakeholders can promote innovation, improve patient results, and adapt effectively and efficiently to the changing healthcare environment.
Sardar et al. [8]	2024	The chapter offers an overview of the various healthcare application domains that utilize these ML techniques, focusing on security and privacy perspectives and addressing the associated challenges in healthcare applications.
Selwy et al. [9]	2023	The paper delves into the deep learning (DL) methods employed in large-scale IoT data applications, discussing their complexities and the challenges they present in Big Data analytics.
Furstenau et al. [4]	2023	The research investigated strategic topics, the structure of thematic networks, and the progression of Big Data in the healthcare sector. Additionally, it pinpointed principal challenges, obstacles, and prospects within this domain.
Palmieri et al. [10]	2023	The present review delved into how artificial intelligence (AI) and Big Data are leveraged to enhance the supply chain and outcomes of heart transplantation (HT), spanning from the determination of heart failure (HF) etiology to post-transplant prognosis.
Tonegawa-Kuji et al. [11]	2023	The paper concentrates on Real-World Data (RWD) studies within the field of cardiovascular medicine, detailing their present situation in Japan. It also explores the possibilities for expanding these studies and addresses the challenges associated with the utilization of medical Big Data and RWD.

Continued on next page

Review paper	Year	Objective
Fang et al. [12]	2023	The article primarily examines the advancements in research on biological Big Data and radiomics within the context of HCC, offering innovative approaches and concepts for the diagnosis, prognosis, and treatment of this condition.
Ahmad et al. [13]	2023	This study seeks to bridge this knowledge gap by presenting an in-depth analysis of how the integration of these technologies is poised to shape the future landscape of healthcare systems.
Shim et al. [14]	2023	This review will highlight the significance of Big Data research for the public good and present key public medical Big Data resources in Korea. Furthermore, the authors will demonstrate the application of systematic review methodologies in research involving public medical Big Data.
Ting et al. [15]	2023	This review aims to provide a succinct overview of Big Data's role and clinical significance in ophthalmology, with a particular emphasis on corneal diseases and cataracts. It explores how the integration of Big Data, AI technologies, the Internet of Things, mHealth, and wearable smart devices can enhance the field.
Karatas et al. [16]	2022	This study's primary goal is to present readers with an analysis of publications that intersect with Industry 4.0, Big Data, and healthcare operations, while also offering insights into future trends and developments in these areas.
Galetsi et al. [17]	2022	This study provides a detailed examination of the critical issues and complex challenges that Covid-19 has posed to public health and society, as well as the potential solutions proposed through the lens of information systems and technology.
Lokesh et al. [18]	2022	This article explores the development of integrated healthcare systems, emphasizing the significance of mobile cloud applications and Big Data analysis. It examines the impact of cloud computing adoption in the healthcare sector, highlighting how it has driven the innovation and enhancement of connected health software and services.
Ahmed et al. [19]	2022	The purpose of this paper is to investigate the utilization of big data in the field of mental health. It particularly focuses on the aspects of data volume, velocity, veracity, and variety, along with the methods of data acquisition, storage, management, and maintenance of privacy and security.
Guo et al. [20]	2021	TThe paper introduces an innovative architecture for core dataset extraction, designed as a universal framework to extract the essential dataset from unlabeled medical Big Data specifically for lesion localization.
Supriya et al. [21]	2021	This paper outlines how the foundational concepts of machine learning (ML), big data, and blockchain technology are employed and their significance in fields such as medicine, healthcare, public health monitoring, and case predictions during the COVID-19 pandemic and other epidemic situations.

These studies predominantly focus on illustrating the application of Big Data concepts in healthcare, detailing their use not only across the broader medical field but also in the targeted management of specific diseases. They provide comprehensive insights into the integration of Big Data methodologies to improve diagnostic processes,

treatment strategies, and overall disease management in various medical contexts. Our paper aims to compose an informative review that provides a comprehensive overview of the application of Big Data in both diagnostic and medical domains. This includes exploring how Big Data is currently being used in the healthcare sector, with a focus on its impact on medical research, disease pattern analysis, treatment effectiveness, and the development of personalized medicine and predictive analysis. The structure of the paper is as follows: Section two offers an overview of Big Data technology, section three delves into its impact on medical research, section four discusses medical data management, section five elucidates the concepts of medical data, section six examines the integration of Big Data with AI, section seven clarify ethical implications of using Big Data in healthcare, section eight provides a discussion of our study's findings, and finally, the paper concludes with a summary and conclusions in the last section.

2. Overview of Big Data

The idea of Big Data originated in the 1990s as a means to manage datasets that were too intricate for storage, processing, and analysis by conventional software tools [2]. The term "Big Data" refers to the explosion in data production in the digital era. Brynjolfsson and McAfee [22] regard the Big Data phenomenon as the innovation of the modern era. An innovation whose effects impact all fields of activity, ranging from science and research to the world of business, government service management, and even the arts. Chris Anderson [23] suggests that not only do our capabilities to answer fundamental questions progress with the growth of facts, but Big Data even renders traditional scientific methods obsolete. Lev Manovich [24], representing the digital humanities school, notes that the term Big Data is poorly suited. In scientific terms, it was used to refer to datasets large enough to require supercomputers, while vast datasets can now be analyzed on desktop computers using standard software. The novelty of Big Data lies not only in the quantity of data produced and processed daily by individuals, organizations, and objects but also in their ability to be linked to other data, connecting several databases. Due to efforts to exploit and aggregate data, Big Data is fundamentally linked to networks that allow datasets concerning an individual, individuals linked to others, groups of people, or simply concerning the structure of information itself, to interconnect within a single set.

2.1. Characteristics of Big Data

To characterize the Big Data phenomenon, multiple definitions often revolve around the model of the three "V"s: Volume, Velocity, and Variety [25]. Other "V" models have been developed subsequently to encompass all the characteristics of the phenomenon: Volume: The quantity of data generated will continue to increase exponentially in the era of social media and connected devices. Velocity or Speed: Massive data is generated increasingly rapidly and in continuous streams. Their analysis requires more real-time processing. Variety: This data is variable, appearing in different formats (like text, image, and sound). It can be structured, semi-structured, or unstructured. Additional attributes such as veracity (Big Data accumulates detailed and comprehensive data) and value (Big Data offers in-depth information on a subject of discussion) have also been suggested to fully capture the real nature and values of Big Data, known as the five "Vs" [15]. From another viewpoint, Kitchin [26] summarized the characteristics of Big Data with seven "Vs," adding terms like variability (data whose meaning is constantly changing) and valence (Big Data connects common domains to bring together different datasets) [9], [27].

2.2. Big Data Technology

Today, the applications of Big Data have vastly expanded to accommodate a variety of needs. Big Data facilitates several aspects of problem-solving, such as predictive analysis, which is particularly useful in preventive maintenance, sales forecasting, and inventory management. Real-time data analysis is another critical application of Big Data. Various Big Data technologies have been developed to cater to these requirements [28]:
MapReduce: Originating from the laboratories of Google Corp, a massively parallel processing method and technology has been developed, featuring fault tolerance management and a unique file system management approach, known as the Google File System [29]. This method entails processing data across thousands of

machines, which are distributed in cluster formations.
Hadoop: A framework developed by the Apache Software Foundation to generalize the use of storage and massively parallel processing of MapReduce and Google File System [29]. Hadoop has its limitations but has become a widely used Big Data solution for analyzing vast amounts of data. It consists of several components: a storage system (HDFS), a treatment planning system (YARN), and the processing framework (MapReduce). One of the most well-known use cases of Hadoop is the data lake.
NoSQL Databases: Traditional relational databases are adept at handling structured data but fall short in storing and rapidly processing data at a large scale. In contrast, NoSQL databases present a novel approach to data storage, characterized by greater flexibility, adaptability to changes, and a lower likelihood of system failures [30]. These databases support redundancy, enhancing their ability to meet demands for flexibility, fault tolerance, and scalability.
Cloud Computing: Big Data requires exceptional hardware capabilities, encompassing both storage and processing resources. Cloud Computing, as a service delivery model, facilitates access to shared computational resources like servers, storage solutions, databases, networks, software, and various services over the Internet, commonly referred to as "the cloud" [31]. It is essential to differentiate between private cloud and public cloud, internal from external, and hybrid clouds combining multiple solution types. Moreover, it is also important to distinguish the service levels of each solution: IaaS, PaaS, SaaS.
Column-Oriented Databases like Cassandra and Hbase: These are data management systems. They are highly efficient databases in reading and writing large volumes of data. This type of database can handle incremental scaling without sacrificing existing features [32].
Batch Processing: It allows processing data until it is depleted at the system's input. The treatments are continuous and incremental, meaning the architecture takes into account new data each time without having to reprocess the old ones [33]. To maintain consistency in processing this data, the results are only visible and accessible at the end of the process. Examples of Big Data batch processing include MapReduce in its Hadoop version or Apache Spark.
Real-time Processing (Streaming): This method stands in contrast to batch processing. In this approach, waiting for the entire data processing to complete before accessing results is not necessary. It offers a straightforward implementation and leads to more efficient processing times [34]. These are often used as the basis to implement scalable solutions.
Lambda Architecture: This approach merges batch and real-time processing techniques. Through batch processing, the architecture achieves a balance in latency, throughput, and system fault tolerance. It ensures accurate data views are maintained while integrating real-time data, leading to more precise outcomes [35].
Saagie, the "technological gem" of Big Data: Saagie offers a Big Data as a Service platform, integrating all existing Big Data technologies. It provides an end-to-end platform for storing, processing, analyzing, and exposing data, along with intelligent applications embedded with components and algorithms for specific business needs.
This overview presents a comprehensive understanding of the current state of Big Data technology and its applications, demonstrating its critical role in modern data management and analysis.

2.3. Fields of Application of Big Data

The fields of application for Big Data technologies are numerous, showcasing its extensive impact across various domains and its contribution to decision-making, operational efficiency, and innovation [25]. For example, in healthcare [4], analyzing massive medical data enhances diagnosis, medical research, electronic health records management, and public health monitoring [16]. In finance [36], financial institutions use Big Data for fraud detection, risk analysis, market modeling, and personalized financial services. Another example is in marketing and advertising [37], where Big Data aids businesses in understanding consumer behavior, targeting ads more accurately, and measuring campaign effectiveness. For industry [38], key applications of Big Data include optimizing manufacturing processes, predictive maintenance, and supply chain management. In transportation and logistics [39], Big Data enhances route planning, fleet management, vehicle condition monitoring, and warehouse management [40]. In the energy and environment sector [41], it is utilized to monitor energy consumption, optimize electricity distribution, and for applications in natural resource management and combating climate change. In education [42], analyzing educational data helps adapt teaching methods, identify struggling students, and enhance

educational program effectiveness. Governments use Big Data for public service management [43], policymaking, national security, combating crime, and tourism [44]. In science [45], researchers employ Big Data for climate modeling, astrophysics research, genomics, and other fields [46]. Finally, technology companies use Big Data to improve product performance, optimize social networks, and develop new technologies. There are also applications of Big Data in surveillance, traceability, and power reconfiguration needs.
Big Data technology refers to the set of tools, techniques, and methodologies used to manage, store, analyze, and leverage Big Data. This technology encompasses infrastructures, databases, analytics software, data processing tools, data flow management techniques, and much more. It plays a crucial role in effectively harnessing massive data to extract valuable insights.

3. Impact of Big Data on Medical Research

The impact of Big Data on medical research has been transformative, ushering in significant advancements and changes in various aspects of the field [45]. Big Data has enabled a deeper understanding of diseases by allowing researchers to analyze large datasets, uncovering patterns and correlations that were not evident before [47]. This comprehensive analysis has been instrumental in the shift towards personalized medicine, where treatments are tailored to individual patients based on their genetic profiles and lifestyle factors [7].
In drug development, Big Data has accelerated the process, enhancing the efficiency of clinical trials. It facilitates the identification of suitable candidates and enables real-time data monitoring, leading to quicker and more informed decisions [48]. Additionally, the predictive analytics capabilities of Big Data tools have become indispensable in public health. They allow for the prediction of disease outbreaks, potential health risks, and the spread of infectious diseases, which is crucial for planning and response strategies.
The field of genomic research has also greatly benefited from Big Data [49]. The analysis of vast genomic datasets has led to a better understanding of genetic disorders and the development of gene-based therapies. Moreover, Big Data's role in healthcare delivery cannot be overstated. By analyzing treatment outcomes and identifying best practices, it contributes to the improvement of healthcare systems and services.
Furthermore, Big Data analysis informs public health policies, leading to more effective strategies that address broader health issues of populations. It also fosters collaborative research efforts, breaking down barriers between researchers, institutions, and even countries, and promoting a more integrated approach to medical research [50].
In conclusion, the introduction and integration of Big Data into medical research have led to more efficient, personalized, and predictive healthcare, marking a significant leap forward in the field. Additionally, the utilization of Big Data for medical data management has been a game-changer. It has revolutionized the way medical data is collected, stored, analyzed, and utilized. Big Data technologies have made it possible to manage the enormous volumes of data generated in healthcare, ranging from patient records and imaging to genomic data. This has not only streamlined data management processes but also enhanced the accuracy and accessibility of medical information.

4. Big Data for Medical Data Management

Big Data holds immense significance in the medical field due to its numerous advantages and its potential to revolutionize disease diagnosis, medical research, and medical data management [51]. Several reasons make Big Data essential in the medical domain:
More Accurate and Early Diagnosis: Big Data allows the analysis of vast medical datasets, facilitating the detection of trends and correlations that might go unnoticed in smaller datasets. This enables early disease diagnosis, which can improve the chances of effective treatment.
Treatment Personalization: By analyzing genetic data, medical histories, health sensor data, and other information, Big Data enables personalized treatments for patients. This means healthcare is tailored to individual needs, enhancing treatment effectiveness.
Advanced Medical Research: Big Data plays a crucial role in medical research by analyzing massive data to identify

new research avenues, discover new drugs and better understand disease mechanisms.
Epidemic Management: Big Data can help prevent and control epidemics by monitoring and analyzing real-time epidemiological data. This is particularly crucial in responding swiftly to threats like pandemics.
Resource Optimization: Big Data is used to efficiently manage medical resources by predicting needs, planning staff allocations, and optimizing logistical processes.
Enhanced Quality of Care: By providing accurate and updated information to healthcare professionals, Big Data contributes to improving the quality of care and reducing medical errors.
Detection of Rare Side Effects: Big Data can help detect rare side effects of medications or medical procedures by analyzing large-scale patient data.
Remote Patient Monitoring: Connected medical devices and health applications enable continuous data collection on patients, facilitating remote monitoring and early intervention in case of health issues.
Electronic Health Record Management: Big Data is essential for managing electronic health records, making their storage, management, access, and secure sharing more accessible.
Big Data transforms the way medical data is collected, managed, and analyzed, significantly impacting the quality of care, medical research, and healthcare system efficiency. A variety of Big Data techniques are used for managing and integrating medical data, including distributed storage such as Hadoop Distributed File System (HDFS), batch processing frameworks like Apache Spark, NoSQL databases like MongoDB and Cassandra for storing unstructured or semi-structured medical data, and cloud services providing Big Data-scale storage and processing for medical data. Additionally, real-time Big Data systems analyze and react to medical data in real-time, which can be critical in emergency healthcare. These techniques are used together to effectively manage and integrate medical data, obtaining valuable insights for clinical decision-making, research, and enhancing disease diagnostic methods.
In essence, big data's contribution to medical data management has not only optimized healthcare operations but also paved the way for innovations that continue to transform the landscape of medical research and patient care.

5. Medical Data

There exist numerous definitions of large medical data (medical big data) proposed by various researchers, and some categorize large medical data based on the individual or entity that owns the data in comparison to traditional clinical data [14]. Fundamentally, accessing large medical datasets often presents challenges, and many researchers in the medical field are hesitant to share their data, concerned about potential misuse by others. Furthermore, large medical datasets are generally well-structured, adhering to specific protocols for collecting individual medical information, which comes in diverse forms.

5.1. Types of medical data

Medical data can be classified into various types or natures based on their content, origin, and usage in the healthcare domain. Table 2 presents some examples of the most common types of medical data.

Table 2. Examples of the most common types of medical data

Type of medical data	**Definition**	**Examples**
Clinical data	Includes patient information	Age, gender, medical history, symptoms, diagnoses, treatments, laboratory test results, medical imaging results, clinical observations

Continued on next page

Type of medical data	Definition	Examples
Genetic data	Encompasses information about an individual's genetic code. They are used in genomics and personalized medicine	DNA, genes, genetic variants, mutations, genetic sequences
Laboratory data	They provide physiological insights and into biochemical parameters	Blood tests, urine tests, tissue analysis, biological fluid analysis, microbial cultures
Prescription data	Involves information about medications prescribed to patients	Medication name, dosage, treatment duration, contraindications
Surveillance data	involve continuous or periodic measurements of vital parameters	Heart rate, blood pressure, oxygen saturation, temperature
Medical research data	These data are generated as part of medical research	Epidemiological data, long-term follow-up data, clinical research data, genomic sequencing data
Administrative data	Comprise information related to the administration	Record management, billing, health insurance, appointment scheduling, healthcare facility management
Telemedicine data	These data are generated in the context of different remote medical	Consultations, teleconferencing, monitoring
Medical device monitoring data	Data from connected medical devices, provide real-time information on patient health	Glucose monitors, blood pressure monitors, pacemakers
Medical imaging data	These data consist of images of the human body obtained through various modalities	X-rays image, ultrasound image, scintigraphy image, radiography image management

These various types of medical data are utilized for diagnosis, treatment, medical research, healthcare management, public health surveillance, and many other applications in the medical field [52]. They are often compiled internationally by governmental organizations, research institutions, public health agencies, and international organizations to assess trends and risk factors. For instance, mortality data, vaccination data, epidemic data, as well as statistics on cardiovascular diseases, diabetes, cancer, and other non-communicable diseases are collected to evaluate trends and risk factors. These statistics are frequently published in annual reports, online databases, and scientific publications by organizations such as the World Health Organization (WHO). Additionally, medical data can be derived from various sources, and in the following section, we'll discuss some sources and challenges related to the collection and normalization of medical data.

5.2. Medical Data Sources

Recently, data collection has become essential for most organizations due to the rapid expansion of data analysis tools, enabling better utilization of collected raw data, providing added value, and positive impact on these enterprises. However, the growth in the quantity of collected data may include personally identifiable information that needs to be protected to comply with relevant laws and regulations. For instance, in the healthcare domain, it's evident that the use of recent information and communication technologies (such as big data, AI, cloud computing, the Internet of Things) enhances communication, facilitates access to the right information at the right time, and ensures better quality of patient care. However, data collected, stored, and processed by these technologies often contain sensitive information, posing new challenges regarding security and privacy protection. Many approaches and solutions are being implemented to mitigate these issues. Specifically, concerning privacy protection, it is widely acknowledged that anonymization techniques are among the most effective approaches.
Access to high-quality medical data sources is crucial for the development and evaluation of AI models aimed at medical diagnosis. Some commonly used data sources in this field include:

Public medical databases: Numerous medical institutions and governments make medical databases accessible to the public. For example, the National Institutes of Health (NIH) [53] in the United States offers medical datasets, as does the World Health Organization (WHO), and other health agencies worldwide. Medical image databases are often available for research purposes.
Electronic Medical Records (EMRs): EMRs store patient information, such as medical histories, test results, X-rays, laboratory reports, etc. Researchers can access anonymized EMRs to train and evaluate their models [54].
Genomic and genetic data: For genetics-related diagnostics, genomic and genetic data, including DNA sequences and functional genomics data, are essential.
Laboratory reports: Laboratory data, such as blood test results, chemical analyses, and metabolic profiles, can be used for diagnosing various medical conditions.
Patient monitoring data: Patient monitoring data collected from wearable medical sensors, home monitoring devices, symptom logs, etc., are useful for tracking disease progression and evaluating treatment effectiveness.
Clinical trial data: Clinical trial data, including information on medical interventions, patient outcomes, and long-term follow-ups, can be used to develop evidence-based diagnostic models.
It's important to note that using medical data comes with challenges of confidentiality and security, and compliance with data protection and patient privacy regulations is crucial. Moreover, accessing this data might be subject to strict regulations. Researchers and healthcare professionals need to closely collaborate with legal and ethical experts to ensure the use of this data complies with all applicable regulations.
The most commonly used medical data in medical practice is medical images, which is why most DL algorithms have focused on this category of information for developing medical applications. The ongoing paper focuses on the analysis and utilization of medical images.

5.3. Medical imaging

The field of medical imaging has seen significant advancements in recent decades, and medical image analysis is an essential branch within the field of computer vision. These images are available in databases like: DICOM Library (Digital Imaging and Communications in Medicine), TCIA (The Cancer Imaging Archive), IDR (Image Data Resource a database of medical images with a focus on biomedical research), Openi (Open Access Biomedical Image Search Engine), MIMD (NIH National Library of Medicine's Medical Image Database). There are also well-known databases that are public and accessible to specialists in the field of scientific research. Table 3 summarizes some medical image datasets used in various studies.

Table 3. Examples of Medical Image Datasets

Disease	**Datasets**	**Description**	**Number of images**	**Reference**
Pulmonary	RSNA	RSNA pneumonia detection challenge	29700	[55]
	X-ray14	Picture Archiving and Communication Systems (PACS)	108948	[56]
	OPEN-I INDIANA	Indiana University School of Medicine	7470	[57]
	MC	Montgomery County's	138	[58]
	SHENZHEN	Shenzhen's Hospital (Guangdong Province, China)	662	[58]
	KIT	Korea Institute of Tuberculosis	108948	[59]
	JSRT	Japanese Society of Radiological Technology	247	[60]
Cancerous	LIDC	lung nodule dataset Cancer Imaging Archive (TCIA) Typology dataset	176	[61]
	INbreast	Mammographic images	116	[62]
	Prostate cancer	Cancer du prostate	1235	[63]

Disease	Datasets	Description	Number of images	Reference
	BRATS	Brain tumor	1639	[64]
	Pap smear	Cervical cancer dataset	963	[65]
	LUNA16	Lung nodule analysis	888	[66]
	SIPaKMeD	Cervical cancer dataset images of isolated cells	4049	[67]
	Herlev	Cervical cancer dataset isolated single-cell images	917	[68]
	ISBI 2016	Analysis of skin lesions	1250	[69]
Ocular	DRISHTI-GS	Retinal images for diabetic retinopathy screening	101	[70]
	RIM-ONE	Retinal images for optic nerve assessment	169	[71]
	IDRiD	Indian Diabetic Retinopathy Image Dataset	516	[72]
	STARE	Structured Analysis of the Retina	385	[73]
	RIGA	Retinal fundus images for glaucoma analysis	750	[74]
	APTOC	APTOC's Asia Pacific Tele-Ophthalmology Society 2019 blindness detection	5590	[75]
	EyePACS	Diabetic Retinopathy Detection	35126	[76]
Cardiovascular	IVUS	IVUS Image Dataset	175	[77]
	SPECTF	SPECTF Heart Data Set	267	[78]
	ACDC	Automatic Cardiac Diagnosis Challenge	150	[79]

These datasets of medical images are freely available for scientific research, complemented by other private databases that require permission for use. Although the number of images in each database might seem limited, a well-known method exists to increase the image count [80]. This method involves applying various image transformations to the original images, thereby generating multiple transformed copies of each image [81]. Each copy differs from the others in some aspects, depending on the augmentation techniques applied, such as shifting, rotating, and flipping. These techniques not only expand the database size but also introduce variation into the dataset, helping the AI model to better handle unseen data. Additionally, the model becomes more robust when trained on these slightly altered new images.

Data augmentation is a strategy primarily aimed at enhancing the diversity and volume of data without acquiring new samples [82]. It includes applying different image manipulation techniques to the original data or creating new samples using generative models. In scientific research, many studies investigate various data augmentation methods to mitigate overfitting in DL models caused by limited data, as noted in [83], DL models often require extensive datasets to prevent overfitting. By artificially enlarging databases with the methods detailed in this study, it becomes feasible to use large datasets even in areas with scarce data. However, given the unique nature of medical images, data augmentation approaches need to be customized for this specific domain. Despite this, data augmentation remains a highly valuable technique for enhancing database construction. Our research work suggests many types of augmentations, typically categorized into data warping or oversampling techniques.

5.4. Acquisition and modalities of Medical Images

Biomedical imaging has transformed medical practice by providing unparalleled abilities to diagnose diseases. It achieves this through high-resolution visualizations of the human body and detailed observations of cells and

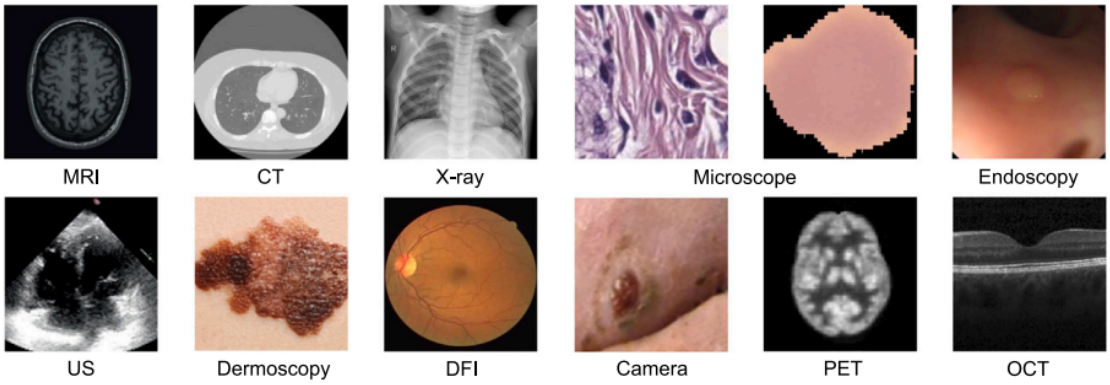

Figure 1. Examples of medical imaging modalities.

pathological specimens [84]. Generally, medical images are produced by the interaction of electromagnetic waves of various wavelengths (energies) with biological tissues, except for ultrasound, which utilizes mechanical sound waves. Images formed using high-energy radiation at short wavelengths, such as X-rays and gamma rays, are ionizing. In contrast, images at longer wavelengths (optical) and even longer wavelengths (MRI and ultrasound) are non-ionizing [85].

Medical images encompass a diverse range of modalities, including radiographs (X-ray), CT scans, MRI (Magnetic Resonance Imaging), ultrasound (US), histological images, Positron Emission Tomography (PET), Optical Coherence Tomography (OCT), Digital Fundus Imaging (DFI), among others [86]. These images are crucial for creating visual representations of the body's interior, significantly aiding in medical diagnosis, treatment monitoring, and research. Figure 1 illustrates examples of medical images from various modalities: To avoid any loss of important information in medical images, preprocessing of these images must be carefully performed. The choice of preprocessing techniques will depend on the imaging modality, the type of data under study, the analysis objectives, and the nature of the noise present in the images.

5.5. Methods for preprocessing and processing medical images

Preprocessing of medical images is an essential step in their analysis and treatment. Several techniques are employed to enhance the quality and usability of images, particularly in the context of medical imaging. Sharpness enhancement focuses on emphasizing contours and details by increasing image sharpness, making features more distinct. Noise filtering is crucial for removing undesirable elements like Gaussian or speckle noise, thereby improving image clarity. Contrast normalization and histogram equalization are methods used to adjust contrast levels and redistribute grayscale levels, respectively, both aiming to enhance feature visibility. Size reduction is another technique, which decreases image dimensions to speed up processing while retaining important features.

Color normalization is essential, especially when dealing with images from diverse sources, as it homogenizes colors. Illumination correction helps in reducing lighting variations to enhance the robustness of image analysis. Frequency filtering is used to filter specific frequency components, highlighting particular features in an image. Artifact removal is important for eliminating unwanted artifacts that might arise due to equipment imperfections or patient movements.

Resizing images to change their resolution is often necessary to meet the requirements of specific models or applications. Image registration involves aligning multiple images from the same patient or different modalities, facilitating joint analysis. Finally, segmentation is a critical process that divides an image into distinct regions or objects, which can significantly enhance the accuracy of subsequent analyses. Each of these techniques plays a vital role in preparing images for effective analysis and interpretation in various applications, including medical diagnosis and research. They can be used individually or in combination, depending on the specific application needs and image characteristics. The preprocessing of medical images aims primarily to obtain higher quality

images, making them ready for use in specific medical applications such as diagnosis, disease detection, treatment planning, etc. It's also an essential step in processing medical images.
Medical image processing comes after preprocessing and involves more advanced analyses, such as anatomical structure detection, organ segmentation, feature measurement, pathology classification, and more. Medical image processing techniques and medical data analysis are crucial in the field of modern medicine [87]. They extract valuable insights from medical images and clinical data to assist healthcare professionals in diagnosis, disease monitoring, treatment planning, and medical research.

6. Big data and IA

The integration of Big Data and AI in the medical domain has indeed brought about radical changes with transformative potential for the healthcare landscape [88]. The advent of Big Data presents immense opportunities for enhancing the management and integration of medical data, thereby contributing to improved healthcare quality, advancing medical research, and streamlining healthcare systems [89]. Advances in AI have enabled the fusion and analysis of multimodal data, marking a new era in extracting meaningful information from complex Big Data sets. Despite these advancements, challenges remain in effectively curating and comprehensively analyzing this data, as well as in maximizing its utilization. Nonetheless, the effective analysis of this data using AI techniques and algorithms remains a crucial component of this technological evolution [90]. Several key examples of Big Data and AI applications in the medical domain can be cited:
Diagnostic Assistance: Big Data and AI contribute to more precise disease diagnoses by analyzing extensive patient data, encompassing medical images, genomic information, and clinical records. AI algorithms are adept at identifying patterns and anomalies that may not be readily noticeable to human diagnosticians.
Personalized Medicine: They enable the development of personalized treatment plans based on an individual's genetic makeup, medical history, lifestyle, and other factors. This tailored approach enhances treatment effectiveness and minimizes adverse effects.
Drug Discovery and Development: Big Data analytics and AI streamline drug discovery processes by analyzing large datasets to identify potential drug candidates more efficiently. AI models help in predicting the effectiveness of drugs and in designing new compounds.
Healthcare Management and Optimization: AI and Big Data assist in managing healthcare systems by optimizing hospital operations, resource allocation, and staff scheduling. They also aid in predicting disease outbreaks and planning for public health emergencies.
Remote Monitoring and Telemedicine: They facilitate remote patient monitoring and telemedicine services, allowing healthcare providers to monitor patients' health status, offer consultations, and provide medical care from a distance.
Predictive Analytics and Preventive Care: Big Data and AI predict potential health risks and diseases by analyzing patterns from patient data. This enables early interventions and preventive measures to maintain patient health.
Natural Language Processing (NLP) in Healthcare: NLP is used to analyze unstructured medical data like doctor's notes, patient records, and research papers. It helps in extracting valuable information for research, diagnosis, and treatment.
Genomics and Precision Medicine: Big Data and AI play a crucial role in analyzing genomic data to understand genetic predispositions to diseases, allowing for more accurate diagnoses and targeted treatments.
These applications illustrate the wide-ranging impact of Big Data and AI in revolutionizing healthcare, encompassing improvements in diagnostics, the transformation of treatment strategies, and enhancements in healthcare management. In particular, AI methods, especially ML and DL, have achieved significant milestones recently, showcasing their potential to augment and automate aspects of medical practice [91]. However, fully and securely integrating these methods into clinical workflows necessitates a collaborative, multidisciplinary approach involving computer science, information technologies, and medical expertise. This collaboration is essential for developing the next generation of high-performing AI methods, ensuring they are robust and interpretable [92].
Moreover, when crafting regulations for the ethical use of AI in healthcare, policymakers must prioritize

transparency to ensure patients and healthcare providers understand how AI algorithms operate. Privacy concerns should be addressed by implementing robust measures to secure patient data throughout its lifecycle. Policymakers must confront issues of bias and fairness in AI algorithms, requiring regular assessments and audits to mitigate potential disparities. Accountability mechanisms should be established, defining responsibilities in cases of errors or adverse outcomes and providing avenues for recourse. Continuous monitoring and evaluation of AI systems, along with international collaboration and adaptability to technological advancements, are essential considerations to ensure ethical and responsible AI implementation in healthcare.

7. Ethics of Big Data in Healthcare

Big Data plays an important role in public health by enabling more effective epidemiological surveillance and crisis management. For example, during the COVID-19 pandemic, real-time data analysis helped track the spread of the virus, identify infection hotspots, and allocate medical resources optimally. Big Data can also be used to analyze the social determinants of health, identify at-risk populations, and develop targeted interventions to reduce health inequalities. Additionally, public health data can be integrated with environmental information to study the impacts of climate change on human health and plan appropriate responses. However, these applications must be accompanied by rigorous measures to ensure data confidentiality and security and to overcome challenges related to interoperability and algorithmic biases.

The use of Big Data in healthcare presents significant ethical implications that must be considered [93]. One of the primary concerns is data privacy. Medical records contain highly sensitive information and protecting them from data breaches is crucial. Robust security measures must be implemented to ensure that patient data is only accessible to authorized individuals and that the information is not misused.

Data security is another critical dimension. Big Data systems must be designed to withstand cyberattacks and unauthorized access. This includes the use of advanced encryption techniques and the implementation of rigorous security protocols to protect data both in transit and at rest.

Informed consent is also fundamental. Patients must be fully informed about how their data will be collected, stored, used, and shared. They should have the option to give or withdraw their consent at any time. This requires transparent and accessible mechanisms for managing consent, as well as clear and honest communication with patients about the potential risks and benefits associated with the use of their data.

Additionally, it is important to consider algorithmic biases and the inequalities they can exacerbate. Big Data models can reflect and amplify biases present in the training data, leading to unfair diagnoses and treatments. Therefore, it is essential to develop transparent and fair algorithms and maintain continuous oversight to identify and correct potential biases.

Finally, the ethical use of Big Data in healthcare must balance technological innovation with respect for patient rights, ensuring that medical advancements benefit everyone without compromising individual integrity and dignity.

8. Discussion

At present, the promotion of medical technology and the ongoing evolution in AI methods, particularly the adaptation of the latest innovations in ML and DL from computer vision to medical applications, are a testament to the significant research focus in this area [84]. This shift demonstrates the immense capabilities and practical benefits of these methods in enhancing clinical practice. The incorporation of advanced AI techniques, originally developed for computer vision, into medical diagnostics and treatment, has shown considerable promise in improving the accuracy and efficiency of medical procedures [94]. As technology continues to bridge the gap between computer science and medicine, the future of AI in medical applications heavily relies on the willingness of the medical community to embrace these technological advancements. This means not only adopting AI technology but also integrating domain-specific medical knowledge into these cutting-edge methods. Such integration is crucial for ensuring that AI tools are fine-tuned to the unique requirements and challenges of medical practice. One of the most promising areas of application for AI in medicine is in early disease diagnosis. Recent

advancements in AI offer clinicians and physicians powerful diagnostic tools, capable of identifying various diseases in their initial stages. This early detection is crucial, as it can lead to more effective treatment and significantly better patient outcomes. For instance, in oncology, AI-driven tools can detect subtle patterns in imaging data that might be indicative of early-stage cancers, which are often missed by the human eye. Moreover, the integration of AI in early disease diagnosis doesn't just stop at detection. These tools also hold the potential for predicting disease progression and patient outcomes, enabling healthcare providers to make more informed decisions about treatment strategies. This aspect of predictive analysis is particularly important in managing chronic diseases, where understanding the disease trajectory can lead to more personalized and effective care plans. However, realizing the full potential of AI in medical applications is not without challenges. Issues such as data privacy, the ethical implications of AI decisions, and the need for robust datasets for training AI models are just some of the hurdles that need to be addressed. Furthermore, there's a need for continuous collaboration between technologists and medical professionals to ensure that AI tools are not only technically sound but also clinically relevant and reliable. Indeed, the evolution of AI methods from computer vision to medical applications marks a significant stride in healthcare technology. The continued adaptation and integration of these methods into clinical practice have the potential to bring about a revolution in patient care, particularly in early disease diagnosis. However, the successful implementation of these technologies requires a concerted effort from both the technology and medical communities to overcome existing challenges and harness the full potential of AI in medicine.

Alternatively, the analysis of medical Big Data entails the application of sophisticated analytics and computational methods to extensive and complex data sets within the medical field [95]. This process involves the extraction of crucial information and identifying patterns and correlations in a broad range of medical data such as patient records, diagnostic images, genomic data, and treatment outcomes. The primary objective of this analysis is to glean significant insights that can improve patient care, assist in diagnosing diseases, inform treatment planning and drug development, and advance medical research [96]. Frequently, this analysis incorporates a variety of approaches, including machine learning, deep learning, natural language processing, and statistical methods, to produce useful information that can enhance the quality of healthcare services.

Moreover, leveraging medical Big Data for analysis significantly bolsters the predictive modeling capacity in healthcare [97]. It enables medical professionals to predict the emergence or progression of diseases by analyzing a patient's health history and other pertinent data. Such an approach facilitates a proactive healthcare strategy, allowing both patients and practitioners to engage in preventative actions, which could avert or mitigate the impact of age-related conditions. The analytical goals of medical Big Data include prediction, modeling, and inference. Techniques like classification, clustering, and regression are typically used in these analyses. Classification, a type of supervised learning, is often used in predictive modeling where the outcome is a categorical variable. This method is useful for creating decision support systems that can provide a diagnosis from various possibilities, or for building models that predict outcomes based on the analysis of a range of biomarkers. Clustering, an unsupervised learning technique, is utilized to identify patterns within data using distance metrics. It's especially relevant in microarray data analysis or phylogenetic studies and can be applied to redefine diseases based on their pathophysiological mechanisms, offering more tailored therapeutic approaches. Regression, another form of supervised learning with a continuous outcome variable, is a statistical tool used to quantify the relationship between dependent and independent variables, thereby revealing trends within the data.

On the other hand, future research directions in the field of Big Data in healthcare will focus on leveraging emerging technologies such as AI for disease diagnosis and prediction. Real-time analysis and continuous patient monitoring through wearable devices will enable early detection of epidemics and rapid response. Health information technologies, such as interoperable electronic health records and blockchain, will secure data transactions and facilitate information exchange between institutions. Personalized medicine, public health, epidemiology, and the optimization of healthcare services represent potential growth areas where Big Data can improve treatments and interventions. However, unresolved challenges such as data privacy and security, interoperability and standardization of formats, biases in algorithms, and the acceptance of new technologies by clinicians must be addressed to maximize the benefits of these technological advancements.

Furthermore, the limitations of using Big Data in healthcare include major concerns about data privacy and security, particularly the protection of patients' sensitive information from cyberattacks and unauthorized access.

Interoperability between different health systems remains a challenge, with the need to standardize data formats and integrate information from disparate sources. Biases in AI algorithms can lead to disparities in care, necessitating ongoing efforts to detect and correct these biases to ensure equity. Additionally, the adoption and acceptance of Big Data technologies by healthcare professionals require adequate training to ensure effective and beneficial use of these advanced tools.

9. Conclusion

In conclusion, our paper presents an informative review that offers a comprehensive overview of Big Data applications in diagnostics and medical domains. We investigate Big Data's current role in healthcare, emphasizing its impact on medical research, disease pattern analysis, treatment efficacy, and the development of personalized medicine and predictive analysis. While Big Data and AI hold immense promise for revolutionizing medical data analysis and healthcare delivery, fully realizing this potential necessitates overcoming challenges in data integration, management, privacy, and ethics. This will require a collaborative effort from various stakeholders in the healthcare ecosystem. Our work focuses on how Big Data advancements are transforming medicine and healthcare, acknowledging the need to address critical issues like data privacy, regulatory compliance, and the interpretability of Big Data techniques. Despite these challenges, the synergy of Big Data and AI in medicine is paving the way for substantial advancements in healthcare, research, and disease diagnosis. The interplay between Big Data and AI not only provides the necessary data for intelligent decision-making but also drives innovation and digital transformation across multiple industries.

Acknowledgement

The authors express their sincere gratitude to Professor Abdelfettah Daoudy, an ESL teacher at Amideast and a researcher in the field of Applied Linguistics at Mohammed V University, for proofreading this article.

REFERENCES

1. D. Tosi, R. Kokaj, and M. Roccetti, *15 years of Big Data: a systematic literature review*, Journal of Big Data, vol. 11, no. 1, 2024.
2. M. Mallappallil, J. Sabu, A. Gruessner, and M. Salifu, *A review of big data and medical research*, SAGE Open Medicine, vol. 8, 2020.
3. B. A. Ojokoh et al., *Big data, analytics and artificial intelligence for sustainability*, Scientific African, vol. 9, p. e00551, 2020.
4. L. B. Furstenau et al., *Big data in healthcare: Conceptual network structure, key challenges and opportunities*, Digital Communications and Networks, vol. 9, no. 4, pp. 856–868, 2023.
5. E. P. Adeghe, C. A. Okolo, and O. T. Ojeyinka, *The role of big data in healthcare: A review of implications for patient outcomes and treatment personalization*, World Journal of Biology Pharmacy and Health Sciences, vol. 17, no. 3, pp. 198–204, 2024.
6. J. O. Arowoogun, O. Babawarun, R. Chidi, A. O. Adeniyi, and C. A. Okolo, *A comprehensive review of data analytics in healthcare management: Leveraging big data for decision-making*, World Journal of Advanced Research and Reviews, vol. 21, no. 2, pp. 1810–1821, 2024.
7. D. O. Ogundipe, *The Impact of Big Data on Healthcare Product Development: a Theoretical and Analytical Review*, International Medical Science Research Journal, vol. 4, no. 3, pp. 341–360, 2024.
8. T. H. Sardar, K. Amina, S. Souvik, A. Yusuf, and A. Tabassum, *Machine Learning in the Healthcare Sector and the Biomedical Big Data: Techniques, Applications, and Challenges*, Big Data Computing, 2024.
9. H. A. Selmy, H. K. Mohamed, and W. Medhat, *Big data analytics deep learning techniques and applications: A survey*, Information Systems, vol. 120, no. October 2023, p. 102318, 2024.
10. V. Palmieri et al., *Artificial intelligence, big data and heart transplantation: Actualities*, International Journal of Medical Informatics, vol. 176, no. May, p. 105110, 2023.
11. R. Tonegawa-Kuji, K. Kanaoka, and Y. Iwanaga, *Current status of real-world big data research in the cardiovascular field in Japan*, Journal of Cardiology, vol. 81, no. 3, pp. 307–315, 2023.
12. G. Fang, J. Fan, Z. Ding, and Y. Zeng, *Application of biological big data and radiomics in hepatocellular carcinoma*, iLIVER, vol. 2, no. 1, pp. 41–49, 2023.
13. H. F. Ahmad, W. Rafique, R. U. Rasool, A. Alhumam, Z. Anwar, and J. Qadir, *Leveraging 6G, extended reality, and IoT big data analytics for healthcare: A review*, Computer Science Review, vol. 48, p. 100558, 2023.
14. S. R. Shim, J. H. Lee, and J. H. Kim, *Medical Application of Big Data: Between Systematic Review and Randomized Controlled Trials*, Applied Sciences (Switzerland), vol. 13, no. 16, 2023.

15. D. S. J. Ting, R. Deshmukh, D. S. W. Ting, and M. Ang, *Big data in corneal diseases and cataract: Current applications and future directions*, Frontiers in Big Data, vol. 6, 2023.
16. M. Karatas, L. Eriskin, M. Deveci, D. Pamucar, and H. Garg, *Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives*, Expert Systems with Applications, vol. 200, no. March, p. 116912, 2022.
17. P. Galetsi, K. Katsaliaki, and S. Kumar, *The medical and societal impact of big data analytics and artificial intelligence applications in combating pandemics: A review focused on Covid-19*, Social Science and Medicine, vol. 301, no. April, p. 114973, 2022.
18. S. Lokesh, S. Chakraborty, R. Pulugu, S. Mittal, D. Pulugu, and R. Muruganantham, *AI-based big data analytics model for medical applications*, Measurement: Sensors, vol. 24, no. October, p. 100534, 2022.
19. A. Ahmed et al., *Overview of the role of big data in mental health: A scoping review*, Computer Methods and Programs in Biomedicine Update, vol. 2, no. March, p. 100076, 2022.
20. K. Guo, Y. Wang, J. Kang, J. Zhang, and R. Cao, *Core dataset extraction from unlabeled medical big data for lesion localization*, Big Data Research, vol. 24, p. 100185, 2021.
21. M. Supriya and V. K. Chattu, *A review of artificial intelligence, big data, and blockchain technology applications in medicine and global health*, Big Data and Cognitive Computing, vol. 5, no. 3, 2021.
22. E. Brynjolfsson and A. McAfee, *The Big Data Boom is the Innovation Story of Our Time*, The Atlantic, pp. 1–5, 2011.
23. C. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, Wired Magazine, vol. 16, no. 07, pp. 1–2, 2008.
24. L. Manovich, *Trending: The Promises and the Challenges of Big Social Data*, Debates in the Digital Humanities, pp. 460–475, 2015.
25. L. Rabhi, N. Falih, A. Afraites, and B. Bouikhalene, *Big Data Approach and its applications in Various Fields: Review*, Procedia Computer Science, vol. 155, no. August, pp. 599–605, 2019.
26. R. Kitchin, *The real-time city? Big data and smart urbanism*, GeoJournal, vol. 79, no. 1, pp. 1–14, 2014.
27. K. Ahaidous, M. Tabaa, and H. Hachimi, *Towards IoT-Big Data architecture for future education*, Procedia Computer Science, vol. 220, pp. 348–355, 2023.
28. V. Shobana and N. Kumar, *Big data - A review*, International Journal of Applied Engineering Research, vol. 10, no. 55, pp. 1294–1298, 2015.
29. S. K. Baliarsingh, S. Vipsita, A. H. Gandomi, A. Panda, S. Bakshi, and S. Ramasubbareddy, *Analysis of high-dimensional genomic data using MapReduce based probabilistic neural network*, Computer Methods and Programs in Biomedicine, vol. 195, p. 105625, 2020.
30. N. Roy-Hubara, A. Sturm, and P. Shoval, *Designing NoSQL databases based on multiple requirement views*, Data and Knowledge Engineering, vol. 145, no. February, p. 102149, 2023.
31. Y. Kumar, J. Kumar, and P. Sheoran, *Integration of cloud computing in BCI: A review*, Biomedical Signal Processing and Control, vol. 87, no. PA, p. 105548, 2024.
32. M. J. Suárez-Cabal, P. Suárez-Otero, C. de la Riva, and J. Tuya, *MDICA: Maintenance of data integrity in column-oriented database applications*, Computer Standards and Interfaces, vol. 83, no. February 2022, 2023.
33. Y. Zhou and F. Gao, *Smart batch process: The evolution from 1D and 2D to new 3D perspectives in the era of Big Data*, Journal of Process Control, vol. 130, no. August, p. 103088, 2023.
34. H. Hazem, A. Awad, and A. Hassan Yousef, *A distributed real-time recommender system for big data streams*, Ain Shams Engineering Journal, vol. 14, no. 8, p. 102026, 2023.
35. M. Gribaudo, M. Iacono, and M. Kiran, *A performance modeling framework for lambda architecture based applications*, Future Generation Computer Systems, vol. 86, pp. 1032–1041, 2018.
36. A. Subrahmanyam, *Big data in finance: Evidence and challenges*, Borsa Istanbul Review, vol. 19, no. 4, pp. 283–287, 2019.
37. A. Jabbar, P. Akhtar, and S. Dani, *Real-time big data processing for instantaneous marketing decisions: A problematization approach*, Industrial Marketing Management, vol. 90, no. September 2019, pp. 558–569, 2020.
38. N. Ellili et al., *The applications of big data in the insurance industry: A bibliometric and systematic review of relevant literature*, Journal of Finance and Data Science, vol. 9, no. February, p. 100102, 2023.
39. Y. Lian, G. Zhang, J. Lee, and H. Huang, *Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles*, Accident Analysis and Prevention, vol. 146, no. April, p. 105711, 2020.
40. D. Bianchini, V. De Antonellis, and M. Garda, *A big data exploration approach to exploit in-vehicle data for smart road maintenance*, Future Generation Computer Systems, vol. 149, pp. 701–716, 2023.
41. H. Tamiminia, B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco, *Google Earth Engine for geo-big data applications: A meta-analysis and systematic review*, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 164, no. May, pp. 152–170, 2020.
42. Y. Li and X. Zhai, *Review and Prospect of Modern Education using Big Data*, Procedia Computer Science, vol. 129, pp. 341–347, 2018.
43. M. Abuljadail, A. Khalil, S. Talwar, and P. Kaur, *Big data analytics and e-governance: Actors, opportunities, tensions, and applications*, Technological Forecasting and Social Change, vol. 193, no. May, p. 122612, 2023.
44. J. Lyu, A. Khan, S. Bibi, J. H. Chan, and X. Qi, *Big data in action: An overview of big data studies in tourism and hospitality literature*, Journal of Hospitality and Tourism Management, vol. 51, no. March, pp. 346–360, 2022.
45. J. Zhang, D. Wolfram, and F. Ma, *The impact of big data on research methods in information science*, Data and Information Management, vol. 7, no. 2, p. 100038, 2023.
46. L. Tang, J. Li, H. Du, L. Li, J. Wu, and S. Wang, *Big Data in Forecasting Research: A Literature Review*, Big Data Research, vol. 27, p. 100289, 2022.
47. H. M. Rai, *Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data*, Applied Intelligence, 2021.
48. G. Singh, D. Schulthess, N. Hughes, B. Vannieuwenhuyse, and D. Kalra, *Real world big data for clinical research and drug development*, Drug Discovery Today, vol. 23, no. 3, pp. 652–660, 2018.

49. K. Y. He, D. Ge, and M. M. He, *Big data analytics for genomic medicine*, International Journal of Molecular Sciences, vol. 18, no. 2, pp. 1–18, 2017.
50. K. Batko and A. Ślęzak, *The use of Big Data Analytics in healthcare*, Journal of Big Data, vol. 9, no. 1, 2022.
51. C. H. Lee and H. J. Yoon, *Medical big data: Promise and challenges*, Kidney Research and Clinical Practice, vol. 36, no. 1, pp. 3–11, 2017.
52. S. Aminizadeh et al., *The applications of machine learning techniques in medical data processing based on distributed computing and the Internet of Things*, Computer Methods and Programs in Biomedicine, vol. 241, no. March, p. 107745, 2023.
53. J. Lawson, E. M. Ghanaim, J. Baek, H. Lee, and H. L. Rehm, *Aligning NIH's existing data use restrictions to the GA4GH DUO standard*, Cell Genomics, vol. 3, no. 9, p. 100381, 2023.
54. M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara, and K. Sadatdiynov, *A survey of data partitioning and sampling methods to support big data analysis*, Big Data Mining and Analytics, vol. 3, no. 2, pp. 85–101, 2020.
55. L. Pereira and J. Chang, *RSNA Pneumonia Detection Challenge – Winning Model Documentation*, pp. 1–9, 2021.
56. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, *dataset noisy labels ChestX-ray8 multidisease ChestX-ray14*, pp. 2097–2106.
57. D. Demner-Fushman et al., *Preparing a collection of radiology examinations for distribution and retrieval*, Journal of the American Medical Informatics Association, vol. 23, no. 2, pp. 304–310, 2016.
58. S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, *Two public chest X-ray datasets for computer-aided screening of pulmonary diseases.*, Quantitative imaging in medicine and surgery, vol. 4, no. 6, pp. 475–7, 2014.
59. S. Ryoo and H. J. Kim, *Activities of the Korean Institute of Tuberculosis*, Osong Public Health and Research Perspectives, vol. 5, no. S, pp. S43–S49, 2014.
60. K. Sugimoto et al., *Assessment of arterial hypervascularity of hepatocellular carcinoma: Comparison of contrast-enhanced US and gadoxetate disodium-enhanced MR imaging*, European Radiology, vol. 22, no. 6, pp. 1205–1213, 2012.
61. K. Clark et al., *The cancer imaging archive (TCIA): Maintaining and operating a public information repository*, Journal of Digital Imaging, vol. 26, no. 6, pp. 1045–1057, 2013.
62. N. Dhungel, G. Carneiro, and A. P. Bradley, *Deep learning and structured prediction for the segmentation of mass in mammograms*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9349, pp. 605–612, 2015.
63. R. J. Nowling et al., *Classification before segmentation: Improved u-net prostate segmentation*, 2019 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2019 - Proceedings, pp. 1–4, 2019.
64. S. Bakas et al., *Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge*, 2018.
65. E. Hussain, L. B. Mahanta, H. Borah, and C. R. Das, *Liquid based-cytology Pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions*, Data in Brief, vol. 30, 2020.
66. A. A. A. Setio et al., *Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge*, Medical Image Analysis, vol. 42, pp. 1–13, 2017.
67. M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, *Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images*, Proceedings - International Conference on Image Processing, ICIP, pp. 3144–3148, 2018.
68. J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, *Pap-smear Benchmark Data For Pattern Classification*, Proc. NiSIS 2005, Albufeira, Portugal, no. January 2006, pp. 1–9, 2005.
69. N. C. F. Codella et al., *Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)*, Proceedings - International Symposium on Biomedical Imaging, vol. 2018-April, no. Isbi, pp. 168–172, 2018.
70. J. Sivaswamy, S. R. K. Gopal, D. Joshi, M. Jain, U. Syed, and A. E. Hospital, *DRISHTI-GS: RETINAL IMAGE DATASET FOR OPTIC NERVE HEAD (ONH) SEGMENTATION IIIT, Hyderabad, India*, pp. 53–56, 2014.
71. F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, *RIM-ONE: An open retinal image database for optic nerve evaluation*, Proceedings - IEEE Symposium on Computer-Based Medical Systems, no. July, 2011.
72. P. Porwal et al., *Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research*, Data, vol. 3, no. 3, pp. 1–8, 2018.
73. K. Mohamad Almustafa, A. Kumar Sharma, and S. Bhardwaj, *STARC: Deep learning Algorithms' modelling for STructured analysis of retina classification*, Biomedical Signal Processing and Control, vol. 80, no. P2, p. 104357, 2023.
74. A. A. Almazroa et al., *Retinal fundus images for glaucoma analysis: the RIGA dataset*, no. March 2018, p. 8, 2018.
75. P. Chowdhury, M. R. Islam, M. A. Based, and P. Chowdhury, *Transfer Learning Approach for Diabetic Retinopathy Detection using Efficient Network with 2 Phase Training*, 2021 6th International Conference for Convergence in Technology, I2CT 2021, no. April, 2021.
76. B. Graham, *Kaggle Diabetic Retinopathy Detection competition report*, Kaggle, pp. 1–9, 2015.
77. S. Balocco et al., *Standardized evaluation methodology and reference database for evaluating IVUS image segmentation*, Computerized Medical Imaging and Graphics, vol. 38, no. 2, pp. 70–90, 2014.
78. M. Mohammadi, R. V Pawar, and P. S. Dhabe, *Heart Diseases Detection Using Fuzzy Hyper Sphere Neural Network Classifier*, no. November, 2010.
79. O. Bernard et al., *Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved? The dataset contains data from 150 multi-equipments CMRI recordings with reference measurements and classification*, pp. 1–12, 2018.
80. F. Garcea, A. Serra, F. Lamberti, and L. Morra, *Data augmentation for medical imaging: A systematic literature review*, Computers in Biology and Medicine, vol. 152, no. July 2022, p. 106391, 2023.
81. M. H. Hesamian, W. Jia, X. He, and P. Kennedy, *Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges*, Journal of Digital Imaging, vol. 32, no. 4, pp. 582–596, 2019.

82. S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, *Image Data Augmentation for Deep Learning: A Survey*, 2022.
83. C. Shorten and T. M. Khoshgoftaar, *A survey on Image Data Augmentation for Deep Learning*, Journal of Big Data, vol. 6, no. 1, 2019.
84. A. Azizi, M. Azizi, and M. Nasri, *Artificial Intelligence Techniques in Medical Imaging: A Systematic Review*, International Journal of Online and Biomedical Engineering (iJOE), vol. 19, no. 17, pp. 66–97, 2023.
85. A. S. Panayides et al., *AI in Medical Imaging Informatics: Current Challenges and Future Directions*, IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, pp. 1837–1857, 2020.
86. Z. Liu, Q. Lv, Z. Yang, Y. Li, C. H. Lee, and L. Shen, *Recent progress in transformer-based medical image analysis*, Computers in Biology and Medicine, vol. 164, no. March, p. 107268, 2023.
87. S. R. Karanam, Y. Srinivas, and M. V. Krishna, *WITHDRAWN: Study on image processing using deep learning techniques*, Materials Today: Proceedings, no. January 2021, 2020.
88. A. Holzinger, K. Keiblinger, P. Holub, K. Zatloukal, and H. Müller, *AI for life: Trends in artificial intelligence for biotechnology*, New Biotechnology, vol. 74, no. February, pp. 16–24, 2023.
89. J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, *A survey of machine learning for big data processing*, Eurasip Journal on Advances in Signal Processing, vol. 2016, no. 1, 2016.
90. S. Bag, P. Dhamija, R. K. Singh, M. S. Rahman, and V. R. Sreedharan, *Big data analytics and artificial intelligence technologies based collaborative platform empowering absorptive capacity in health care supply chain: An empirical study*, Journal of Business Research, vol. 154, no. September 2022, p. 113315, 2023.
91. N. Mizik and D. Hanssens, *Machine Learning and Big Data*, Handbook of Marketing Analytics, pp. 253–254, 2018.
92. A. Azeroual, M. Chala, B. Nsiri, R. O. Haj Thami, I. Nassar, and B. Benaji, *Artificial Intelligence Applied to COVID-19 Lung Infection Segmentation from CT Images*, International Journal of Engineering Trends and Technology, vol. 71, no. 7, pp. 124–131, Jul. 2023.
93. Adekunle Oyeyemi Adeniyi, Jeremiah Olawumi Arowoogun, Chioma Anthonia Okolo, Rawlings Chidi, and Oloruntoba Babawarun, *Ethical considerations in healthcare IT: A review of data privacy and patient consent issues*, World Journal of Advanced Research and Reviews, vol. 21, no. 2, pp. 1660–1668, 2024.
94. A. Veltman, D. W. J. Pulle, and R. W. De Doncker, *The Transformer*, Power Systems, no. Nips, pp. 47–82, 2016.
95. M. M. Rathore, S. A. Shah, D. Shukla, E. Bentafat, and S. Bakiras, *The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities*, IEEE Access, vol. 9, pp. 32030–32052, 2021.
96. S. Rana and A. K. Gautam, *Online and Biomedical Engineering*, International Journal of Online and Biomedical Engineering, vol. 19, no. 9, pp. 122–130, 2023.
97. I. E. Agbehadji, B. O. Awuzie, A. B. Ngowi, and R. C. Millham, *Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing*, International Journal of Environmental Research and Public Health, vol. 17, no. 15, pp. 1–16, 2020.