Mesopotamian journal of Big Data
Vol. **(2024)**, 2024, **pp**. 11–22
DOI: https://doi.org/10.58496/MJBD/2024/002 ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

Research Article

Leveraging AI and Big Data in Low-Resource Healthcare Settings

Ahmed Hussein Ali [1], Saad Ahmed Dheyab [2], Abdullah Hussein Alamoodi [3, *], Aws Abed Al Raheem Magableh [4], Yuantong Gu [5]

[1] *College of Education, Aliraqia University, Baghdad, Iraq.*

[2] *College of Engineering, University of Information Technology and Communications, Baghdad, Iraq*

[3] *Universiti Tenaga Nasional, Kajang, Malaysia.*

[4] *Yarmouk University, Irbid, Jordan.*

[5] *Queensland University of Technology, Brisbane, Australia.*

ARTICLE INFO

Article History

Received 20 Nov 2023
Accepted 23 Jan 2024
Published 14 Feb 2024

Keywords

Big data

artificial intelligence

Poor healthcare

Spark

Machine learning

ABSTRACT

Big data and artificial intelligence are game-changing technologies for the underdeveloped healthcare industry because they help optimize the entire supply chain and deliver more exact patient outcome information. Machine learning approaches that have recently seen more growing popularity include deep learning models that have brought revolution within the healthcare system in the previous years due to more complicated data compared to previous years . Machine learning is an essential data analysis procedure to describe efficient and effective methods to extract hidden information from large amounts of data that it would take logical analytics too long to manage. Recent years have seen an expansion and growth of advanced intelligent systems that have been able to learn more about clinical treatments and glean untapped medical information emanating from vast quantities of data when it comes to drug discovery and chemistry. The aim of this chapter is, therefore, to assess which big data and artificial intelligence approaches are prevalent in healthcare systems by investigating the most advanced big data structures, applications, and industry trends today available. First and foremost, the purpose is to provide a comprehensive overview of how the artificial intelligence and big data models can allocation in healthcare solutions fill the gap between machine learning approaches' lack of human coverage and the healthcare data's complexity. Moreover, current artificial intelligence technologies, including generative models, Bayesian deep learning, reinforcement learning, and self-driving laboratories, are also increasingly being used for drug discovery and chemistry . Finally, the work presents the existing open challenges and the future directions in the drug formulation development field. To this end, the review will cover on published algorithms/automation tools for artificial intelligence applied to large scale-data in the case of healthcare .

1. INTRODUCTION

It is estimated that a considerable amount of data is generated in the healthcare business each year from a variety of sources, such as medical records and patient-related information. Every item of information should be digitized in today's digital environment, because it is more efficient [1, 2]. The most efficient and least expensive data analysis methods should be used to produce the best potential results while solving new challenges with big data[3]. As well as the private sector, the government sector generates a significant volume of information every day. It is therefore necessary to develop a solution that can deal with this huge data stream in real time without compromising on overall efficiency. Because of this, citizens will be able to achieve greater outcomes. Big data, through the application of various machine learning algorithms, aids in the supply of meaningful judgments by detecting data patterns and linkages between different data sets. Over the last few years, the health-care industry has undergone a paradigm shift, transitioning from a "clinicalcentric" care model to a more consumer-oriented "patient-centric" care model. There has been a transition away from a provider-centric strategy and toward a patient-centric approach, with the emphasis shifting from a provider-centric experience to a patient-centric experience in the healthcare industry[4]. In the conventional paradigm, which placed hospitals and healthcare practitioners at the center of the system, there was a vast amount of data to be found. This information was readily available in the form

*Corresponding author. Email: alamoodi.abdullah91@gmail.com

of medical records and files. Volume, diversity, value, velocity, and truthfulness are only some of the characteristics that distinguish big data from other types of information[5]. The five Vs of big data are depicted in Figure 1.

Volume

The amount of data that has been generated and saved is referred to as its volume. When it comes to big data, the amount of the data is what defines whether or not it is regarded such.

Variety

It has everything to do with the type and nature of the data. Text, photos, audio, video, and fusion data are all forms of data that combine numerous different types of information.

Velocity

In computing, the velocity of data is the rate at which it is generated and processed. When compared to tiny data, big data is produced on a more frequent basis[6]. The ability to move quickly is not only essential for big data generation; it is also required for any operations that use large amounts of data.

Veracity

It is a more comprehensive definition of big data that takes into account the data's overall quality. The veracity of data refers to the quality or trustworthiness of the information.

Value

It is the monetary value of the information that is being sought that we are referring to when we use the term "value."

Fig. 1. The Big Data Five Vs.

Nowadays, big data begins to be discussed in every possible place[7, 8] . Among the leading companies, the understanding of its importance as a valuable strategic asset for their work has also begun to be distributed. Therefore, probably, we should start with the clarification of the concept of everyone's agreement. In general, it can be stated that big data is vast and complex data, which, accordingly, is difficult or impossible to analyze from the standpoint of the storage and processing technologies used. According to practitioners, big data is defined as data volumes from 30 to 50 terabytes to petabytes and more. In recent years, the healthcare sector has developed into largest and most promising sectors, including research on diseases, pharmacology, new ways of treatment, technological innovation on diagnostic and surgery, etc. The industry has shifted from the early age local speciality hospitals controlled by few specialists to ultra modern big clusters multispecialty hospitals owned and managed by corporate. Healthcare management focuses more on patient-driven treatment than the diseases. The delivery model for health care is based on value than volume. The advancement of computing facilities and the access to new quantifiers of processors has clustered and stocked massive data. The rhythm with which this data has been produced and accumulated within the last two ages is amassing. More than 90% of the world's data was produced in the last two ages[12] . The advent of Electronic Health Records (EHR), personalized medicine, and administrative data has played a huge part in ensuring that health data is being recorded in vast and varied quantities according to reports. Big data is a complex subject and it is therefore hard to give a simple and all-embracing explanation but some key characteristics of the data have been identified that represent both major potential and potential challenges. These data are known to contain a large amount of the data that is involved within these datasets, the rate at which the data is being produced and collected, is generated and acquired, and the type of data that can be used is generated as well (data variety) . Other definitions put a fourth of veracity, emphasizing the importance that the data's quality is extensively reviewed. We are going to see in this chapter, however, that it is not a decent feature for the information as it might seem. As we want to propose in more detail on, nevertheless, big data is much more than just the datasets. Nonetheless, as frequently mentioned, big data annotates to sets of materials that have been used initially for one aim, now are used, collectively, in new secondary analyses or the combination of new datasets collected for new aims or for the data generated in the course of ordinary business, by far the highest part of which is assembled and stocked in an autonomous or pseudonymous manner.

2. BIG DATA AND HEALTH

The issue of Big Data's applicability in biomedical and health sciences has become a subject of extensive discussions in the past decade . These results open up great possibilities in diagnosing, treating, and preventing various diseases and creating new medicines for better general health outcomes. Apart from the evident reveal in sensitivity with this data disclosure, there is also an unveiling of people needing intervention or therapy as a result of their increased vulnerability to the threat from the previous concern . Genome sequencing is one of the technological evolutions in the health and bio-medical sector, combined with the digitalization of imaging, the growth and development of huge patient data repositories, the ever-increasing speed of bio-medical knowledge, and the increasing patient engagement, involving the collection of personal data, in treatment form the core contributors of the Big Data revolution to the health sector. Big data is applied by various companies in today's corporate world to learn what their consumers want, or to save more money, as well as to reduce risk, control fraud, enhance operational productivity, and assist make quicker and better decisions, which include services and product creation among other things. Many studies conclude that traditional systems have failed to effectively capture and analyse comprehensions of consumers, competitors, and products, a challenge that big data addresses . The current systems fail to find out the cheapest and best use of the available capacities. Its features, life cycle, and uses are among the topics that this part will discuss. The Big Data Life Cycle describes[1] the set of actions that are included in the conventional life cycle model when working with datasets that have a high velocity, variety, and volume. A sophisticated technique, because it stores large volumes of data in a meaningful and compact format, is considered to be a data warehouse. Each stage of the Big Data Life cycle is broken down into nine steps, which include the following steps: business case review; data identification; acquisition and filtering; data extraction; validation and cleansing; data aggregation and representation; statistical analysis; statistical visualization; and data management. As defined by the definition, "big data" refers to a large amount of data that researchers and teachers can use in medical education for a variety of processes such as developing, planning, estimating, and analyzing, as well as to ensure the delivery of teacher's activities. "Big data" is also referred to as "big information" or "big data mining." A wide range of applications are possible by combining data analysis and extraction techniques, information and knowledge representation, evaluating the cognitive strength of persons, and recognizing visual patterns. The use of visual analytics is getting more and more widespread. The medical data component comprises all of the information essential for recognizing a specific illness at the most fundamental level. These characteristics are static values that are exhibited in huge data systems, such as the ones used by the government. These components will be combined in order to perform a specific purpose. In addition to supporting doctors in administering the most appropriate treatment for a single patient suffering from a certain condition, these data analytics are useful in the identification of a specific disease among humans.

2.1 Spark

A big data platform for big data processing and analysis, Spark is an in-memory cluster computing platform that is designed for processing and analyzing enormous volumes of data in a distributed setting[15, 16]. For the goal of processing large amounts of information quickly and efficiently, it provides a basic programming interface that enables an application developer to efficiently leverage the processing power, memory, and storage resources available across a cluster of machines[17]. MapReduce provides a more complex programming model, whereas Spark provides a more straightforward programming model[18, 19]. Making a distributed data processing program with Spark is substantially easier than making the same application with another framework such as MapReduce, for example. When compared to Hadoop MapReduce, Spark is orders of magnitude faster. The speed of this method can be hundreds of times faster than that of Hadoop MapReduce if the data is tiny enough to fit in memory. The Apache Spark programming model, which processes data that does not fit in memory, is up to ten times faster than the Hadoop MapReduce programming model. When dealing with large datasets, processing speed is important to the success of the project. When a data processing operation takes days or hours, it causes decision-making to be put on hold until the problem is resolved. It has a negative impact on the monetary value of data. Extending the time it takes to perform the same processing to a faster rate opens the door to an endless number of novel applications and possibilities. It is now possible to develop new data-driven applications that were previously impossible to develop in the past. A variety of factors contribute to Spark being much faster than Hadoop MapReduce in various situations. First and foremost, it enables in-memory cluster computing. Second, it has an intuitive execution engine that is both creative and user-friendly. When compared to previous methods, Spark's in-memory cluster computing capabilities provide an order of magnitude larger gain in overall performance. While reading data from memory, the sequential read throughput is 100 times greater than the sequential read throughput when reading data from a hard drive, when comparing the two methods of reading data. Consequently, data may be retrieved from memory at a rate that is 100 times faster than data retrieved from disk. When a small dataset is read and processed by an application, the difference in read speed between disk and RAM may not be immediately apparent to the end user due to the small dataset size. In spite of this, when a program reads and processes terabytes of data, I/O latency (the time it takes to load data from disk to memory) becomes a significant factor to the total time necessary to finish a job.

Spark [20, 21]allows an application to cache data in memory so that it can be processed by other apps at a later time. Thus, the amount of disk I/O that the application performs can be reduced by the application. Using MapReduce, tasks are constructed that read input from disk, process it, and then write the results back to disk in a sequential order, much like a traditional data processing pipeline. As a result, a complicated data processing program that is implemented using MapReduce may receive data from disk and write data to disk several times during the course of its execution. Given the fact that Spark allows data to be cached in memory, an identical application that is designed using Spark only needs to read data from disk a single time. Cached data is data that has been temporarily stored in memory so that further operations can be conducted directly on the data that has been cached there. This means that applications can reduce I/O latency, which, as previously described in this blog article, can be a significant contributor to overall task execution time.

More than 80 data processing operators are available in Spark, which makes it a strong tool for data scientists looking to build large data applications. As defined by the IEEE, streaming data is characterized as a large amount of data created by thousands of sensors and transmitted to a central point all at the same time. In order to extract usable information from this data, it must be processed on a record-by-record basis, which takes time. A variety of filtering and correlation options are available, as well as sampling and aggregate options depending on the needs of the company. Presented in a manner that is applicable to a wide range of company and consumer scenarios, this study is a valuable resource. Using social media streams, the industry may track the most popular products among consumers based on similar comments and likes, or it can track consumer feelings based on specific scenarios and then take rapid smart steps to remedy the situation. In order to further enhance the insights derived from data collecting, stream processing techniques are being used more and more frequently. It is possible to train machine learning algorithms on historical data and then apply them to streaming data by utilizing the Apache Spark Machine Learning Library (MLlib), which is provided as part of the Apache Spark project (see Figure 2 Apache Spark). There are only a few streaming machine learning algorithms that are capable of both training from streaming data and testing the model on streaming data at the same time, and these algorithms are quite rare indeed. For example, streaming K-Means is an example of a streaming approach since it dynamically updates the cluster as new information is introduced into it. In addition to streaming techniques such as Streaming Linear Regression, which changes the parameters of a model in real time as the data is being collected and processed, there is also Streaming Linear Regression. The following instance illustrates the use of Streaming Linear Regression on streaming data and the implementation of this technique.

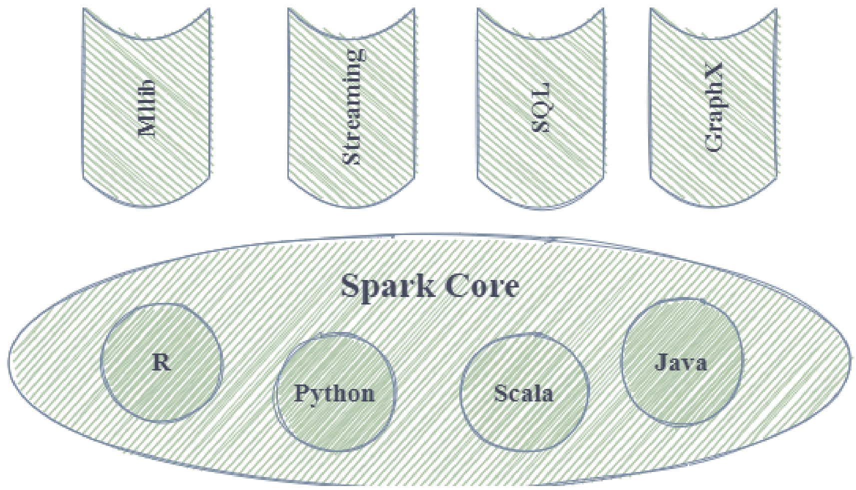

Fig. 2. Apache Spark

2.2 Hadoop

Big data processing problems can be solved with the help of Hadoop[22, 23], which is a distributed and scalable framework. Hadoop is a distributed computing system developed by Doug Cutting and Mark Cafarella in the Java programming language. The fact that it can be installed on a cluster of commodity hardware and that it scales horizontally on distributed systems makes it an excellent choice for cloud computing. Programming is simple. Originally based on a Google research article, Hadoop was developed and put into operation. Hadoop is a cost-effective solution due to its ability to run on commodity technology and its ability to scale. Fault-tolerance will be a concern as long as we are working with commodity hardware. Hadoop, on the other hand, is a fault-tolerant data storage and computation system, and it is this fault-tolerant feature that has contributed to Hadoop's widespread adoption. Hadoop is comprised of two pieces, which are illustrated in Figure 3. Figure 3: Hadoop is composed of two parts. First and foremost among these components is the Hadoop Distributed File System (HDFS)[24, 25]. MapReduce is the second component, and it is what we'll be talking about today. HDFS is utilized for distributed data storage in this example, while MapReduce is used to do computation on the data that has been saved in HDFS by the Hadoop cluster. Massive amounts of data can be stored in HDFS, which operates on a distributed and fault-tolerant architecture.

Designed to run on low-cost commodity hardware, HDFS[26] is written in Java and runs on a Java virtual machine. It was sparked by a Google search.

The Google File System is the subject of a research paper now being written (GFS). Writing once and reading many times is a strategy that works well when dealing with large amounts of information at the same time. HDFS is made up of two components: the NameNode and the DataNode, which work together to store data. It is the responsibility of Java daemon processes to ensure that these two components are operational. Typically, a NameNode, which is responsible for storing the metadata of files that are disseminated throughout a cluster, serves as the master for a large number of DataNodes in a distributed computing environment. HDFS divides a large file into small blocks, each of which is saved on a distinct DataNode in the cluster. HDFS is a distributed file system. The actual file data blocks are kept in DataNodes, which are nodes in the file system. HDFS includes a set of commands that are quite similar to those found in a Unix shell environment. When dealing with large files, we can take advantage of the Java file system API provided by HDFS to operate at a more granular level than would otherwise be possible. Fault tolerance is done by storing several copies of the same data on different servers. A single thread process can be used to access HDFS files, or several threads can be used to access the same file in a parallel mode. An extremely useful utility known as distcp is included with HDFS and is often used to transfer data in parallel from one HDFS system to another. HDFS additionally includes a number of helpful utilities that are not included in the base distribution. Parallel map jobs are used to transfer data from one location to another throughout the copying procedure. With Hadoop, you may run MapReduce files in the form of map reduce jobs, which follow the MapReduce programming paradigm (Key, Value). In order to convert these input text files into the form of maps, it is essential to transmit the values from the input split to the mapper through another pre-defined interface called Record Reader (Key, value). The Record Reader is invoked repeatedly on the input until the entire InputSplit file has been processed, with each invocation of the RecordReader resulting in another call to the map() method of the Mapper, which then stores the intermediate result in a variable called idx.

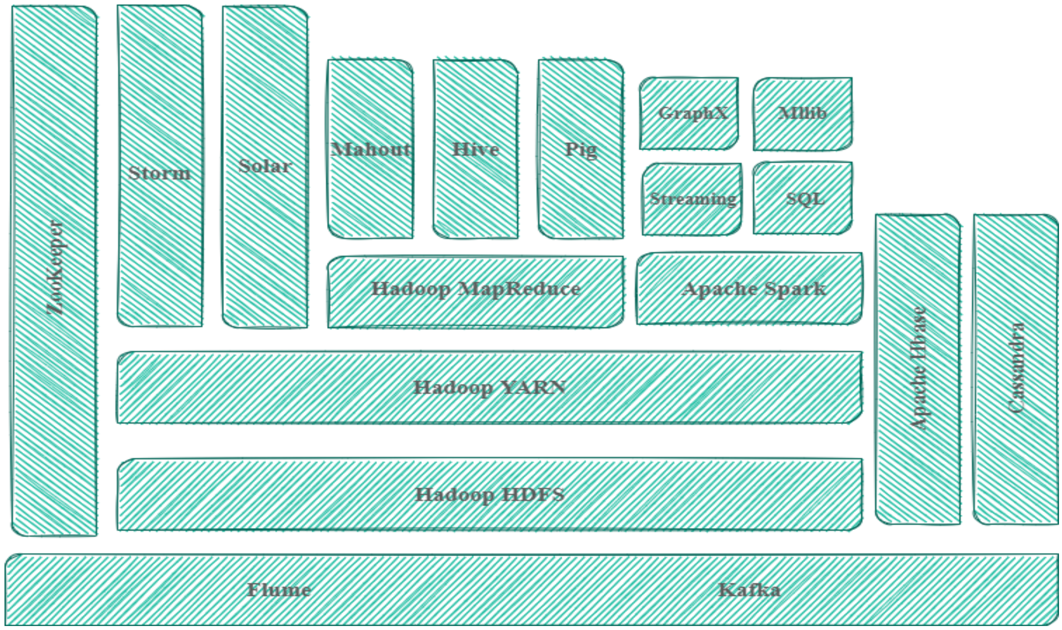

Fig. 3. Apache Hadoop Ecosystem

2.3 Health and Big Data in Low- and Middle-Income Countries

A poll was conducted utilizing data from studies that specifically focused on the use of Big Data for health delivery in LMICs. A majority of the time when health delivery occurs in those contexts is via other mechanisms as convenient as clinical care, known as vertical programs . High numbers of the CDs employed in the health chain are community health workers, mostly unregistered and containing no formal qualifications. A few of the examples of such several programs include those for TB, HIV, malaria, and so forth. The first significant benefit harvested from the two studies is that some of the immediate and logistical challenges that were formerly insurmountable are now an option. For instance, new kinds of data collection means, like a smartphone, a tablet, or a portable computer, real-time data capture involving devices have become possible ranging from a BP machine to a glucometer, an electrode machine, or even a portable ultrasonograph that retrieves data to a central server every minute. Low- and middle-income countries commonly leapfrog some of the procedures and advancements that have already occurred in wealthy countries. In Sub-Saharan Africa and a few other countries, for example, the vast penetration of the phone linkages has been accompanied by other resources indicating good growth. A few studies and product releases have connected cell phone use with good growth and health indicators in LMICs. A cell phone's rapid penetration can be attributed to the lack of well-developed mobile infrastructures, which results in a rapid rollout at the expense of fixed-line equipment that has characterized the telecom industry in developing nations until now, which will happen anyway.

Because there was no established pattern or recognized possession asset to contend with, it was possible to work more efficiently this way. In other terms, access to many technologies has progressed much sooner and more unexpectedly than anybody could have predicted and will demand far more years of acceptance than any other technology in completely industrialized nations or regions.

It is also hoped that the successful use of Big Data analytics in low- and middle-income countries will ensure high-quality healthcare delivery, reveal illness risk factors early on, and promptly identify In the current service delivery, governance, and data coordination are all inadequate action, even the smallest change in each may have a significant impact on gains. As a result, the most limited resources are use productively. In these resource-constrained environments health systems are centered on specific diseases, while the integrative nature of Big Data may facilitate a shift to a more integrated, horizontal approach to disease pattern investigation, prevention and control, as well as understanding the root causes of poor health and responses to it. Although it remains critical to have access to safe water, nutritious food, and adequate sanitation, Big Data analytics may be as successful in promoting human development as incline health. Even if the current and already-developed infrastructure and abilities could be usefully utilize in this endeavor, the influence of good health on development and vice versa grows in a positive way. At the global level, it is imperative to address the present inherently weak governance of global health in order to harness these potential benefits entirely. Data gathering must be carefully informed, thoughtful, and appropriately fund, and it must also undergo appropriate oversight and stewardship by the best authorities. The United Nations established the Global Pulse initiative in 2009, which purpose to "fast investigate, develop, and scale up Big Data technology adoption for sustainable development and humanitarian initiatives." The United Nations Development Programme pays for the initiative. Several health-related projects have been part of the program's comprehensive reach. There have been several such projects, and some of the most ones are: An online dashboard that use in Uganda to monitor mother-to-child HIV prevention project implementation, with real-time indicator data to fill it collects from health centers around the country. So, when data showed that there were bottlenecks in the move to Option B+ treatment. This treatment was generally available and offer for all HIV-infected expectant mothers, regardless of whether they had a CD4 t-lymphocyte count test. The data available indicates a leakage of the medicine into the informal market.

- Primarily, by the use of a database of more than 280,000 tweets that were sent between January 2019 and December 2020, this study was aimed at capturing social media's popular opinion about immunization across the world. Utilizing content analysis and filters, relevant tweets were found; then the revelation of this experiment is that social media was being utilized to disseminate important information to immunization, and the distribution could be monitored in real time. Share only information that was pertinent and accurate, the knowledge of a core of twitter influencers that could be used for rapid response communication if necessary, and who could then share accurate and relevant tweets, was significant.
- So, it is possible to draw a meaningful picture of public awareness and public sentiment using social media and news content analysis in the Republic of India, Kenya, Nigeria, and Pakistan. This could be performed based on the data taken from Twitter and Facebook, and their messages and posts, as well as the information received from the traditional media. There is a link between the growth of the content analyzed and certain key events . In order to identify the crucial influencers within social networks, it was possible to use network analysis and to concern demographic data on users. Overall, it is expected that these efforts would help to discover the degree to which social media monitoring can be efficient in terms of understanding the public sentiment.
- Using the data from the Facebook pages and U-report platform of the UNICEF in Uganda from 2018 to 2020, the approach made it possible for the researchers to analyze the public attitudes towards teenage pregnancy and contraception. In order to ensure that the messages were related to contraception and family planning, the Growth Hacking approach aimed to anonymize the Facebook data by filtering the posts without identifying the sources . As intended, the information was compiled in a user-friendly interactive dashboard that is presented to general audience at http://familyplanning.unglobalpulse.net/uganda/ . This application was able to provide real-time data on the change in public opinion on family planning and contraception, informing any public health program that involved these concerns in the future.
- Alternatively, we might be more likely to gain a more detailed comprehension of public perceptions of cleanliness through the use of social media data and an analysis of the findings on a social media data analytics platform. Generally, data volume trends, influencers and pertinent hashtags were recognized and reported, and then all analyzed.

The findings of this study demonstrate that it is feasible to monitor changing social media discussions on sanitation by monitoring baseline indicators over time, thereby assess the extension and efficacy of educational campaigns, especially those that prompt the general public to participate more actively in the conversations around sanitation. Seasonal mobility patterns were also analyzed using mobile phone data from Senegal. The data was anonymized, and the location of individuals was used for statistical and comparative reasons over time, in that way, showing the seasonal mobility patterns over the year. Moreover, it was possible to discriminate between daily motions and month-long movements. The narrowing down on the individuals' locations and seasonal mobility has the potential of being very useful in health surveillance and outbreak response, as well as resource mobilization and distribution decisions . The use of new (especially online) data sources such as social media and their correlation to epidemiological environmental data to produce geopositioned disease tailorable outbreak maps as opposed to static maps that are available and updated all the time presents new opportunities for creating more relevant tools. The current tools only provide a static overview of a disease's spread and impact, but accurately knowing where and when a disease is distributed is critical in a low and middle-income country when the current gap of understanding a disease biases funding meant the fight against global health issues.

3. AI IN POOR HEALTHCARE

The generation of large digital datasets acquired through next-generation sequencing (NGS), the application of image processing algorithms, the use of patient-related health records, the analysis of data from large clinical trials, and the prediction of disease will all have a significant impact on the future of healthcare[30, 31]. When it comes to utilizing artificial intelligence for the treatment of all cancers, oncology has been at the forefront of the advancement[32]. This includes, among other things, early identification, individualized or focused therapy based on the genetic information of the patient, and projections of future outcomes (Figure. 4). In order to collect meaningful clinical information, artificial intelligence can perform pattern recognition and complex algorithms[33]. This will reduce the number of errors involved with diagnosis and therapy. Machine learning (ML) [34]is a key technology in the field of cancer, with numerous applications in the field of precision medicine. A complex neural network may be used to generate diagnostic images and genetic analysis data, which can subsequently be utilized to predict the likelihood of a disease and the outcome of its treatment with great accuracy. Deep learning is the most frequently used artificial intelligence tool in radiomics, a field of machines that harvests diagnostic images to find malignant tumors that are not visible to the naked eye. Deep learning is the most frequently used artificial intelligence technique in radiomics. Deep learning is the most widely utilized artificial intelligence technique in radiomics. It is anticipated that the combined efforts of radiomics and deep learning would lead to improved diagnostic picture processing in diagnostic imaging. The field of artificial intelligence's machine learning is a crucial component of the overall discipline. Ultimately, the purpose of machine learning is to discover new knowledge and make intelligent decisions in a wide range of scenarios. Unsupervised learning algorithms are the most common type of machine learning algorithm. They can be divided into three categories: supervised learning, semi-supervised learning, and unsupervised learning. When dealing with enormous amounts of information, it is critical to scale up machine learning algorithms. Machine learning can also be classified according to the results of a machine learning system, which is another classification. Classification, regression, clustering, and density estimation are just a few examples of the types of outputs that might be generated. Machine learning, which makes use of artificial intelligence, is able to transcend the limitations imposed by the human capacity for reasoning and for detecting correlations in enormous amounts of data. Machine learning algorithms are capable of detecting and revealing patterns buried inside vast volumes of data, and they are becoming increasingly popular. The automated approaches used in large-scale data analysis are relied upon by these companies. First and first, in order to make accurate forecasts, it is necessary to define a direction for the improvement of algorithms that are speedy, cost-effective, and efficient in real-time data processing. Large amounts of both organized and unstructured data are typically referred to as "big data," and they are particularly tough to process with the help of traditional software approaches and the database concept. Using machine learning, massive amounts of data are being processed, which is having a favorable impact on technological discoveries and value creation. Many different types of models are used in machine learning, and they all make predictions in response to a sequence of tests that are given to them. When performing a supervised learning algorithm, the input data is referred to as training data, and it is used to search for different types of examples within the feature representation assigned to specific data focuses. If you are performing a supervised learning algorithm, the input data is referred to as test data, and it is used to search for different types of examples within the feature representation assigned to specific data focuses. An strategy based on training is used to develop a model in cases when forecasting is necessary. If the forecasts are not as accurate as predicted, the model will make adjustments to the training data. Using this strategy, a target/result variable (also known as a target variable) is predicted from a certain sequence of action of indicator values, which is then used to predict other variables (self-governing factors). In accordance with this game plan of elements, we set a limit for commitments, guiding them toward the expected yields. Unlike supervised learning, which requires a named dataset and a known outcome, unsupervised learning does not require either. Machine learning systems generate a swarm of clusters, and there are no names linked with information centers that can be used to identify these clusters as they are generated. Models are constructed by utilizing intuition structures that emerge in the information. Semi-supervised learning makes use of data that is a mixture of labeled and unlabeled data in the event that such an event should occur throughout the course of the experiment. Having a requirement for a prediction problem that must take into consideration the structures that are used to sort out the data and formulate wants is achievable. Several issues in depiction are tackled, including categorization and relapse.

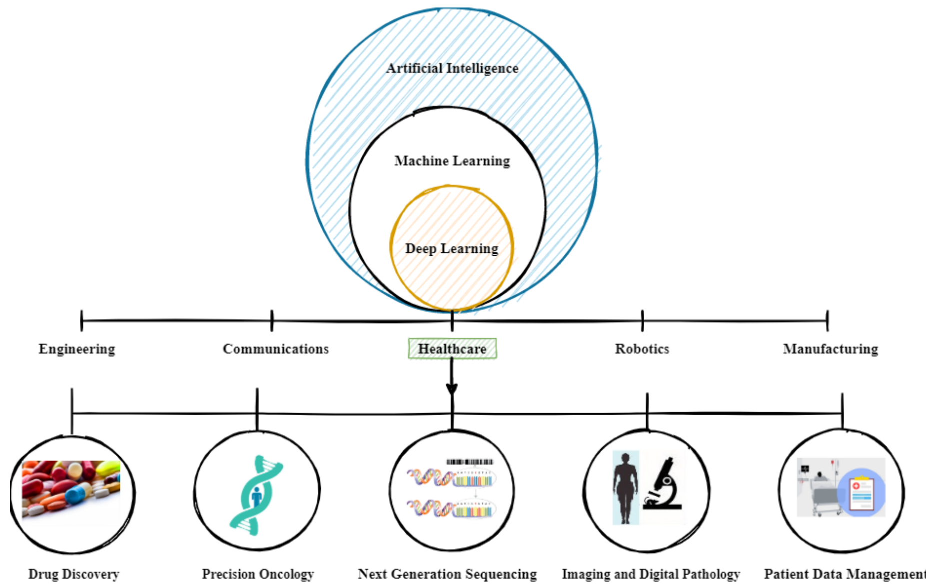

Fig. 4. Application of artificial intelligence in resource-poor healthcare

In this subset of machine learning, which is concerned with how programming operators should treat behaviors in a particular environment in order to augment some notion of aggregate reward, behavioral brain research serves as the driving force. Comparing Reinforcement Learning (RL) to conventionally supervised learning, Reinforcement Learning (RL) differs in that neither the correct information/yield sets nor the flawed activities are ever introduced nor are they ever definitively corrected. Individuals often use the most commonly used supervised and unsupervised learning methodologies when crunching information to make a business choice in order to make the best possible decision. Unsupervised learning approaches in applications such as image classification, where there are vast datasets but only a small number of named outlines to work with, are now one of the most fascinating topics in AI research today. An algorithm can demonstrate the relationship between an issue and an experience or condition, or whatever we choose to call the data, in a number of different ways. There are just a few major learning approaches or models that a technique might employ in order to gain valuable expertise in producing solutions for a variety of problem types that they are best equipped to solve, and these are listed below. This scientific categorization or method for sorting out machine learning algorithms is beneficial due to the fact that it forces you to consider the parts of the information and model readiness process and select the ones that are most appropriate for one's concern while keeping the end goal in mind to obtain the best outcome in mind.

4. BIG DATA IMPLEMENTATION IN RESOURCE-POOR HEALTHCARE AND RECOMMENDATIONS FOR THE FUTURE

This data was collected and ingested into the system during the year 2020 from more than 2000 POC devices in three African countries in order to test the system's installation. The system had also been periodically fed with huge volumes of simulated data, which was absorbed into it in order to evaluate its real-time and batch processing capabilities. It was chosen to employ a number of different data analytics methodologies. The sections that follow include details on the data collection, analysis, methodologies, and metrics that were employed in the system that was put in operation. Because POC machine data recordings are recorded and processed as events, the table comprises two types of analytics: one for temporal analytics and another for spatiotemporal analytics. This is owing to the fact that POC machine data recordings are saved and processed as events (time series observations). A key distinction between the two types of data analysis is that in spatiotemporal data analysis, location (position) and spatial relationships (such as distance, direction, and connectivity between locations of the devices) are used to group and cluster the data, whereas time is used primarily for this purpose in temporal data analysis. Because of the above-mentioned analytics, a series of interactive dashboards has been created in the deployed system that may be utilized to assist in decision-making in a range of situations. Each dashboard is made up of a mix of interactive visualizations and reports that may be tailored to meet the needs of the user. The following images show various visualizations and reports generated by some of the dashboards, as well as how certain relevant information can be inferred from the POC generated data, as seen in the following illustrations.... Lab tests accounted for slightly more than half of all tests (54.7 percent), with the remaining 27.3 and 18 percent being performed in hospitals and mobile labs, respectively. It is worth noting that, despite the fact that the number of tests performed in labs is about three times bigger than the number of tests completed in mobile labs, the percentage of failed tests in mobile labs is higher than the percentage of failed tests in laboratories (and hospitals as well). This line can be understood as follows as a result of any one or a combination of the causes listed below: (1) A non-uniform distribution of human resources is seen. We believe that in order to fully fulfill the promise of Big Data in the healthcare setting and to make healthcare more outcome-oriented, it is necessary to address and overcome the following fundamental hurdles.

- It is being explored to build national centers of excellence for big data to investigate the applications, advantages, and hazards involved with big data. These centers would be responsible for conducting research into the applications, advantages, and risks related with big data. Modern information technology (IT) systems are inadequate for handling Big Data in terms of both volume and structuring and analysis; they are also inadequate in terms of heterogeneity and the ability to employ machine learning techniques. Aside from that, dealing with enormous amounts of data needs a multidisciplinary approach that integrates a number of abilities in the field of healthcare, as previously indicated. Big Data research competence centers are needed, where the fundamental structures of the industry may be developed and built, as well as where new technologies – particularly in the area of information security – can be reviewed and compared to existing ones. An important step in this regard has been taken lately by the German Federal Ministry of Education and Research (BMBF) with the launch of an initiative to promote centers of excellence.
- Capital expenditures for information security technology of the future Given the widespread lack of confidence in the areas of data security and privacy – with the National Health Service (NHS) of the United Kingdom serving as a poor example in the domain of healthcare earlier this year – it is vital that we also engage in research to improve information security. It is not required to accomplish 100 percent IT security all of the time in order to achieve 100 percent IT security — this would be an impossible goal. To ensure that safety is continually improved, it is necessary to bring the level of protection as close as technically possible to the highest level of protection imaginable.
- It is necessary to continue efforts to build widely recognized open standards of interoperability, as well as to make additional investments in the implementation of electronic patient records, in order to achieve success. It will be impossible to achieve data sharing and interoperability without the adoption of openness (for everyone) and standardization (by everyone). Meanwhile, we must significantly accelerate the speed with which clinical information systems are introduced and integrated into practice. The results of surveys carried out around Europe have showed that there is still more work to be done in this field of research.
- With the exception of the United States, we do not yet have a viable and sustainable business model in the essential healthcare markets that supports the use of Big Data for transformation in the direction of outcome orientation, notwithstanding the fact that it is technically possible to overcome the challenges of Big Data mentioned above. The model, as well as a conceptualization of how such a model would be implemented in practice and how payment might be based on the value or outcome, are both lacking in this respect.

5. CONCLUSION

As with any new and promising technology or method, there is the risk of viewing Big Data as excessively useful and relevant to all areas of science and human behavior, as has been the case with many other new and promising technologies or methodologies. This is especially true for Big Data, which has the potential to be applied to a wide range of scientific and behavioral fields. It is necessary to consider the possibility that the sheer volume and variety of data used in Big Data analysis will give the impression that these analyses are all "objective" and "value less," or that they will be the most likely to uncover "truths," as a result of the sheer volume and variety of data used. This article provides a very brief introduction of some of the opportunities and challenges involved with Big Data, particularly in terms of ethical considerations. It then goes on to explore some of the major topics that must be addressed or studied as this area of research develops. While big data holds immense promise and opportunity for the information age, it also faces a number of challenges, including coping with the sheer volume, rapidity, and complexity of big data as well as issues about privacy and security. In this chapter, we discussed the role that big data technologies like as Hadoop and Spark play in today's data-driven world, as well as their limitations. This study provides a brief overview of some of the applications of big data technologies in resource-constrained healthcare and the public sector, as well as some of the promise and limitations associated with these applications in resource-constrained environments. The potential impact of Big data technologies in the fields of comprehensive and preventive healthcare is so enormous that it is difficult to ignore them, despite the fact that there is still a long way to go before these technologies can be fully utilized in healthcare due to data access restrictions imposed by a variety of regulatory policies and competitive pressures. Governments and resource-strapped healthcare organizations must collaborate in order to find a middle ground while simultaneously tackling the challenges that have arisen as a result of the use of Big Data in the healthcare business, according to the World Health Organization. Experience with big data and its associated life-cycle management will become increasingly crucial in the near future, and these skills will be critical to ensuring the successful implementation of big data solutions. To make full use of Big Data across disciplines, it will be necessary to solve technological issues such as scale, computational complexity, data diversity, and data security as they arise. When common standards are established and numerous technical, analytical, and ethical challenges are resolved, it is expected that the use of Big Data in resource-poor health will make significant contributions to the development of a more personalized approach to medicine and the development of smarter, adaptive health strategies. In essence, we are in the midst of the second wave of Big Data.

Conflicts Of Interest

The author asserts that there are no conflicts of interest that could have affected the study design, methodology, or results.

Funding

The absence of funding details in the author's paper suggests that the research was entirely self-funded.

Acknowledgment

The author acknowledges the institution for the intellectual resources and academic guidance that significantly enriched this research.

References

[1] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," Journal of Big Data, vol. 6, no. 1, pp. 1-25, 2019.
[2] S. Shilo, H. Rossman, and E. Segal, "Axes of a revolution: challenges and promises of big data in healthcare," Nature medicine, vol. 26, no. 1, pp. 29-38, 2020.
[3] S. Senthilkumar, B. K. Rai, A. A. Meshram, A. Gunasekaran, and S. Chandrakumarmangalam, "Big data in healthcare management: a review of literature," American Journal of Theoretical and Applied Business, vol. 4, no. 2, pp. 57-69, 2018.
[4] R. Pastorino et al., "Benefits and challenges of Big Data in healthcare: an overview of the European initiatives," European journal of public health, vol. 29, no. Supplement_3, pp. 23-27, 2019.
[5] S. Bahri, N. Zoghlami, M. Abed, and J. M. R. Tavares, "Big data for healthcare: A survey," IEEE access, vol. 7, pp. 7397-7408, 2018.
[6] L. Bote-Curiel, S. Munoz-Romero, A. Gerrero-Curieses, and J. L. Rojo-Álvarez, "Deep learning and big data in healthcare: A double review for critical beginners," Applied Sciences, vol. 9, no. 11, p. 2331, 2019.
[7] A. Mir and S. N. Dhage, "Diabetes disease prediction using machine learning on big data of healthcare," in 2018 fourth international conference on computing communication control and automation (ICCUBEA), 2018, pp. 1-6: IEEE.
[8] A. H. Ali, M. Z. Abdullah, S. N. Abdul-wahab, and M. Alsajri, "A Brief Review of Big Data Analytics Based on Machine Learning," Iraqi Journal For Computer Science and Mathematics, vol. 1, no. 2, pp. 13-15, 2020.
[9] A. H. Ali, Z. F. Hussain, and S. N. Abd, "Big Data Classification Efficiency Based on Linear Discriminant Analysis," Iraqi Journal For Computer Science and Mathematics, vol. 1, no. 1, pp. 7-12, 2020.
[10] S. N. Abd and H. R. Ibraheem, "Rao-SVM Machine Learning Algorithm for Intrusion Detection System," Iraqi Journal For Computer Science and Mathematics, vol. 1, no. 1, pp. 23-27, 2020.
[11] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," International journal of medical informatics, vol. 114, pp. 57-65, 2018.
[12] P. Kaur, M. Sharma, and M. Mittal, "Big data and machine learning based secure healthcare framework," Procedia computer science, vol. 132, pp. 1049-1059, 2018.
[13] D. V. Dimitrov, "Medical internet of things and big data in healthcare," Healthcare informatics research, vol. 22, no. 3, pp. 156-163, 2016.
[14] E. Baro, S. Degoul, R. Beuscart, and E. Chazard, "Toward a literature-driven definition of big data in healthcare," BioMed research international, vol. 2015, 2015.
[15] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Big data in healthcare: Challenges and opportunities," in 2015 International Conference on Cloud Technologies and Applications (CloudTech), 2015, pp. 1-7: IEEE.
[16] A. H. Ali, "A survey on vertical and horizontal scaling platforms for big data analytics," International Journal of Integrated Engineering, vol. 11, no. 6, pp. 138-150, 2019.
[17] A. H. Ali and M. Z. Abdullah, "Recent trends in distributed online stream processing platform for big data: Survey," in 2018 1st Annual International Conference on Information and Sciences (AiCIS), 2018, pp. 140-145: IEEE.

[18] N. D. Zaki, N. Y. Hashim, Y. M. Mohialden, M. A. Mohammed, T. Sutikno, and A. H. Ali, "A real-time big data sentiment analysis for iraqi tweets using spark streaming," Bulletin of Electrical Engineering and Informatics, vol. 9, no. 4, pp. 1411-1419, 2020.
[19] R. A. Hasan, R. A. I. Alhayali, N. D. Zaki, and A. H. Ali, "An adaptive clustering and classification algorithm for Twitter data streaming in Apache Spark," Telkomnika, vol. 17, no. 6, pp. 3086-3099, 2019.
[20] A. H. Ali and M. Z. Abdullah, "A novel approach for big data classification based on hybrid parallel dimensionality reduction using spark cluster," Computer Science, vol. 20, no. 4, 2019.
[21] A. H. Ali and M. Z. Abdullah, "A parallel grid optimization of SVM hyperparameter for big data classification using spark Radoop," Karbala International Journal of Modern Science, vol. 6, no. 1, p. 3, 2020.
[22] T. White, Hadoop: The definitive guide. " O'Reilly Media, Inc.", 2012.
[23] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in 2010 IEEE 26th symposium on mass storage systems and technologies (MSST), 2010, pp. 1-10: Ieee.
[24] S. R. Zeebaree, H. M. Shukur, L. M. Haji, R. R. Zebari, K. Jacksi, and S. M. Abas, "Characteristics and analysis of hadoop distributed systems," Technology Reports of Kansai University, vol. 62, no. 4, pp. 1555-1564, 2020.
[25] Z. Haiyun and X. Yizhe, "Sports performance prediction model based on integrated learning algorithm and cloud computing Hadoop platform," Microprocessors and Microsystems, vol. 79, p. 103322, 2020.
[26] P. Tallada et al., "CosmoHub: Interactive exploration and distribution of astronomical data on Hadoop," Astronomy and Computing, vol. 32, p. 100391, 2020.
[27] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," IEEE Reviews in Biomedical Engineering, vol. 14, pp. 156-180, 2020.
[28] I. Y. Chen, E. Pierson, S. Rose, S. Joshi, K. Ferryman, and M. Ghassemi, "Ethical Machine Learning in Healthcare," Annual Review of Biomedical Data Science, vol. 4, 2020.
[29] A. Souri, M. Y. Ghafour, A. M. Ahmed, F. Safara, A. Yamini, and M. Hoseyninezhad, "A new machine learning-based healthcare monitoring model for student's condition diagnosis in Internet of Things environment," Soft Computing, vol. 24, pp. 17111-17121, 2020.
[30] S. Kaushik et al., "AI in healthcare: time-series forecasting using statistical, neural, and ensemble architectures," Frontiers in big data, vol. 3, p. 4, 2020.
[31] H. Greenspan, R. S. J. Estépar, W. J. Niessen, E. Siegel, and M. Nielsen, "Position paper on COVID-19 imaging and AI: From the clinical needs and technological challenges to initial AI solutions at the lab and national level towards a new era for AI in healthcare," Medical image analysis, vol. 66, p. 101800, 2020.
[32] T. Panch, H. Mattie, and L. A. Celi, "The "inconvenient truth" about AI in healthcare," NPJ digital medicine, vol. 2, no. 1, pp. 1-3, 2019.
[33] A. Panesar, Machine learning and AI for healthcare. Springer, 2019.
[34] R. Shah and A. Chircu, "IoT and AI in healthcare: A systematic literature review," Issues in Information Systems, vol. 19, no. 3, 2018.