# A Review Of Big Data Analytic In Healthcare

**Nouf Ali AL Ajmi[1], Mehmet SabihAksoy[2]**
[1] PHD Student, Department of Information Systems, College of Computer and Information Sciences, King Saudi University, Riyadh 11451, Saudi Arabia
naalajmi@ksu.edu.sa

[2]Professor, Department of Information Systems, College of Computer and Information Sciences
King Saudi University, Riyadh 11451, Saudi Arabia
msaksoy@ksu.edu.sa

**Article History**: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

**ABSTRACT**:: Big data analytics (BDA) has been proven to be useful in various fields such as sports, and healthcare data. Healthcare data is increasingly being digitized today and the data collected today is coming in from modern devices. It has reached a significant volume all over the world. There is increased interest in the reuse of Big Data technology in the healthcare industry to manage massive collections of heterogeneous health datasets. BDA is of paramount importance in healthcare aspects such as patient diagnostics, fast epidemic recognition, and improvement of patient management.This paper aims to provide an overview of the BDA concept in the healthcare field and to show different aspects of BDA in healthcare like architectural framework, analysis models, tools, and applications.

**Keywords:** Healthcare analytics,Big Data, Big Data Analytic Regular Research Paper

## 1. Introduction

The healthcare industry historically has generated vast amounts of data-driven by record keeping, compliance, andregulatory requirements, and patient care.

These data were shifted from paper-based to digital format. As a result, electronic health records (EHR) appeared. An electronic health record gathers, creates, and stores the health record electronically.

Benefits of an electronic health record include a gain in healthcare efficiencies, large gains in quality and safety. Also, it lowers healthcare costs for consumers.Especially, that we have different healthcare resources are availableas shown in the nextfigure:


*Figure 1: different health data sources*

The health industry sector has been confronted by the need to manage the big data being produced by various sources.Nowadays, technology such as Big Data analyticmakes it possible to analyze health data from different resources[1][2].

### 1.1. Big DataAnalytics

Big Data is a new driver of world economic and societal changes. It refers to the vast, rich, and available data. Big Data's characteristics can be defined with 8V's: Volume, Velocity, Variety, Value, Variability, Viability,visualization, and Veracity[3].

*Table 1: Eight common Vs of Big Data.*

| Attribute | Description |
|---|---|
| Volume | Massive amount of data collected |
| Velocity | Rate at which data arrives,is stored and retrieve for processing |
| Variety | Drives structure and forms of data |
| Veracity | Trustworthiness of data |
| Variability | Changing nature of data |
| Viability | Relevance of data |
| visualization | Comprehensibility of data |
| value | Data translated into learning ,knowledge creation and/or economic gain |

In current scenarios, Big Data is associated with core technologies and various enterprises including Google, Facebook, and IBM, which extract valuable information from the huge volumes of data collected. [1][4].

This massive amount of data is produced every day by businesses and users. It creates a valuable resource that needs to utilize. One way of handling massive data is Big Data Analytics (BDA).Big Data analytics (BDA) is a method of analysis of the available huge amount of electronic data. BDA considers a knowledge discovery method[5].BDA can be defined as the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured, and unstructured data, from different sources, and in different sizes from terabytes to zettabytes.Big Data plays an active role in practices and knowledge not only regionally but globally. By globalizing data, it is made more widely accessible and providers may access new information from all regions.Big Data analytics includes various methods, including text analytics, multimedia analytics, data mining, and machine learning.

The generation of Big Data is taking place in medical healthcare units. This field is rich in data from different resources[6].

### 1.2. Big Data Analytics in Healthcare:

The Big Datasets on healthcare are contained both structured and unstructured data. Therefore comprehensive approach where structured and unstructured data coming from both clinical and non-clinical resources are exploited for a better perception is needed. Big Data scientists are always trying to discover the associations and hidden patterns among these big medical data to improved care and treatment programs[6].

Big Data analytics can be used to predict the outcomes of decisions made by physicians. An example, the outcome of a heart operation for a condition based on the patient's age, current condition, and health status. Essentially, we can say that the role of Big Data in the health sector is to manage data sets related to healthcare, which are complex and difficult to manage using current hardware, software, and management tools[1].

BDA allows early detection of diseases, which aids in clinical objectives related to achieving improved treatments and higher patient outcomes.BDAenables the appropriate use of evidence-based medicine and helps health care providers make more informed decisions. This, in turn, improves the quality of care provided to the patients[7].

In general, healthcare providers useBDA to develop decisions and processes that improve patient outcomes, reduce spending, and increase operational efficiency[8].

This paper is structured as follows, second Section describes the related work. While the third section shows BDA architectural framework. The fourth section shows BDA analysis types. The fifth section presents applications of Big Data analytics in healthcare. Challenges faced Big Data analytics in healthcare in the sixth

section. Then, the Future of Big Data Analytics in Healthcare on the seventh section. The last section is the Conclusion.

## 2. Related Work

The story of how data became big starts many years before the current buzz around Big Data. Seventy years ago, the first attempts to quantify the growth rate in the volume of data appear in 1941. According to the Oxford English Dictionary,it has popularly been known as the "information explosion" [9].

Within the time, the term 'Big Data' has been used for this. Although it is not exactly known who first used the term, most people credit John R.Mashey (who at the time worked at Silicon Graphics) for making the term popular[10].

Later, Big Data has been used in many fields. Here is, an overview ofstudies that relatedBig Data to healthcare analytics:

McGregor in [11]discusses the benefits and effectiveness of the use of Big Data in neonatal intensive care units. He claims that it will lead to the earlier discovery of fatal medical conditions. Also, the capability to process multiple high-speed physiological data streams from numerous patients in numerous places and real-time could considerably improve both healthcare competence and patient outcomes.Patel et al. on the "Big Data for better health planning"[12]discuss the reasons for using data in healthcare and the results of various surveys to demonstrate the influence of Big Data in healthcare.Also,the study presents some case studies on Big Data analytics in healthcare industries. Different tools for handling Big Data problems are discussed.Ojha et al presented in [13]insight into how Big Data analytics tools like Hadoop can be used with healthcare data. The study discusses how meaningful information can be extracted from EHR (electronic health records). This work conducted experiments on central India's major government hospital. It used different data mining techniques such as clustering classification, and association. This implies a steady and efficient medical Big Data analysis technique is available.Koppad et al [14]introduced an application of Big Data analytics in the healthcare system. It predicts the diagnosis of COPD (chronic obstructive pulmonary disease). It used the decision tree technique, a data mining technique. The authors present an encouraging accuracy in diagnosing COPD patients and the efficacy of the proposed system through experimental results.

## 3. BDAArchitectural Framework

An architectural framework is an important requirement for any knowledge discovery process[6].Big Data architecture is designed to handle the ingestion, processing, and analysis of data that is too large or complex for traditional database systems. Below suggested architectural:

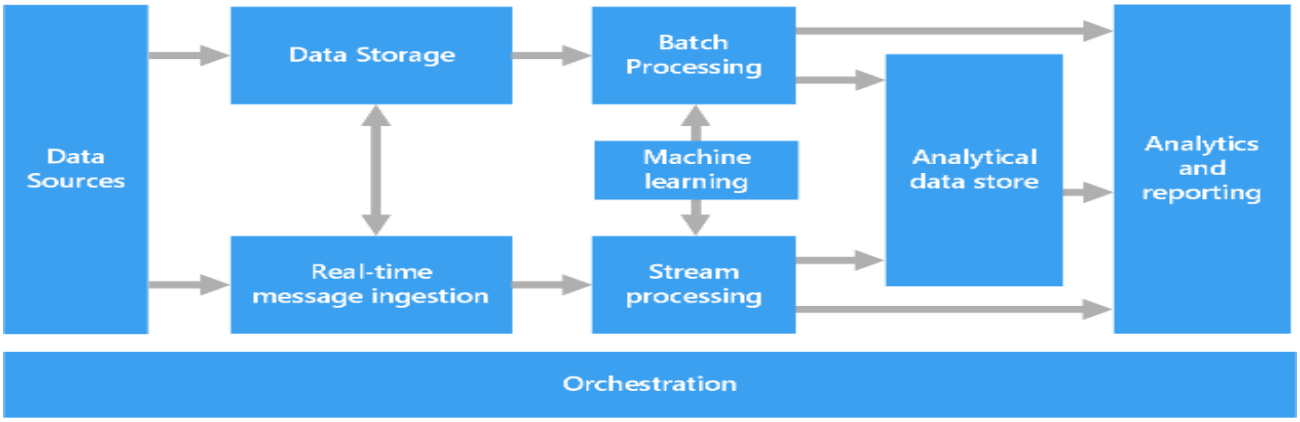


Figure 2: BDAarchitectural framework[15].

Most Big Data architectures include some or all of the components on Figure 3 are as follows:

**Data sources** as all Big Data solutions start with one or more data sources, as static files produced by applications, such as web server log files.

**Data storage** is wheredata for batch processing operations are typically stored.This is distributed file store that can hold high volumes of large files in various formats. This kind of store is often called a data lake.

**Batch processing** is used to running batch jobs to filter, aggregate, and prepare the data for analysis. Usually, it involves reading source files, processing them, and writing the output to new files. It is needed because the data sets are so large. Often, a Big Data solution must preprocess data files before using them.
**Real-time message ingestion** helps if the solution includes real-time sources. As a result,the architecture must include a way to capture and store real-time messages for stream processing. This might be a simple data store, where incoming messages are dropped into a folder for processing.
**Stream processing** comes to process real-time messages by filtering, aggregating and preparing the data for analysis. The processed stream data is then written to an output sink.
**Analytical data store** used to prepare data for analysis and then serve the processed data in a structured format that can be queried using analytical tools on next stage.
**Analysis and reporting** to provide insights into the data through analysis and reporting. It might also support modeling and visualization technologies. It is the core of BDA. The analysis stage can apply many analysis techniques and answers different questions.Analysis types discuss in the next section.
**Orchestration** is used to repeat data processing operations, encapsulated in workflows, that transform source data, move data between multiple sources and sinks, load the processed data into an analytical data store, or push the results straight to a report or dashboard. [6][15].

**4.BDA Analytics Types in Healthcare**

The diverse origins and forms of Big Data are challenging the healthcare informatics community, which led to using a variety of analytical techniques.Different data analytics in healthcare can be used, as in Figure 4.

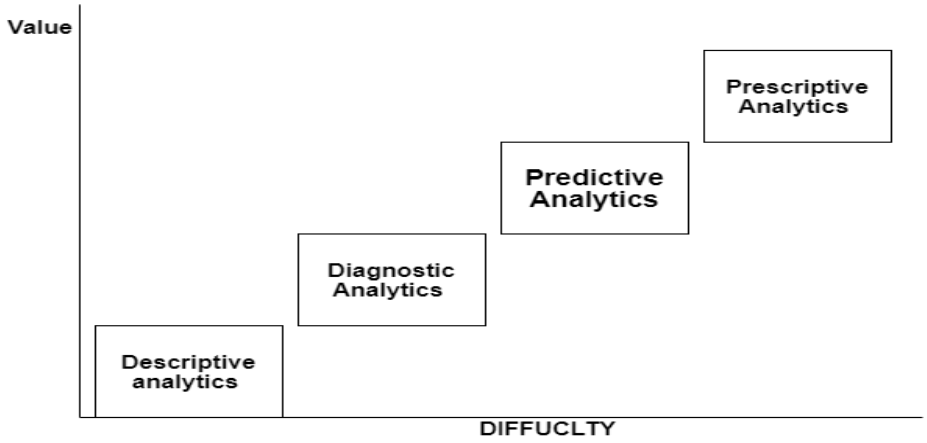


*Figure 4.BDA analyticsTypes*

**Descriptive analytics in healthcare**:It is the ability to quantify events and report on them in a human-readable way. This can help with population health management tasks such as identifying how many patients are living with diabetes, benchmark outcomes against government expectations, or identify areas for improvement on clinical quality measures or other aspects of care[16].It uses statistical measures within the historical data. A simple statistical measure can be used like counts, percentages, averages, and standard deviation. Descriptive analytics informs that this is happening and provides real-time data with all the corresponding statistics as the date of occurrence, volume, and patient detail.In general, this descriptive-analytical is used at the beginning to get a better historical understanding[17] [18] [4].
**Diagnostic Analytics in healthcare:**It is a form of advanced analytics that examines data or content to answer why something happened. It needs extensive exploration and directed analysis of the existing data. It is characterized by techniques such as drill-down, data discovery, data mining correlation, and visualization techniques[17].

Diagnostic analytics would explore the data and make correlations. For instance, it may help you determine that all of the patients' symptoms—high fever, dry cough, and fatigue—point to the same infectious agent,You now have an explanation for the sudden spike in volume at the emergency room ER[18][4].

**Predictive Analytics in Healthcare:**Predictive analysis has been recognized as one of the major business intelligence. It is a type of technology that combines machine learning and business intelligence with historical as well as real-time data to make projections about future events. This data can help doctors to make important patient care decisions[8].

The success of predictive analytics and healthcare lies in identifying the most promising use cases, capturing quality data, and applying the best model to uncover meaningful insights.

This can improve various areas of healthcare like helpon Triage which estimate the risk of complication when patient first present on the hospital [19].

**Prescriptive Analytics in healthcare:**Prescriptive analytics moves beyond the ability just to predict an upcoming event and provides the capability to do something about it. Prescriptive analytics is an advanced analytics concept. It is based on Optimization that helps achieve the best outcomes and stochastic optimization. Ithelps understand how to achieve the best outcome and identify data uncertainties to make better decisions.It actively suggests how organizations can take action to avoid or mitigate a negative circumstance[4] Prescriptive analytics is the most advanced healthcare analytic because it allows us to make specific recommendations on patient care delivery interventions.The future of prescriptive analytics is nearly unlimited in its scope and depth as developers dream up the technologies of the future [18]. In summary, healthcare analytics is a tool that can be used to answer any questions about patients in a data-driven manner. It is used in healthcare to understand performance, improve quality, advance research, and risk management.

Hopefully, with the development in the technology aspect, healthcare will do more advanced analytics [18].

**ToolsforBig Data Analytics in Healthcare:**

Many different platforms and tools are used in Big Data analytics in healthcare data. Some ofthe frequently used tools are listed here.

**The Hadoop Distributed File System (HDFS):**This is the prime data storage system used by Hadoop applications. HDFS is a distributed file system that handles large data sets running on commodity hardware. It is used to scale a single Apache Hadoop cluster to hundreds (and even thousands) of nodes. The main drawback is complexity in the usage. Also, it is not suited for small data. HDFS is used for handling input data from different resources[6][20][21].

**Apache Flume:**Apache Flume is a highly reliable and distributed system for collecting, aggregating and moving massive quantities of log data. Then, moving large volumes of data from independent machines to HDFS. Often data transport involves several flume agents that may traverse a series of machines and locations. Flume is often used for log files, data generated by social media, and email messages. It used fordata storage[1]

**MapReduce:** Apache Hadoop is often associated with MapReduce computing. The MapReduce computation model is a very powerful tool used in many health applications. Its underlying concept is very simple. In MapReduce, there are two stages: a mapping stage and a reducing stage.The mapping stage performs filtering and sorting, while, reducing stage performs a summary operation. MapReduceis usedfor analyzing patients' data[1][21].

**Apache Pig:**Apache Pig is one of the available open-source platforms being used for analyzing large data sets. Pig is an alternative to the MapReduce programming tool. First developed by the Yahoo web service provider as a research project, Pig allows users to develop their user-defined functions and supports many traditional data operations such as join, sort, and filter. It isused for analyzing data[21][1][22].A large number of different analytical software tools are available in the market which can be used by the application of Big Data analysts to improve healthcare. Some applications will be listed in the next section.

**5. Application of Big Data Analytics in Healthcare**

**Preventing Epidemics** Big Data was used in preventing the spread of the Ebola virus in Africa by spotting population movements and providing insights into the best areas for setting up treatment centers and bring in movement restrictions.

**Detecting Side Effects of Drugs**Patient data and their medical conditions over some time can lead to insights on whether there are noticeable side effects to a drug.

**Identifying Disease Risks from the DNA**Family history on diseases and medical conditions can be profiled and DNA analysis can provide insights on medical conditions a patient is likely to encounter.

**Avoid Preventable Deaths**Constant monitoring of medical conditions can provide alerts to take preventive action. Medical data can be compared with existing data sets to come to insights on whether there is a risk[23][24].
**Medical Imaging**Medical images are rich information resources. It includes a wide spectrum of different image acquisition methodologies for different purposes. They can vary from two, three, to four-dimensional images. Big Data analysis helps on analyze and supports image decision.Higher resolution and dimensions result in the volume of data growing exponentially which requires advanced analytical methods and high-performance computing.

Region of Interest (ROI) analyst can be used in image analyst.ROI analysis involves extracting the signal from specified regions by selecting clusters of pixels or voxels in the image.it can be used for analysis within one subject or across multiple ones. In ROI each image is divided into two parts, foreground; the areas that carry diagnostically important information, and background; the rest of the image. Then, feature extraction and selection are applied to be used on the classification. This will help the physician decide on the image.

In spite of the big role of BDA in developing thehealthcare field, it faces some challenges which are listed in section 6.

**6.Challenges Faced Big DataAnalytics in Healthcare**

There are major challenges that are faced byBig Data analytics in healthcare.

One of them is that the healthcare data are not in a standardized format, so it needs to be standardized before use.The other major challenge faced by BDA is the real-time processing issue. Real-time Big Data analytics is an important requirement in the healthcare industry. To address this issue, the delay between data acquisition and data processing should be dealtwith quickly[6].Moreover,Big Data analytics should be maintained efficiently,so that data inconsistency should not occur at any instant or any cost asotherwise, this will lead to a serious problem of affecting the whole healthcare industry concerned.

Due to the constant evolution of technology, there exist populations of individuals lacking specific skills. Health care workers must be also kept up to date with the use of constantly changing technology, techniques, and a constantly moving standard of care.[7]

As in all analytic fields, security is an important issue. Healthcare organizations must frequently remind their staff members of the critical nature of data security protocols and consistently review those who has access to high-value data assets to prevent malicious parties from causing damage[25].

**7.The Future ofBig Data Analytics in Healthcare**

New technologies are continuing to emerge that push the boundaries of how healthcare analytics can be used. From artificial intelligence to machine learning to natural language processing, the true future of healthcare is in wielding these technologies for greater impact.

Quality of care will also be improved by reducing the waste of information, which will reduce inefficiencies. This will also assist in analyzing real-time resource utilization productivity.

**8.Conclusion**

Big Data and the use of advanced analytics have the potential to advance how providers leverage technology to make informed clinical decisions and luxury healthcare. This paper review many resources to clearly defined big data analytic, a good architectural frameworkused for BDA system, and how it has been used in the healthcare industry.This paper provided an in-depth description of healthcare system, which plays a significant role in healthcare informatics.The most famous tools for BDA are briefly illustrated. It demonstrated the exponential growth in the amount of the used on Big Data in health. The fields where it can be used are constantly expanding and it will help to conduct more accurate analyses in the coming years.

The future will held more powerful BDA, more tools will be available. The challenges of BDA may be solved. The use of IOTNet will increase the volume of data and the need for BDA.

**References**

1. Sunil Kumar and Maninder Singh, "Big Data Analytics for Healthcare Industry:Impact, Applications, and Tools," BIG DATA MINING AND ANALYTICS, vol. 2, p. p p 4 8– 5 7, 2018.
2. D. T. J. Seymour, "Electronic Health Records (EHR)," American Journal of Health Sciences (AJHS) , 2014.
3. By Mattew J. Mazzei and David Noble, "Big Data and Strategy: Theoretical Foundations and New Opportunities," *IntechOpen*, Jan,2019.
4. J. Bresnick, "Healthcare Big Data Analytics: From Description to Prescription," Health it analytics, 2015.

[Online]. Available: https://healthitanalytics.com/news/healthcare-big-data-analytics-from-description-to-prescription.

5. Panagiota Galetsi & Korina Katsaliaki, "A review of the literature on big data analytics in healthcare," *Journal of the Operational Research Society,* 2019.
6. D. J. Bora, BIG DATA ANALYTICS IN HEALTHCARE: A CRITICAL ANALYSIS, India: School of Computing Sciences, Kaziranga University, Jorhat.
7. Clemens Scott Kruse1, MBA, MHA, MSIT, PhD Author Orcid Image ; Rishi Goswamy1, MHA Author Orcid Image ; Yesha Jayendrakumar Raval1, MHA Author Orcid Image ; Sarah Marawi1, MHA Author Orcid Image, "Challenges and Opportunities of Big Data in Health Care: A Systematic Review," *JMIR MEDICAL INFORMATICS,* vol. vol 4, 2016 .
8. p. K. PRZYSTALSKI, "Predictive Analytics in Healthcare: Use Cases, Benefits & Challenges," *ARTIFICIAL INTELLIGENCE TECHNOLOGY,* April 14, 2020.
9. G. Press, "A Very Short History Of Big Data," *Forbes,* 2013.
10. F. X. Diebold, "On the Origin(s) and Development of the Term 'Big Data'," *SSRN Electronic Journal,* 2012.
11. C.McGregor, "Big dat in neonatal intensive care," *computer(46),* 2013.
12. Jigna Ashish Patel; Priyanka Sharma, "Big data for better health planning," *International Conference on Advances in Engineering and Technology Research (ICAETR),* 2014.
13. K. M. Mimoh Ojha, "Proposed application of big data analytics in healthcare at Maharaja Yeshwantrao Hospital," *IEEE,* 2016.
14. Shaila H Koppad,Anupamma Kumar, "Application of big data analytics in healthcare system to predict COPD," *IEEE,* 2016.
15. "Big data architectures," Microsoft, 02 12 2018. [Online]. Available: https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/. [Accessed 2 2021].
16. B. BRINKMANN, "Comparing Descriptive, Predictive, Prescriptive, and Diagnostic Analytics," *logi.*
17. lMohamedKhalif,aIbrahimZabani, "Utilizing health analytics in improving the performance of healthcare services: A case study on a tertiary care hospital," *Journal of Infection and Public Health,* vol. vol 9, no. Issue 6, 2016.
18. P. Len Usvyat, "4 Types of Healthcare Analytics to Use in Your Practice," *FMCNA ,* 2019.
19. k. Morblitzer, "The use of predictive analytics in healthcare," *UNC,* 2017.
20. "HDFS," IBM, [Online]. Available: https://www.ibm.com/analytics/hadoop/hdfs. [Accessed 2021].
21. arah Shafqat· Saira Kishwer ·Raihan Ur Rasool· Junaid Qadir,Tehmina Amjad· Hafiz Farooq Ahmad, "Big data analytics enhanced healthcare systems:a review," *springer Science,* 2018.
22. "Welcome to Apache Pig!," Apache, [Online]. Available: https://pig.apache.org/. [Accessed 2021].
23. S. R. R. Sudipta Chandra, "Big Data Security in Healthcare," *IEEE 7th International Advance Computing Conference,* 2017.
24. Muhammad Qasim Shabbir* and Syed Babar Waheed Gardezi, "Application of big data analytics and organizational performance: the mediating role of knowledge management practices," *Journal of Big Data,* (2020.
25. J. Bresnick, "Top 10 Challenges of Big Data Analytics in Healthcare," Health IT Analytics, June 12, 2017.