IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

Exploring the Path to Big Data Analytics Success in Health Care

Preshita Mahendra Panshikar[1] and Prof. Divakar Jha[2]
Student, Department of Computer Application[1]
Mentor, Department of Computer Application[2]
Late Bhausaheb Hiray S S Trust's Hiray Institute of Computer Application, Mumbai, India

Abstract: *Like Oxygen, the world is surrounded by data today. The quantity of data that we harvest and eat up is thriving aggressively in the digitized world. Increasing use of new innovations and social media generate vast amount of data that can earn splendid information if properly analyzed. This large dataset generally known as big data, do not fit in traditional databases because of its rich size. Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Organizations need to manage and analyze big data for better decision making and outcomes. So, big data analytics is receiving a great deal of attention today. In healthcare, big data analytics has the possibility of advanced patient care and clinical decision support. In this paper, we review the background and the various methods of big data analytics in healthcare. This paper also elaborates various platforms and algorithms for big data analytics and discussion on its advantages and challenges. This survey winds up with a discussion of challenges and future directions.*

Keywords: Big Data, Cloud Computing, Hadoop, Big Data Mining, Predictive Analytics.

I. INTRODUCTION

The new advances in Information Technology (IT) guide to smooth creation of data. Healthcare sector also has produced huge amount of data by maintaining records and patient care. Contrary of storing data in printed form, the fashion is digitizing those limitless data. Those digitized data can be used to improve the healthcare delivery quality at the same time reducing the costs and hold the promise of supporting a wide range of medical and healthcare functions.

Also it can provide advanced personalized care, improves patient outcomes and avoids unnecessary costs. By description, big data in healthcare refers to electronic health datasets so large and complex that they are difficult to manage with traditional software, hardware, data management tools and methods

Healthcare big data includes the clinical data, doctor's written notes and prescriptions, medical images such as CT and MRI scans outcomes, laboratory records, drugstore documents, insurance files and other administrative data, electronic patient records (EPR) data; social media posts such as tweets, updates on web pages and numerous amount of medical journals. So, huge amount of healthcare data are available for big data scientists. By understanding stencils and trends within the data, big data analytics seems to improve care, save lives and reduce costs. Therefore, big data analytics applications in healthcare take advantage of extracting insights from data for better decisions making purpose. Analytics of big data is the process of inspecting enormous amount of data, from different data sources and in various formats, to deliver insights that can enable decision making in real time. Various analytical concepts such as data mining and artificial intelligence can be applied to analyze the data. Big data analytical approaches can be employed to recognize anomalies which can be found as a result of integrating vast amounts of data from different data sets.

In the rest of this paper, firstly we introduce the common background, definitions and properties of big data. Then various big data platforms and algorithms are discussed. Eventually, the challenges, future directions and conclusions are presented.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

II. APPLICATIONS OF BIG DATA IN HEALTH CARE

2.1 Drug Discovery and Development

Introducing new drugs to the market is a daunting task that is often expensive requiring upfront investments running into millions. Several different techniques such as machine learning and visual analytics tools help streamline simulations across data sets.

2.2 Enhancing Patient Engagement

Many consumers use smart devices that record each step and vitals such as heart rates, as well as sleeping quality permanently. This vital information when combined with other trackable data can help identify potential health risks. Patients can directly monitor their health and get incentives from health insurance that can lead them to lead a healthy lifestyle.

2.3 Electronic Health Records (EHRs)

Every patient has a personal digital record including details such as demographics, medical history, allergies, and laboratory test results. Records are shared through secure information systems and are available for providers from both the public and private sectors. Every record consists of a file that allows doctors to implement changes over time with no paperwork and no danger of data replication.

EHRs can also generate warning signals and reminders for a patient when it is time to get a new lab test or check routine prescriptions to find if a patient has been following doctors' orders.

2.4 Anticipate and Treat Illnesses

Access and analyzing voluminous data sets can improve our ability to anticipate and treat illnesses. This data can help track individuals who are at considerable risk for critical health problems. The ability to effectively use big data to identify waste in the healthcare system can significantly reduce the cost associated with healthcare across the board.

2.5 Fitness Tracking Devices

Keeping patients healthy and avoiding illness and disease comes across as the top priority list. Consumer health and fitness products like the Fitbit activity tracker and the Apple Watch keep tabs on the physical activity levels of individuals and can report on specific health-related patterns. The data collected is delivered to cloud servers, providing information to physicians who use this vital information for overall health and wellness programs.

Apple's HealthKit, CareKit, and ResearchKit leverage the technology embedded in Apple's mobile devices to help patients manage their conditions and enable researchers to collect data from hundreds of millions of users worldwide.

2.6 Eliminating Prescription Errors

Adverse drug events (ADEs) account for more than 3.5 million physician office visits and 1 million emergency department visits each year. It is believed that preventable medication errors impact more than 7 million patients and cost almost $21 billion annually across all care settings.

Healthcare organizations are using decision support tools empowered by big data analytics that help identify fatal prescription errors even before they occur. Big data analytics solutions allow, preventing a wide range of prescription errors relating to wrong drug, wrong patient, drug interactions, dosages, and allergies with higher precision than the existing systems.

2.7 Innovation in Healthcare Industry

A. Wearables and IoT

Wearables are now a rage everywhere, allowing the collection and measurement of vital information from sensors placed on human bodies. This information is relevant to maintaining the health of a person. A wearable device or sensor will provide a real-time feed of health records, which allows medical staff to monitor and later consult with the patient, either face-to-face or remotely.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

B. Precision medicine

Precision medicine aims to understand a person's genetics and lifestyle choices/habits to determine the best approach to either avoid or treat a disease. The long-term goals of the Precision Medicine Initiative focus on revamping healthcare on a large canvas.

C. Machine Learning

Component of artificial intelligence and one that depends on big data is already helping physicians improve patient care. Machine learning together with healthcare big data analytics, multiply caregivers' ability to enhance patient care.

III. RELATED TECHNOLOGIES

3.1 Big Data Platforms

As in big data uses distributed storage technology based on cloud computing rather than local storage. Some big data cloud platforms are Google cloud services, Amazon S3 and Microsoft Azure. Google's distributed file system GFS (Google File System) and its programming model Mapreduce are the lead in the field. The performance of mapreduce has received a valid amount of attention in large scale data processing. So many organizations use big data processing framework with map reduce. Hadoop, an influential aspect in big data was developed by Yahoo and Hadoop enables storing and processing big data in distributed environment on large clusters of hardware. Enormous data storage and faster processing are supported by hadoop. Hadoop Distributed File System (HDFS) provides reliable and scalable data storage. HDFS makes multiple copies of each data block and distributes them on systems on a cluster to enable reliable access. HDFS supports cloud computing through the use hadoop, a distributed data processing platform. Another one, 'Big Table' was developed by Google in 2006 that is used to process huge amount of structured data. It also supports map reduce.

Amazon developed Dynamo, a key-value pair storage system. It is a scalable distributed data store built for Amazon's platform. It gives high reliability, cost effectiveness, availability and performance. Tom white elaborates various tools for big data analytics. Hive, a framework for data warehousing on top of hadoop. It was built at Facebook. Hive with hadoop for storage and processing meets the scalability needs and is costeffective. Hive uses a query language called HiveQL which is alike on SQL.

A scripting language for exploring large datasets is called 'Pig'. An opinion of map reduce is that writing of mappers and reducers, compiling the package and code are tough and so the development cycle is long. Hence working with mapreduce needs experience. Pig overcomes this criticism by its simplicity. It allows the developers to write simple Pig Latin queries to process the big data and thereby save the time. A distributed column oriented database Hbase built on top of Hadoop Distributed File System. It can be used when we need random access of very large datasets. It speeds up the performance of operations. Hbase can be accessed through application programming interfaces (APIs) like REST (Representational State Transfer) and java. Hbase does not have its own queries, so it depends on Zookeeper. Zookeeper manages huge amount of data. This allows distributed process to manage through a namespace of data registers. It is a data mining and machine learning library.can be categorized as collective filtering, categorization, clustering and mining. It can be executed by Mapreduce in a distributed mode. Big data analytics is not only based on platforms but also analytics algorithms plays a significant role.

3.2 Algorithmic Techniques

Big data mining is the method of winnowing hidden, unknown but useful information from massive amount of data. This information can be used to predict future situations as a help to decision making process. Helpful knowledge can be found by the usage of data mining techniques in healthcare applications like decision support system. The big data produced by healthcare organizations are very complicated and vast to be handled and analyzed by usual methods. Data mining grants the procedure to transform those bundles of data into useful information for decision support. Big data mining in healthcare is about learning models to predict patients' disease. For example, data mining can help healthcare insurance organizations to detect hypocrites and misuse, healthcare institutions make decisions of customer relationship management, doctors identify effective treatments and best practices, and patients get improved and more economical healthcare services. This predictive analysis is widely used in healthcare institutions make decisions of customer

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

relationship management, doctors identify effective treatments and best practices, and patients get improved and more economical healthcare services. This predictive analysis is widely used in healthcare.

There are various data mining algorithms discussed in 'Top 10 algorithms in data mining'. It discussed variety of algorithms along with their limitations. Those algorithms encompass clustering, classification, regression, statistical learning which are the issues in data mining investigation. The ten algorithms discussed include C4.5, k-means, Apriori, Support Vector Machines, Naïve Bayes, EM, CART, etc.

Big data analytics includes various methods such as text analytics, multimedia analytics and so on. But as given above, one of the crucial categories is predictive analytics which includes various statistical methods from modeling, data mining and machine learning that analyze current and historical facts to make prediction about future. In hospital context, there are predictive methods used to identify if someone may be at risk for readmission or is on a serious recession. This data helps therapists to make important care decisions. Here it is necessary to know about machine learning since it is widely employed in predictive analysis.

The process of machine learning is very much alike of data mining. Both of them hunt through data to look for patterns. But, rather than extracting data for human understanding as in data mining applications, machine learning model uses that data to improve the program's own understanding. Machine learning programs finds patterns in data and alters program functions respectively. Machine learning The process of machine learning is very much alike of data mining. Both of them hunt through data to look for patterns. But, rather than extracting data for human understanding as in data mining applications, machine learning model uses that data to improve the program's own understanding. Machine learning programs finds patterns in data and alters program functions respectively. With the increasing knowledge in the area of big data, the variety of techniques for analyzing data is represented in 'Data Reduction Techniques for Large Qualitative Data Sets'. It describes that the selection for the particular technique is based on the type of dataset and the way the pattern are to be analyzed. They aimed at efficiently analyzing large dataset in a minimal amount of time.

C5.0 is the classification algorithm which is applicable for big data sets. It overcomes C4.5 on the speed, memory and the performance. C5.0 method works by splitting the sample based on the field that gives the maximum information gain. The C5.0 system can split samples regarding of the biggest information gain field.

IV. IMPORTANCE OF BIG DATA IN HEALTHCARE

This study identified that aptness of BDA to add significant value to healthcare can be classified into six themes, namely, conceptual evolution, data governance, decision support, disease prediction, strategy formulation, and technology development. It is expected that the findings of this study will be useful to healthcare practitioners, policymakers, and service developers, as subsequently discussed. First, healthcare practitioners, particularly hospital administrators, should take note of the innovative ways presented herein to improve efficiency in healthcare service delivery using BDA. These innovative approaches include, for example, the supervision of patients with specific medical conditions, medication assignment, and pre-admission testing**.** Second, policymakers will find inputs from the current study's findings for formulating healthcare policies, optimising public funds usage, and developing legal frameworks. Suitable public policies may deliver efficient decision-support systems, infrastructure development, and technological advancement in healthcare. Third, service developers may do well to follow our study findings when exploring opportunities to develop new services for the healthcare sector using state-of-the-art technologies. For instance, the application of augmented reality, quantum computing, and digital twins have the potential to maximise the value added by BDA to healthcare in the future. Fourth, at present, we are facing a tough challenge from the Covid-19 outbreak. BDA can help medical professionals, scientists, epidemiologists, public health officials, and policymakers fight this pandemic. For example, scientists and policymakers can use BDA to comprehend and trace the impact of the Covid-19 pandemic. BDA not only helps in locating the fast spread of the Covid-19, but it can also aid various efforts undertaken to control and prevent its spread.

Big Data reportedly helps hospital management improve their efficiency in delivering healthcare services and in providing customised care to patients. Future research is invited to empirically study the role of Big Data Analysis is improving service quality in hospitals. It is hoped that scholars will also explore further means of providing more sophisticated personal assistance to individuals, especially senior citizens and patients suffering from chronic diseases. Overall, this set of future research agendas reiterates the future research scope of extending the present study

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

V. CHALLENGES ASSOCIATED WITH BIG DATA IN HEALTHCARE

Big data analytics not only provides charming opportunities but also faces lot of challenges. The challenge starts from choosing the big data analytics platform. While choosing the platform, some criteria like availability, ease of use, scalability, level of security and continuity should be considered. The other challenges of big data analytics are data incompleteness, scalability and security. Since cloud computing plays a major role in big data analytics, cloud security should be considered. Studies show that 90% of big data are unstructured data. But the representation, analytics and access of numerous unstructured data are still a challenge. Data timeliness is also critical in various healthcare areas like clinical decision support for making decisions or providing information that guides to take decisions. Big data can make decision support simpler, faster and more accurate because decisions are based on higher volumes of data that are more current and relevant. This needs scalable analytics algorithms to produce timely results. However, most of the current algorithms are inefficient in terms of big data analytics. So the availability of effective analytics algorithms is also necessary. Concerns about privacy and security are superior, although these are increasingly being attempted by new authentication approaches and policies that better protect patient identifiable data. One of the most complicated challenges is the aggregation of data. The information of patients is usually spread across various administrators, hospitals, servers and file cabinets. Proper planning is needed to amalgamate and arrange all these together for future collaboration. Moreover, all the collaborated organizations should have a clear idea and agree to the formats and types of big data they want to analyze. Additionally, the cleansing and governance of data are also required to maintain the accuracy and quality of data.

The processes and policies that protect health data are often involved with various issues, after aggregating and validating the data, these issues should be addressed.

If there is no proper track of where the data is being saved, the IT department may face money, security, and performance-related problems. The amount of healthcare data increases over time, therefore the management team sometimes faces issues in handling the expense and effects of premise data storage. Maintaining an on-site server for storing data can be a difficult and costly affair. Also, this way the process of spreading the data across different departments can be messy.

Data security is very important for health organizations. If the health organizations start implementing big data, the healthcare givers should have access to that data. And when access is granted by the organization to some specialists there is no problem. But if there is a big team gets involved with patients' personal information, the problem related to privacy breach may arise. To avoid such problems healthcare organizations should bank on trusted data vendors with a safe and well-structured distribution to receive secure big data solutions.

And lastly realizing the future of big data in healthcare needs organizations to modify how they do business. Data scientists are required together with the proficient IT staff to perform the analytics. Without proper IT infrastructure organizations may find struggles in making the most out of big data.

Despite all the hurdles, the application of big data in the healthcare industry is growing as it helps the medical practitioners provide a proactive approach to their patients in a fast and affordable way.

VI. CONCLUSION & FUTURE SCOPE

Large amounts of heterogeneous medical data have become available in various healthcare organizations. The rate of electronic health record (EHR) adoption continues to climb in both inpatient and outpatient aspects. Those data could be an enabling resource for deriving insights for improving patient care and reducing waste. Analyzing the massive amount of healthcare information that is newly available in digital format should enable advanced detection of powerful treatment, better clinical decision support and accurate predictions of who is likely to get sick. This requires high performance computing platforms and algorithms. This paper reviews the various big data analytics platforms and algorithms and challenges are discussed. The current study intended to address four research questions related to the application of BDA in healthcare. These questions have been answered following a standard protocol for reviewing resources from key databases. The prior literature on the application of BDA in healthcare has focused on five main themes, namely health awareness, stakeholders of the healthcare ecosystem, hospital management practices, specific medical conditions, and healthcare service delivery through technology use. The study has identified the gapsin the existing literature and provided an actionable research agenda for future research on the utilisation of big data in the healthcare sector. However, despite the significant contributions of this current study, it suffers from three main

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

limitations: first, book chapters, magazine articles, and thesis studies have been excluded from the scope of this study; second, journal articles and conference studies not available in English were not considered; third, studies not available in the four databases were not reviewed unless they appeared in the forward and backward searches. Future research is invited to overcome these limitations. Also, we recommend that scholars study the application of BDA in services provided by, for example, banking and financial institutions, media and broadcast channels, and the travel and hospitality industry by adopting the protocol followed in the current study. Similarly, the application of new technologies, such as blockchain, cloud computing, and machine learning, in healthcare provides promising avenues of exploration.

Healthcare and big data are intensively involved with each other. The implementation of big data has become necessary to boost success for the healthcare industry. It doesn't only help the healthcare givers provide better care to patients but also make the whole industry's marketing touch points more defined and integrated.

VII. LITERATURE REVIEW

We conclude this SLR (System Literature Review) with a call for theory development regarding the specific applications of BDA and the general integration of technology in the healthcare sector. From drug testing and patient care to genetic testing, there are several opportunities to leverage big data analytics solutions in improving outcomes. It is now possible for healthcare and life science organizations to apply fast and actionable analytics.

Research on the application of BDA in healthcare is gaining popularity, particularly within the domains of information systems and medical studies. As one of the earliest comprehensive reviews on the topic, the present study offers three major implications to theory, as subsequently discussed. First, this study presents a current research profile on the applications of BDA in healthcare. This research profile includes information about the key contributors, prominent publication outlets, and common methodologies prevalent in the reviewed studies. Second, the current study has identified the themes of healthcare contexts in which BDA is applied and where BDA can deliver value. Our review of prior literature indicated that the contexts can be synthesised into five broad themes addressing health awareness, healthcare ecosystems, hospital management, specific medical conditions, and technology aspects. A thematic identification organised prior literature and aimed to catalyse future research in various related domains of study. Third, the present study proposed a comprehensive framework that captures the interplay among the process of health data accumulation, derivation of the insights from the data, and application of these insights to healthcare. The comprehensive framework also offers future research agendas for advancing the application of BDA in healthcare.

.

VIII. ACKNOWLEDGEMENT

We would like to acknowledge the University of Mumbai, Mumbai, India to give us the opportunity to do the research work under the title "**Exploring the Path To Big Data Analytics Success in HealthCare**". Also, we would like to acknowledge the college L.B.H.S.S. T's ICA Bandra East, Mumbai, India to support us during the research process.

REFERENCES

[1]. Alexandros Labrinidis and H. V. Jagadish, "Challenges and opportunities with big data," Proc.pp. 2032-2033, August 2012.

[2]. Amir Gandomi, Murtaza Haider, "Beyond the hype: Big data concepts, methods, and analytics," International Journal of Information Management 35, pp. 137-144, 2015.

[3]. Chaitrali, S., D. Sulabha and S. Apte, "Improved study of heart disease prediction system using data mining classification techniques," 2012.

[4]. Wullianallur Ragupathi and Viju Raghupathi, "Big data analytics in healthcare promise and potential",2016

[5]. Xindong Wu, Xingquan Zhu, Gong-Qing Wu, Wei Ding, "Data mining with big data," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, 2014.

[6]. Wu, J., H. Li, S. Cheng, and Z. Lin. 2016. "The promising future of healthcare services: when big data analytics meets wearable technology." Information & Management 53 (8): 1020–1033. doi:10.1016/j.im.2016.07.003.

[7]. Yasin, S. A., and P. P. Rao. 2018. "A framework for decision making and quality improvement by data aggregation techniques on private hospitals data."

[8]. ARPN Journal of Engineering and Applied Sciences 13 (14): 4337–4345. Zaragoza, M. G., H. K. Kim, and Y.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, June 2022
Impact Factor: 6.252

Chung. 2017. "U-healthcare Big data analytics process control." International Journal of Control and Automation 10 doi:10.14257/ijca.2017.10.11.15.

[9]. Zhang, F., J. Cao, S. U. Khan, K. Li, and K. Hwang. 2015. "A task-level adaptive mapreduce framework for real-time streaming data in healthcare applications." Future Generation Computer Systems 43: 149–160.