

## Big Data Analytics for Healthcare Organization, BDA Process, Benefits and Challenges of BDA: A Review

Siva Sankara Reddy Donthi Reddy\*,<sup>1</sup>, Udaya Kumar Ramanadham<sup>2</sup>

<sup>1</sup>Department of Computer Science & Engineering, BIHER, Bharath University, Chennai, Tamilnadu, India

<sup>2</sup>Department of Information Technology, BIHER, Bharath University, Chennai, Tamilnadu, India

### ARTICLE INFO

Article history:

Received: 10 July, 2017

Accepted: 05 September, 2017

Online: 09 October, 2017

Keywords:

Big Data

Big Data Analytics

Healthcare

BDA process

BDA Advantages

### ABSTRACT

Day by day, data grows exponentially large using advanced technology and it requires effective analytical techniques to analyze the unknown and useful facts, patterns, associations and new trends which will provide new way for giving treatment to diseases and to provide good quality healthcare at low cost for everyone. This paper describes uncover valuable insights, various lifestyle choices, some social determinants, clinical and financial factors that it may effect the overall health of an individual. It also presents how to analyze the facts by using big data analytics to improve the healthcare in the world and also describes the various steps involved in Big Data Analytics process and discusses its advantages and challenges which show impact on healthcare organization.

## 1. Introduction

In the digital world, data are generated as large sets from various sources. The fast transition from conventional to digital technologies has contributed to the growth of big data. It provides evolutionary breakthroughs in many fields with collection of large datasets. Big Data is generated everyday by diverse segments of industries like business, finance, manufacturing, healthcare, education, research and development etc. In general, it refers to the collection of large and complex datasets which are difficult to store and process using traditional database management tools or data processing applications. So there is need of developing and using an effective, innovative tools and technologies offered by Big Data. Data can be of structured, unstructured and semi-structured type. Different variety of data include the text, audio, video, log files, sensor data etc. in petabytes and beyond. As the data is too big from various sources in different form, it is characterized as 5 V's. The 5 V's of Big Data are: Volume, Variety, Velocity, Veracity and Value [1]. Volume represent the size of the data - how large the data is. The size of the data can be represented in terabytes and petabytes. Variety represents the data which appears in different forms. Velocity represents the motion of the data and the analysis of streaming of the data. Veracity represents the availability and accountability of various sizes of data. Value represents the high quality of data. The Big Data helps more to healthcare in the world [9]. The healthcare organization has generated large amount of data till date, which is scaled in

petabytes or exabytes. According to [3], with such fast and rapid growth of data, U.S. healthcare alone will soon reach the zettabyte (1021 gigabytes) scale. The main goal of healthcare industry is to analyse this big volume of data for unknown and useful facts, patterns, associations and trends with the help of machine learning algorithms, which can give new innovative techniques for treatment of various diseases. The aim is to provide high quality healthcare at lower cost to all. This can be a beneficial one for the entire world. Big Data sources are showed in the following Figure 1.

## 2. Characteristics of Big Data:

The 5 V's of Big Data relevant to Healthcare are:

- i) **Volume:** As described earlier, healthcare industry produces the variety of data with more growth rate. According to EMC report and the research firm IDC, the healthcare data increases with 48 per cent annually. In 2013 year, the healthcare data was 153 Exabyte's and it may increase to 2,314 Exabyte's by 2020.[1-2]
- ii) **Variety:** In the past, the healthcare organization was generating clinical data of patients with similar symptoms, storing and analysing it to derive the most effective course of treatment for the admitted patient. Now the healthcare industry is focusing on complete healthcare, by providing an effective treatment through analysis of a patient's data from various other sources also. This refers to the variety. Generally, the varied health care data falls into one of the three categories - i.e. structured, semi structured and unstructured. Generally the following data is collected:

\*Corresponding Author: Udaya Kumar Ramanadham, Professor, Department of Information Technology, BIHER, Bharath University, Chennai, Tamilnadu, India, Contact No: (+91) 9789994242, E-Mail: rsukumar2007@gmail.com

clinical data from Clinical Decision Support systems (CDSS) (physician’s notes, genomic data, behavioural data, data in Electronic Health Records (EHR), Electronic Medical Records (EMR)), machine generated sensor data, data from wearable devices, Medical Image data (from CT scan, MRI, X Ray’s etc.), medical claim related data, hospital’s administrative data, national health register data, medicine and surgical instruments expiry date identification based on RFID data[3-6], social media data like Twitter data, Facebook data, web pages, blogs and various articles.[7]

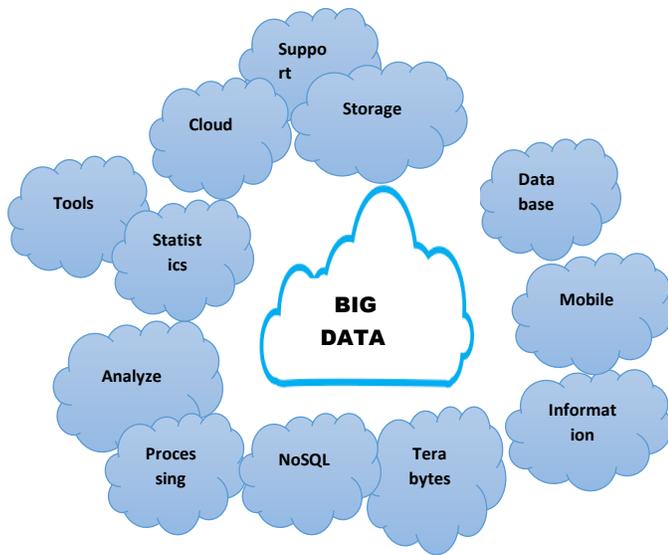


Figure 1: Big Data Sources

**iii) Velocity:** It refers to the frequency and speed at which data is generated, captured and shared. More data is generated by consumers as well as businesses with in shorter cycles, from hours, minutes, and seconds down to milliseconds. The wearable devices and sensor devices collect real time physiological data of patients rapidly. This new data which is being generated every second is posed a complex and critical challenge for data analysts. Social media data is also added to velocity as the users views, posting data, feeding data scale up in seconds to enormous amount in case of epidemics/national disasters.

**iv) Veracity:** It refers to trustworthiness of data. Data is accumulated in real-time and at a rapid pace, or velocity. The continuous flow of new data presents new challenges. Just as the volume and variety of data that is collected and stored has changed, so is the velocity at which speed it is generated and that is necessary for accessing, analyzing, and comparing as well as taking decisions based on the output. Most healthcare data has been traditionally static— paper files, x-ray films, and scripts. Velocity represents regular monitoring, such as more daily diabetic glucose measurements (or more continuous control by insulin pumps), blood pressure readings, and EKGs. Meanwhile, in many medical situations, constant real-time data (trauma monitoring for blood pressure, operating room monitors

for anesthesia, bedside heart monitors, etc.) can mean the difference between life and death.[8-9]  
**v) Value:** It refers to the quality of data. The data of EMR’s and EHR’s are recognized as high value data normally. But it is too difficult to certify the value of data from social media. So, the effective analytical methods are needed for the high value data to lead for better quality, effective healthcare solutions and innovations.

The following Figure 2 depicts the 5 V’s of Big Data in Healthcare.

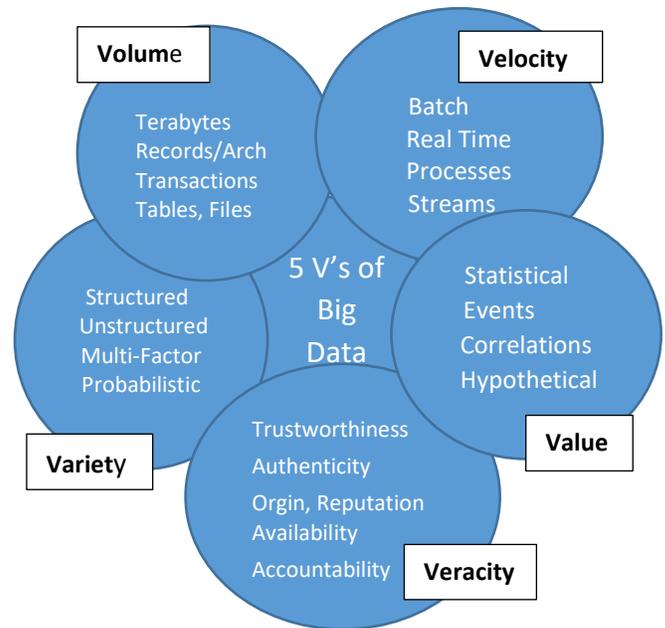


Figure 2: Five V’s of Big Data in Healthcare

### 3. Literature Survey:

By increasing digitization of healthcare information, it is needed to improve the quality of healthcare, results, and reducing the costs. The advanced tools and technologies are used in health care organizations to generate valuable insights of digital healthcare information. The organizations must analyse patient information to more accurately measure the risk and for better outcomes. At the same time, many organizations are working to increase data transparency for producing latest insightful knowledge.

To exchange health information between various providers and payers, some integrated delivery networks can be formed. The pharmaceutical companies are tied up to protect patients’ privacy while making data available to qualified researchers outside the organization.

Kiyana Zolfaghar et al. [17] has presented prediction model to give possible solutions for congestive heart failure incidents using Mahout Framework. The raw data is pre-processed and converted to encoded format which will be given as input to the Mahout framework, using random forest algorithm.

Joseph M. Woodside [18] has presented, inefficient vendors can be identified, and who is poor in the member’s lifestyle

decisions and compliance with preventative care programs. For individuals, intensives can be given, such as cash, gift cards which are considered as one of the recommended changes in the health care system.

Existing analytical techniques can be applied to the vast amount of existing patient related health and medical data to reach a deeper understanding of results, which can be applied at the point of care. Ideally, these data would inform each physician and their patients during the decision-making process and used to identify the appropriate treatment option for that particular patient.

#### **4. BDA Initiatives for Healthcare Industry in the World**

Most of the countries have initiated the number of big data initiatives around the world. Some of the initiatives are described as follows:

New Zealand's Ministry of Health has collaborated with New Zealand Society to Study about Diabetes. These have used SAS. [10]

Data analysis for providing a Virtual Diabetes Register which combines and filters the health information to determine accurately that how many people are diagnosed with the diabetic condition and predicting who can develop diabetes in the future.[12]

McKinley Children's Centre of California provides child welfare services in Los Angeles country. The organization serves for more than 700 children annually, which provides the services such as residential care, foster care and adoptions, special education, and mental health services. The center has launched an innovative big data analytics for initiating the staff identifying the variables that impact each child's success and identifying the right combination of programs to improve outcomes. [13]

The Data Science Institute of Columbia University, New York has collaborated with the New York City Department of Health and Mental Hygiene (NYC DOHMH) for working a project that focuses on the detection of disease outbreaks in New York City restaurants. The main goal of this project is to identify and analyze the unprecedented volumes of user-contributed opinions and comments on social media sites such as Twitter, Face book and Yelp, which host massive amounts of content by users about their real-life experiences and opinions about restaurants. It will help to extract reliable indicators of otherwise-unreported disease outbreaks associated with the restaurants. The NYC DOHMH analyses these indicators, as they are produced, to decide when additional action is required to be taken. This project is developing non-traditional information extraction technology over redundant, noisy, and often ungrammatical text-- for a public health task of high importance to society at large. [14]

WellPoint, Inc. is an Indiana polis-based health Benefit Company wanted to reduce the waste of resources (money) by improving the utilization management (UM) process, which governs the pre-approval of healthcare insurance coverage for many medical procedures. Its goals were to accelerate processing of physicians' treatment requests, save members' time and improve efficiencies in the approval process, while continuing to base UM decisions on medical evidence and clinical practice guidelines(for ensuring consistency in process).It was a very big

challenge considering the volume of data that is analyzed in making UM decisions. WellPoint was teamed the IBM on a new method to UM: using the cognitive system IBM Watson to provide approval guidelines for nursing staff, based on clinical and patient data. WellPoint trained Watson with 25,000 historical cases. The system uses hypothesis generation and evidence-based learning to generate confidence-scored recommendations that help nurses make decisions about UM. The new system provides responses to all requests in seconds, as opposed to 72 hours for urgent pre-authorization and three to five days for elective procedures with the previous UM process. Encouraged with success of the system, today 15,835 healthcare provider offices use it. [15]

Seattle Children's Hospital and Regional Medical Centre is using big data analytics as part of its Clinical Standard Work (CSW) program, which defines patient populations and recommends an ideal protocol for each population, allowing ensuring that every patient at the hospital receives the same standard of care. The CSW program gets the enormous data from enterprise data warehouse (EDW) which currently integrates data from 10 sources across the hospital, including electronic medical records (EMRs) and billing systems. With the help of CSW program, the doctors and nurses get complete information based on thousands of data points of each patient, they get answers to complex queries about potential treatments and procedures, and identify pathways of care for patients with particular needs, regardless of provider. Clinicians can also evaluate treatment protocols for determining the resources which need to be allocated by hospital. [16]

Seton Healthcare Family, based on Texas, Austin. It has used the IBM Content and Predictive Analytics for Healthcare solutions. The system gives an integrated view of relevant clinical and operational information to drive more informed decision making. By teaming unstructured content (History and Physical, Discharge Summaries, Echocardiogram Reports, and Consult Notes) with predictive analytics, Seton is able to identify patients likely for re-admission and introduce early interventions to reduce cost, mortality rates, and improved patient quality of life.

Doctors at UCLA with the help of IBM Watson Foundations have recently started using data streaming technology in order to make more informed decisions about brain functions and abnormalities. Using IBM Watson foundations, physicians are able to gather data from sensors to analyze brain functions in real time. As a result of the use of this technology, patient care can be substantially improved, and doctors have more time to serve for more patients.

The U.K.'s National Health Service uses cloud analytics software to pluck numerical and text data on health-care facilities from spread sheets and databases and presents it in plain English on its website, NHS Choices. This endeavor is helpful to its citizens, as they can make better choice about their care based on information of about 50,000 health care facilities. The software uses natural language generation techniques and to examine through structured data and automatically present it in story form.

US State of North Carolina processes about 88 million claims totaling about \$12 billion annually from 66,000 providers who treat the state's two million Medicaid patients., The State's

Department of health and human services in collaboration with IBM used big data analytics to help identify suspicious billing patterns by healthcare providers. Using three years' worth of North Carolina Medicaid claims data, IBM data mining software, which featured special algorithms and modeling capabilities, was applied to detect common fraud and abuse schemes. Almost 90% reduction in fraud was achieved.

Informatics for Integrating Biology and the Bedside (i2b2) is an NIH-funded National Centre for Biomedical Computing based at Partners HealthCare System. The i2b2 Centre developed a scalable informatics framework that will enable clinical researchers to use existing clinical data for discovery research and, when combined with IRB-approved genomic data, facilitate the design of targeted therapies for individual patients with diseases having genetic origins.

The Human Connectome project which led by Washington University, University of Minnesota, and Oxford University is working to map the human brain by making a comprehensive connectivity diagram. It will produce invaluable information about brain connectivity, its relationship to behavior, and the contributions of genetic and environmental factors to individual differences in brain circuitry and behavior. This will help to figure reasons why certain people have certain brain disorders, help the physicians to easily diagnose and in certain cases prevention of mental or physical illness. Over a 3-year span (2012-2015), the Human Connectome Project (HCP) has scanned 1,200 healthy

adult people and the generated data is made publicly accessible quarter wise.

In January 2015, President Obama announced a new biomedical research project that is "Precision Medicine" which will use the power of big data to help for the development of specialized drugs to cure the diseases like cancer and diabetes. The program shall collect genetic data of one million Americans so scientists could develop drugs and treatments tailored to the characteristics of individual patients.

### 5. Big Data Analytics Process Steps:

There are different steps in Big Data Analytics process.

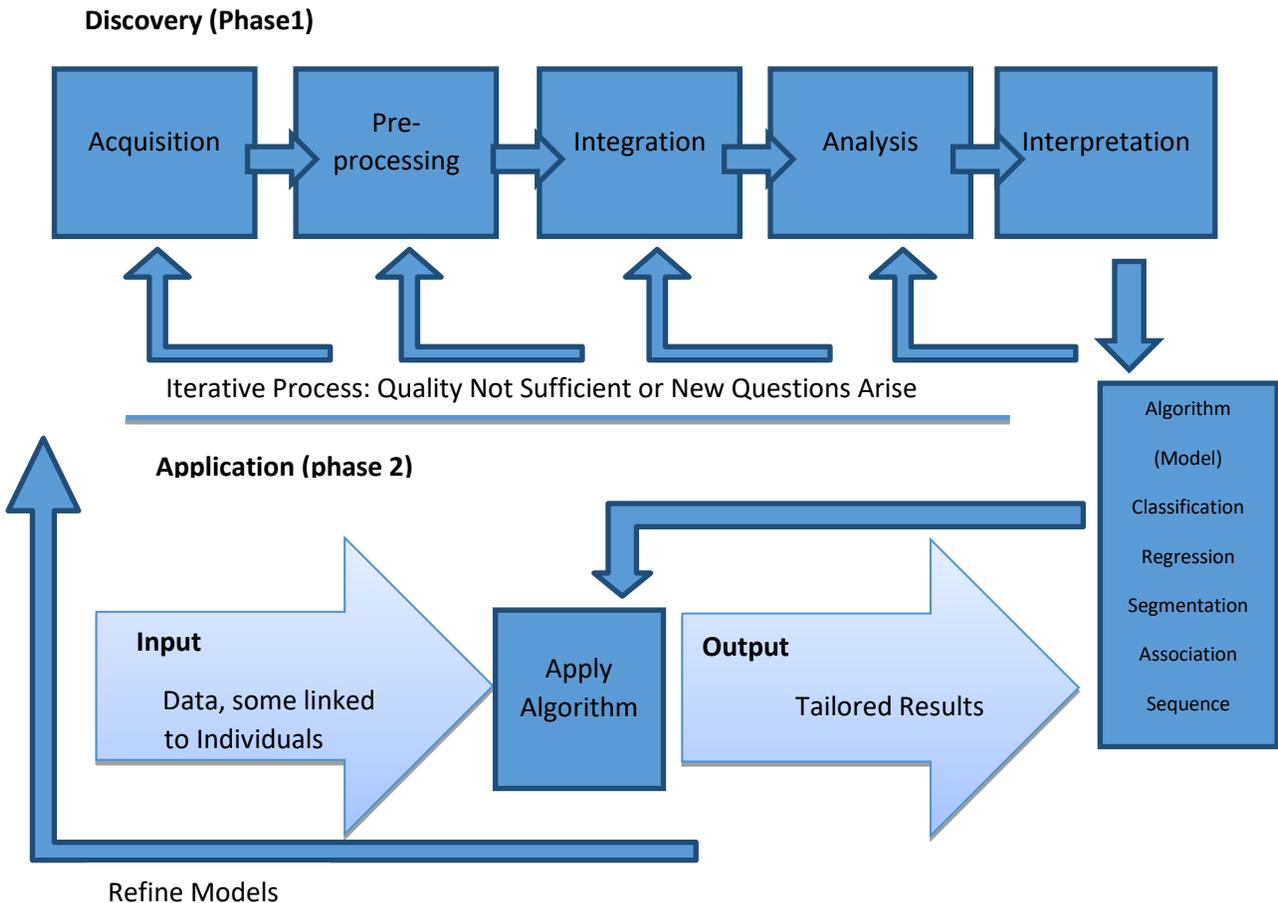


Figure3 Functional Architecture of Big Data Analytics Process

- ii) **Cleaning of Data:** Generally, the healthcare data is seen as flaws like many patients don't share their data completely like data about their dietary habits, weight and lifestyle. In this type of cases the empty fields need to be filled appropriately. Another example, the gender can be either at most one of two values i.e. male or female. In case any other value or no value is present then such entries need to be updated and handled accordingly. The data from sensors, prescriptions, medical image data and social media data need to be expressed in a structured and suitable form for performing effective analysis.[2]
- iii) **Integration of Data:** The BDA process makes use of data where accumulated across various platforms. This data may be varied in metadata (the number of fields, type, and format). The total data can be grouped correctly and consistently into a dataset which can be effectively used for data analysis purpose. This is a very challenging task, considering the big volume and variety of big data.
- iv) **Querying, Analysis and Interpretation of Data:** After cleaning of data and integration, the next step is to query the data. A query can be simple one like what is mortality rate in a particular area? Or complex query like how many patients with diabetes would be likely to develop heart related problems in next 6 years? Based upon the complexity of the query, the data analyst can choose appropriate platform and analytic tool.

The following Figure 3 depicts the functional architecture of Big Data Analytics process steps.

### 5.1 Technology Support for Big Data Analytics in Health Care:

There are large number of open source and proprietary platforms and tools available in the market. Some of them are Hadoop, Map Reduce, Storm, Grid Grain. Big Data Databases like Cassanadra, HBase, Mongo DB, Couch DB, Orient DB, Terrastore, Hive etc. Data Mining tools like RapidMiner, Mahout, Orange, Weka, Rattle, and KEEL etc. File Systems like HDFS and Gluster. Programming Languages like Pig/PigLatin, R, and ECL. Big Data Search Tools like Lucene, Solr etc. Data Aggregation and Transfer Tools like Sqoop, Flume, and Chukwa. Other tools like Oozie, Zookeeper, Avro, and Terracotta. Some open source platforms are also available like Lumify, IKANOW [11].

The criteria for platform evaluation may be varied for different organizations. Generally the ease of use, availability, the capability to handle voluminous data, support for visualization, high quality assurance, cost, security can be some of the variables to decide upon the platform and tool to be used. Some of the platforms and tools are mentioned the following Table

Table 1 Platforms & Tools for Big Data Analytics in Healthcare

Platforms & Tools	Description
The Hadoop Distributed File System (HDFS)	HDFS enables the underlying storage for the Hadoop cluster. It divides the data into smaller parts and distributes it across the various servers/nodes.
MapReduce	MapReduce provides the interface for the distribution of sub-tasks and the gathering of outputs. When tasks are executed, MapReduce tracks the processing of each server/node.
PIG and PIG Latin	Pig programming language is configured to assimilate all types of data (structured/unstructured, etc.). It is comprised of two key modules: the language itself, called PigLatin, and the runtime version in which the PigLatin code is executed.
Hive	Hive is a runtime Hadoop support architecture that leverages Structure Query Language (SQL) with the Hadoop platform. It permits SQL programmers to develop Hive Query Language (HQL) statements akin to typical SQL statements.

Jaql	Jaql is a functional, declarative query language designed to process large datasets. To facilitate parallel processing, Jaql converts “high-level” queries into ‘low-level’ queries” consisting of MapReduce tasks.
Zookeeper	Zookeeper allows a centralized infrastructure with various Services, providing synchronization across a cluster of servers. Big Data analysis applications utilize these services to coordinate parallel processing across big clusters.
HBase	HBase is a column-oriented database management system that sits on top of HDFS. It uses a non-SQL approach.
Cassandra	Cassandra is also a distributed database system. It is designated as a top-level project modeled to handle big data distributed across many utility servers. It also provides reliable service with no particular point of failure ( <a href="http://en.wikipedia.org/wiki/Apache_Cassandra">http://en.wikipedia.org/wiki/Apache_Cassandra</a> ) and it is a NoSQL system.

Oozie	Oozie, an open source project, streamlines the workflow and coordination among the tasks.
Lucene	The Lucene project is used widely for text analytics/searches and has been incorporated into several open source projects. Its scope includes full text indexing and library search for use within a Java application.
Avro	Avro facilitates data serialization services. Versioning and version control are additional useful features.
Mahout	Mahout is yet another Apache project whose goal is to generate free applications of distributed and scalable machine learning algorithms that support big data analytics on the Hadoop platform.

## 6. Big Data Analytics Benefits in Healthcare

The massive amount of data provides the opportunities for researchers in the Healthcare field to use tools and techniques for opening the hidden answers. Big Data Analytics tools and techniques can be applied in effective way on large sets of data then the following benefits will be given:

**i) Individuals/Patients:** Generally, when treatment is given to a patient, then the historical data can be considered such as a set of similar patients about the symptoms, drugs used outcome/response of different patients. With the help of BDA, the specific treatment is given for a patient based on his genomic data, location, weather, lifestyle, medical history, response to certain medicines, allergies, family history etc. When the genome data is fully explored for some kind of relation and it can be established between the DNA and a particular disease. Then the specific line of treatment can be constructed for every individual. The patients will benefit in the following ways:

- Correct and effective treatment can be applied.
- Health related issues will be known in better way.
- Preventive steps can be taken in time.
- Continuous health monitoring at patients location using wearable wireless devices.
- Designing specialized treatment for patient.
- Life expectancy and quality will be found in advance.

**ii) Hospitals:** By using effective BDA techniques on the data availability, the hospitals can get following benefits:

- Predict the patients staying and readmission information.
- New healthcare plans will be developed to prevent hospitalization.
- Various questions can be answered by analyzing the data using BDA tools and techniques regarding disease treatment.
- The hospital management can take and manage administrative decisions in the better way.

**iii) Insurance Companies:** The government is reimbursed the large amount of expenditure for giving medical claims for patients. We can analyze, identify, predict and minimize the possible frauds related to medical claims by using BDA. [3]

**iv) Pharmaceutical Companies:** By using BDA techniques effectively, the R&D can help pharmaceutical companies to produce drugs that may be most effective for treating a specific disease with in the shorter period.

**v) Government:** The BDA can help in improving the public health surveillance and speed up the response to disease outbreaks. The government can use demographic data, historical data of disease outbreak, weather data, data from social media over disease keywords like cholera, flu etc. BDA can analyze this massive data to predict epidemics, finding correlation between the weather and likely occurrence of disease. Therefore preventive measures can be taken to avoid the same. [3]

## 7. Big Data Analytics - Challenges:

The advantages of big data are more for healthcare, but there are number of challenges which can be broken up.

**i) Unstructured and Provenance of Data:** The BDA process can collect data from different sources. Most of the data is unstructured data like medical prescriptions, blogs, tweets, status updates, and comments. It is necessary to generate right metadata for this unstructured data and transform it into a structured format. The image and video data should be structured for semantic content and search. By using data analysis process, the provenance of data along with its metadata should be carried out so it is easy to track the processing steps when error generates [3]. Some intelligent processing techniques should be proposed to deal the data input from sensors and wearable devices. This will help to filter/derive the meaningful data, which can then be stored on permanent storage. Therefore it will save space.

**ii) Missing or Incomplete Data:** Some patients may hide their personal information about his/her life style at the time of filling forms or oral interviews by doctors. Some fields may be empty at the time of storing the data in digital format. Sometimes it may happen that some of the fields produce wrong results. If analysis is done on the empty or wrong fields of data, then it may or may not get processed. In both the cases they produce wrong results. If we leave some records as empty then the analysis may not on cumulative data. If we take wrong value fields then the analysis is incorrect and unreliable. This type of issues will be addressed.

**iii) Quality of Data:** When we consider data from social media, then we need to ensure that data whether it is a valid data or not. So it is great challenge to determine the validation and quality of data.

**iv) Technical Challenges:** There are different technical challenges.

- Data aggregation with different database management systems is also a great challenge in BDA. By dividing certain standard database design practices meant for a

specific domain like healthcare, financial sector etc., it can be made easier [3]. We are required more technological standards and protocols for different database management systems to integrate seamlessly.

- The traditional algorithms can be scaled up to handle the big volume of data in data mining processes or analysis. The processors speed has come to a point beyond which it's hard to increase in parallelogram process. So the trend can be moved towards multi-core processors. In such a scenarios, we need statistical algorithms which can be parallelized otherwise the computing performance will decrease when they handle complex big volume data.[9] The interactive response time is another big problem while apart from this scaling complex query processing techniques to terabytes. [3]
- An analysis is more useful if a non-technical person is able to understand and interpret it. The large volume and variety of data is too hard to represent it visually in a more understandable and easy way. A user should be able to perform the repeated analysis with the different set of assumptions, data sets and parameters. It will help the user to better understand the analysis process and verify whether the system works in a required way or not.
- We need careful evaluation process to use the best platform and tool for market floods.

v) **Data Security:** Data Security is another major challenge as more and more data is digitized. Most of the people are not willing to share their personal data with a fear of security breach. If there is assurance for data security, then the problem can be managed. There should be strict government policies and norms for what data can be shared and what not. Apart from this, strong technological hardware and software level security precautions and measures should be implemented to prevent the hacking and interpreting malicious code.

vi) **Lack of Experts:** There is a more shortage of qualified and experienced data scientists in the world. So it is necessary to create an expertise in the field of data science to turn the promises of big data into reality.

## 8. Innovative Ideas and Solutions:

The following are some possible new innovative ideas and solutions of Big Data in Healthcare industry.

- **Clinical Decision Support:** BDA technologies predict outcomes or recommend alternative treatments to clinicians and patients at the point of care by understanding, analyzing, categorizing and learning from them.
- **Personalized Care:** By predicting and analyzing disease symptoms in advance personalized care is taken (e.g., genomic DNA sequence for cancer care) in real time to highlight best practice treatments to patients. These solutions may offer early detection and diagnosis before a patient develops disease symptoms.
- **Public And Population Health:** BDA solutions that can help in searching and identifying patient population via social media data to predict flu outbreaks based on consumers' search, social content and query activity.

BDA solutions can also help clinicians and epidemiologists performing analyses across patient populations and care venues to help identify disease trends.

- **Clinical Operations:** BDA can produce accurate solutions for clinical operations without waiting for longer time to take fast decisions.
- **Policy, Financial and Administrative:** BDA has supported the decision makers to integrate and analyze data related to key performance indicators on policy and financial aspects.

## 9. Conclusion and Future Work

Big Data Analytics in healthcare is evolving into a promising field for giving new insights from huge data sets and improving results while reducing costs. Its strength is high; however there are more challenges to overcome. Big Data Analytics has the potential to transform the way healthcare providers from traditional ways to more suitable and right tools and technologies to gain insight from their clinical and other data repositories and make constructive decisions. In the future we'll see the rapid, widespread implementation and use of big data analytics across the healthcare organizations and the healthcare industry. To that end, the challenges must be discussed and see the overcoming measures. As big data analytics become more important, more attention will be required, due to some issues such as guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies. Big data analytics and applications in healthcare are at an initial stage of development, but rapid advancements of Big Data platforms and tools can accelerate their maturing process.

## Conflict of Interest

The authors declare no conflict of interest.

## Acknowledgment

I would like to thank to all people who help me prepare this paper completely. I would also thank to my guide who help me and get proper suggestions. I would like to thank to all website and journal papers which I have referred to create my review paper successfully.

The authors would like to thank all reviewers and Prof. Passerini Kazmerski, Editor for his valuable comments on the manuscript.

## References

- [1] Jasleen Kaur Bains, "Big Data Analytics in Healthcare- Its Benefits, Phases and Challenges" , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 4, April 2016, Available online at: [www.ijarcsse.com](http://www.ijarcsse.com)
- [2] Wullianallur Raghupathi and Viju Raghupathi, "Big data analytics in healthcare: promise and potential", *Health Information Science and Systems* 2014, 2:3, Available: <http://www.hissjournal.com/content/2/1/3>
- [3] Vivek Wadhwa, "The rise of big data brings tremendous possibilities and frightening perils", April 2014. Available: <http://www.washingtonpost.com/blogs/innovations/wp/2014/04/18/therise-of-big-data-brings-remendous-possibilities-and-frightening-perils/>
- [4] D. Agrawal et. al, "Challenges and Opportunities with Big Data", Big Data WhitePaper-Computing Research Association, Feb-2012, Available: <http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>
- [5] Nambiar, R. ; Cisco Syst., Inc., San Jose, CA, USA ; Bhardwaj, R. ; Sethi, A. ; Vargheese, R., "A look at challenges and opportunities of Big Data

- analytics in healthcare”, IEEEConference 2013, Available: [http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6691753&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D6691753](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6691753&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6691753)
- [6] Ahmed E. Youssef, “A Framework for Secure Healthcare Systems Based on Big Data Analytics in Mobile Cloud Computing Environments”, *The International Journal of AmbientSystem and Applications* 06-2014, Available: <http://airccse.org/journal/ijasa/papers/2214asa01.pdf>
- [7] J. Archenaa, E.A. Mary Anita, “A Survey of Big Data Analytics in Healthcare and Government”, *Procedia Computer Science*, Elsevier, Volume 50, 2015, Pages 408–413, Big Data, Cloud and Computing Challenges, Available: <http://www.sciencedirect.com/science/article/pii/S1877050915005220>
- [8] Matthew Herland, Taghi M Khoshgoftaar and Randall Wald, “A review of data mining using bigdata in health informatics”, Herland et al. *Journal of Big Data* 2014, Springer, 1:2 Available: <http://www.journalofbigdata.com/content/1/1/2>
- [9] MH Kuo, T Sahama, AW Kushniruk, EM Borycki, DK Grunwell, —“Health big data analytics: current perspectives, challenges and potential solutions”, *International Journal of Big Data Intelligence*, Vol. 1, Issue 1, pp.114-126.
- [10] Bernard Marr, “How Big Data Is Changing Healthcare”, Available: <http://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/>
- [11] “Improve Healthcare Win \$3,000,000”, Available: <http://www.heritagehealthprize.com/c/hhp>
- [12] Cynthia Harvey, “50 Top Open Source Tools for Big Data”, Available: <http://www.datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html>
- [13] “Big Data Provides True Picture of Diabetic Population”, Available: [http://www.sas.com/en\\_us/news/sascom/2014q1/nz-ministry-of-health.html](http://www.sas.com/en_us/news/sascom/2014q1/nz-ministry-of-health.html)  
[http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SWGE\\_YT\\_YT\\_USEN&htmlfid=YTC03753USEN&attachment=YTC03753USEN.PDF](http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SWGE_YT_YT_USEN&htmlfid=YTC03753USEN&attachment=YTC03753USEN.PDF)
- [14] Health Analytics, Available: <http://datascience.columbia.edu/health-analytics>
- [15] [http://www.ibm.com/smarterplanet/us/en/ibmwatson/assets/pdfs/WellPoint\\_Case\\_Study\\_IMC14792.pdf](http://www.ibm.com/smarterplanet/us/en/ibmwatson/assets/pdfs/WellPoint_Case_Study_IMC14792.pdf)
- [16] Linda L. Briggs, “BigData means better care at Seattle's Children Hospital”, Available: <http://tdwi.org/articles/2013/08/13/big-data-analytics-smarter-care.aspx>
- [17] Kiyana Zolfaghar, Naren Meadem, Ankur teredesai, Senjuti Basu Roy, Si-Chi Chin. “Big Data Solutions for Predicting Risk-of-Readmission for Congestive Heart Failure Patients”. 2013 IEEE International Conference on Big Data, 978-1-4799-1293-3/13. <http://dx.doi.org/10.1109/bigdata.2013.6691760>
- [18] Joseph M. Woodside. Virtual Health Management, 2014 11th International Conference on Information Technology New Generations 978-1-4799-3187-3/14. <http://dx.doi.org/10.1109/itng.2014.124>