

# Applications of Big Data Analytics in the Field of Medical Diagnosis : A Review

Mubashir Hussain, Jatinder Manhas

Department of Computer Science and Information technology, BCJU, University of Jammu, India

## ABSTRACT

Big data science has revolutionized almost every field in today's world including healthcare sciences which has completely utilized the potential of BDA in enhancing human healthcare by reducing the number of readmission, better decision making, proper medication etc. meanwhile reducing the healthcare cost. The main focus of this paper is to provide a deep insight into the applications of data sciences in the field of medical sciences by investigating the previous studies made in this field. This study has focused on analyzing relevant previous research articles about medical diagnosis and analytics by searching in several popular databases in order to understand the healthcare topic and conduct our literature review. Methods and technology progress about Big Data are presented in this study that has been used to improve healthcare performance in many areas such as: medical operations, reports, decision making, and prediction and prevention system. This paper provides a general literature survey of recent progress and advances in data science, healthcare and biomedical research. In the final part of the paper, the most significant conclusions and suggestions are offered.

**Keywords :** Big Data Science, Medical Sciences, Big Data Analytics (BDA), Value-based Learning, Healthcare, Statistics.

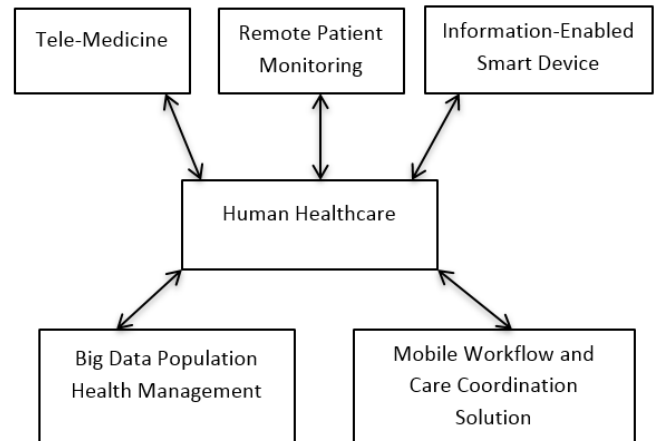
## I. INTRODUCTION

Data science term coined back in 1996 [wiki] is no more a news now. The highly paid job in the next ten years will be statisticians [Hal Varian, Chief Google Economist]. Reports say data from the U.S. healthcare system alone reached, in 2011, 150 exabytes. At this growth rate, big data for U.S. healthcare will soon reach the zettabyte (10<sup>21</sup> gigabytes) scale and, not long after, the yottabyte (10<sup>24</sup> gigabytes) [1]. Kaiser Permanente, the California-based health network, which has more than 9 million members, is believed to have between 26.5 and 44 petabytes of potentially rich data from EHRs, including images and annotations [1]. The U.S will need 140,000-190,000

predictive analysts and 1.5 million managers/analysts by 2018 [McKinsey Global Institute's June 2011].

Data science is an interdisciplinary field in which processes, algorithms and methods are used in collaboration to manipulate, extract and interpret knowledge from tremendous amount of data [wiki]. Data science principles apply to all data-small or big. Theories and techniques from various fields and disciplines are used to investigate and analyze a large amount of data to help decision makers in many industries such as science, engineering, economics, politics, finance and education. The rapidly increasing field of data sciences has influenced every corner of life as mentioned includes health industry which to some extent also has grasped its potential benefits. Over the last few years, healthcare industry

has generated large volumes of data driven by administrative record keeping, biomarker data, clinical trials, compliance and regulatory requirements, and patient care. It has moved from volume based business to value based business leading to more efficiency and productivity. Benefits of health-related Big Data have been demonstrated in three areas, namely to 1) prevent disease, 2) identify modifiable risk factors for disease, and 3) design interventions for health behavior change [2]. Tomorrow's healthcare may look more efficient thanks to things like electronic health records. It also may look a lot more effective. Reduced readmissions, better care, and earlier detection are on the horizon. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of data (known as 'big data') hold the promise of supporting a wide range of medical and healthcare functions, including among others clinical decision support, disease surveillance, and population health management [2][3][4][5]. As for example, Cancer is an incredibly complex disease; a single tumor can have more than 100 billion cells, and each cell can acquire mutations individually. The disease is always changing, evolving, and adapting. By employing the power of big data analytics and high-performance computing and leverage sophisticated pattern and machine learning algorithms patterns that are potentially linked to cancer can be identified. Big data analytics helped to healthcare sector upgrade by implementing epitomize medicine and prescriptive analysis, hospital liability interference and predictive analysis, dissipation and responsibility, changeability reduction, automatic extraneous and constitutional exposure of patient record, regulated health conditions and patient registries and disintegrated end solution [6].



**Fig. 1.** Big Data Revolution in Medical Science

### Need for Big Data Analytics in Healthcare

Analytics is increasingly weaving itself into the fabric of healthcare and will fundamentally shape the future of medicine and care delivery. With opportunities such as the ability to improve the efficiency of healthcare while improving the quality of care, mine genetic data, reduce costs, effectively respond to disasters, and numerous other goals, the application of analytics is broad and far [7]. The need for data analytics in healthcare can be defined as follows:

- 1) **Patient centric healthcare services:** Enhancing the quality of healthcare services and reducing the number of patient re-admission by examining the previously obtained data from similar patients into actionable information that can be used to identify needs; provide services and predict and prevent crises. Determine more clinically relevant and cost effective ways to diagnose and treat patients.
- 2) **Evidence based treatment and medication:** Analyzing various varieties of data (structured, semi-structured and unstructured) obtained through various sources like clinical data, EMR's, genomic data, operational and financial data, in-hospital and in home devices data etc. to match treatment/cure parallel to the outcomes, identification of patients at risk of adverse

outcomes, customized care and personalized medicine.

- 3) **Remote patient disease monitoring:** Collecting and analyzing the real time and large volumes of data from in-home and in-hospital devices for better monitoring, disease outbreak, identifying adverse event and public health surveillance
- 4) **Real time patient disease analysis:** Application of modern methods of data analytics on patient disease profiles to identify advantages for patients who would benefit pro-active care, precise diagnose and elimination of unnecessary and duplicate care and thereby reducing wrong treatment and cost.
- 5) **Earlier disease detection:** Availability of large volumes of medical data and data analyzing technologies enhances earlier disease prediction by analyzing the earlier related cases of patients and hence diagnosing the later stage of a disease at early. For example, medical biosensor system can be used to detect heart attack many hours before it actually happens [8].
- 6) **Monitoring quality of Medical Institutes:** Regular visits to hospitals by governing officials to review hospital quality as per standards set by the various government agencies in order to improve the quality of treatment and medication.

## II. LITERATURE SURVEY

Wullianallur Raghupathi and Viju Raghupathi [9] in their paper discussed how big data analytics is emerging as a prominent and promising field in medical sciences in enhancing the quality and standards by providing an insight into very large data sets and hence improving outcomes and reducing cost. In their paper they discussed the benefits of 6V's (variety, volume, velocity, veracity, variability and value) properties of big data in healthcare and the need of big data analytics in healthcare to help reduce waste and inefficiency in clinical operations, research and development and public health. In their

paper they outlined an applied conceptual architecture of big data analytics and methodology to analyze very large data sets as providers to gain insight for making better informed health related decisions. They also listed and compared various available platforms/tools that have ability to mine large healthcare datasets (structured as well as unstructured) that are captured from various sources. In the end they addressed various challenges being faced by the data scientists in utilizing the full potential of big data analytics.

Fabricio F. Costa [10] in his research discussed about the increasing availability and flow of healthcare big data leading to an opportunity for personalized medicine programs that will lead to better patient care. Big data analytics is leading from the front in medical sciences in improving clinical practices, drug development and healthcare financing process leading to better patient outcome and care while reducing cost. In his paper he has outlined an architectural framework showing hardware and software requirements to analyze very large data sets as providers to gain insight for making better informed health related decisions. He also listed various companies and institutions that are providing services and solutions to store, analyze and deal with complex biomedical information and institutions that provide solutions to generate, interpret and visualize combined omics and health clinical data. In his paper he also pointed out the breakthroughs achieved in collaborating omics and clinical data to achieve betterment in personalized medicine. In the end he also reviewed challenges associated with using big data in biomedicine that need to be addressed.

Ruchita Gujarathi, Fabricio F. Costa [11] in their paper studied the potential of both online networks and big data analytics in the field of population healthcare and concluded that both the fields are dynamic and emerging that will greatly impact healthcare. In their paper they pointed out that

information technology has enhanced the data capturing ability by leveraging the data exchange in a variety of fields which in healthcare field can be transformed into value with the help of data analytic algorithms in order to improve individual and population health, deliver effective therapies and consequently reduce the cost of healthcare. They have also outlined a structure representing the implementation of digital technologies in collaboration with the BDA to facilitate personalized medicine decisions by analyzing clinical reports for translational medicine that can be used in the clinic by health professionals in order to improve patient care. In their paper they also studied the impact of social networking sites by analyzing the behavior of people towards medical and pharmaceutical devices, medicines etc. and the collaboration between third party networking sites and pharmaceutical companies, biotechnology firms, medical device manufacturers etc. and concluded that this obtained information can be used by companies to facilitate drug discovery, execute market research cheaper and faster, providing information to speed up clinical trials, enable personalized medicine programs to be publicized among people and give patients enough basic knowledge to make their own healthcare decisions.

**S. M. Riazul Islam, Daehan Kwak, Md.Humaun Kabir, Mahmud Hossain, And Kyung-Sup Kwak [12]** in their research discussed how internet of things (IoT) revolution is redesigning the modern healthcare with promising technological, economic and social prospects. In their paper they surveyed diverse aspects of IoT based technologies and presented various healthcare network architectures and platforms that support access to IoT backbone. The paper considered various security requirements, research problems and challenges and to propose a model to that can overcome the associated risks. The paper offers a broad view on how recent and ongoing advance technologies have enhanced the quality of

healthcare services by providing affordable healthcare gadgets. The paper presented an intelligent collaborative security model to minimize security risk and also discussed how different innovative technologies like big data analytics, ambient intelligence and wearables can leveraged in enhancing the quality of healthcare. The article also discussed eHealth and IoT policies that can be beneficial to various stakeholders interested in accessing IoT based healthcare technologies.

**Lidong Wang and Cheryl Ann Alexander [13]** in their article studied the applications and importance of big data analytics in healthcare sciences and how it enables doctors to get a 360-degree view of the patient to analyze and predict outcomes. In their paper they discussed the promising benefits of big data as it can improve clinical practices, drug development and healthcare financing process but there are many challenges like consolidating and processing segmented or siloed data, aggregating and analyzing unstructured data, indexing and processing continuously streaming data, privacy, data leakage, information security and lack of infrastructure and unified standards, etc. that need to be addressed. In their paper they outlined an applied conceptual architecture of big data analytics and methodology to analyze very large data sets as providers to gain insight for making better informed health related decisions.

**J.Archenaa and E.A.Mary Anita [14]** pointed out the need for BDA in healthcare and government sectors as these are among the leading areas where large volumes of data are generated from which additional value can be uncovered with proper data analytics methods. In their paper they also laid out architecture to implement big data ecosystem to handle large scale data processing and analytics on it. BDA can play an effective role in analyzing large volumes of real time data to make decisions in emergency situations. They revealed that there is not the lack of data but lack of information that can be

employed to support decision making, strategy and planning. They mention that entire government system can realize the benefits of BDA but there is a need to devote time, allocate budget and resources to visioning and planning.

**Mohammad Ahmad Alkhatib, Amir Talaei-Khoei, Amir Hossein Ghapanchi [15]** in their study conducted a peer review of the research going in the field of healthcare data analytics to encourage professional, doctors, medical staff etc. to adopt and utilize technologies to assist healthcare analytics and to enhance the quality and standards to gain insight for making better informed health related decisions. They searched popular databases for healthcare data analytics article and research papers and presented a summarized progress over past few years in healthcare sciences utilizing big data analytical techniques. A total of 43 papers were reviewed out of which 31 were related to healthcare analytics in which the focus of study was related to decision making, prediction and prevention of disease, clinical delivery, clinical operations, performance monitoring and reporting, diagnosis and treatment and healthcare information exchange. In their paper they also pointed out that most of the research in this field is restricted to western countries particularly US where most number of publication were found based on geographical distribution of research conducted. They also listed various data analytical tools and techniques that are improving healthcare performances in many areas such as medical operations, medication, decision making etc. In the end he also reviewed issues associated with using big data analytics that need to be addressed.

**Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian [16]** in their study discussed major challenges and promising areas of medical big data analytics like signal processing, image processing and

genomics. In their paper they figured out that the data generated using techniques such as medical images (like PET-CT, X-Rays, MRI's, molecular imaging), signal processing and genomics used for diagnosis generate large volumes of data that require large storage capacities and demand fast and accurate algorithms if any decision assisting automation were to be performed using the data. In their research they pointed out that medical big data analytics has the potential to play a significant role in enhancing the standards of healthcare policy, improve care delivery, expand access to healthcare, align payment with performance, and help curb the vexing growth of healthcare costs.

**Liang Y and Kelemen A [17]** in their research article put forward the impact and potential of big data science in healthcare and discussed various opportunities and challenges. In their paper they discussed that healthcare and biomedical sciences have become data-intensive investigator generating and using large, complex, high dimensional and diverse domain specific datasets. They surveyed the recent advances in big data analytics and presented a detailed discussion of basic infrastructure needed, skilled disease expert talents, adaptive and intelligent analytical tools and smart utilization of resources are keys in realizing the benefits of data sciences for actionable healthcare decision making. They also listed and compared various available platforms/tools that have ability to analyze large healthcare datasets. In their paper they also illustrated an example of how big genome data is transforming data into diagnostics, therapeutics and new insights into population health and disease treatment.

**Vishnu S Basuthkar, Chetana Srinivas[18]** in their paper on comprehensive study of cost effective BDA applications in healthcare discussed about the impact of BDA in this field and pointed out various tools and methodology that could be beneficial in mining value from the large amount of data flow in healthcare field. They searched popular IEEE Xplore for healthcare

data analytics article and research papers and presented a summarized progress over past few years in healthcare sciences utilizing big data analytical techniques. In this paper, many tools of BDA are described to solve the issues in HBD and improve its efficiency. In the end they also discussed various research gap that is currently present in employing the full potential of healthcare big data and point out the possible future scope of BDA in healthcare sciences.

**Jingwei Li, Wei Ding, Hsing Kenneth Cheng, Ping Chen, Dehai Di, Wei Huang [19]** in their paper made a comprehensive and systematic literature review of published papers in the field of BDA in healthcare sciences from last 20 years, analyzing 321 articles. They discussed the role previously and currently played by the data analytics in transforming the shape of human healthcare in improving the standards of healthcare policy, improve care delivery, expand access to healthcare, align payment with performance, and help curb the vexing growth of healthcare costs. They presented a summarized analysis of the research done in this field by searching for the related articles in various journals, PhD dissertations, conference proceedings etc. and classified the data (a total of 214 selected papers) on the basis of field of study such as omics, medical specialties, IT healthcare etc. in which it was carried out. They find out that only 14.5% of research was done in omics (genomics, proteomics), 14.1% in medical specialties (neurology, imaging, pharmaceutical etc.) and 73.4% of total research was done in enhancing IT healthcare (HER, bioinformatics, Web mining etc.). In their study they also classified the research articles into 13 different sub-categories in which they were carried out. In the end they suggested that research in healthcare is becoming more and more popular particularly in the field of IT healthcare.

**Siuly Siuly, Yanchun Zhang [20]** in their article case studied the importance of big data analytics in neurological diseases diagnosis and pointed out that

data generated using techniques such as brain wave tests (Electroencephalography or EEG), medical images (like PET-CT, X-Rays, MRI's, molecular imaging), signal processing and genomics used to detect, manage and treat neurological diseases generate large volumes of data that require large storage capacities and demand fast and accurate algorithms if any decision assisting automation were to be performed using the data with high precision in order to improve consistency of diagnosis and increase the success of treatment, save lives and reduce cost and time. They also discussed the various medical technologies used in the neurological disease diagnosis and various challenges faced in medical big data analysis. They also introduced the CAD (Computer Aided Diagnosis) systems and briefly described on how these systems work for neurological disease diagnosis in assistance with big data analytics algorithms.

**Mathieu Guillermin, Thierry Magnin [21]** conducted their research about the ethical issues concerning big data analytics in biomedical sciences and personalized medicine. They pointed out that there is no doubt that BDA is taking biomedical science to new heights but data driven science and their technological applications are raising many serious ethical questions particularly about privacy protection. In their paper they discussed about the ethical threats caused due to data fundamentalism such as reductionism, over-medicalization, or automated but dehumanized medicine and indispensability of human intelligence. Their analysis showed that big data processing can generate value in the field of biomedical research and healthcare only when coupled with human intelligence and its ability to rationally perform non-neutral decisions.

**Muhammad Umer Sarwar, Muhammad Kashif Hanif, Ramzan Talib, Awais Mobeen, and Muhammad Aslam [22]** in their research studied the advancement made in the field of healthcare sciences using BDA. In their paper they presented a comparative study of

various methods of machine learning that are being used to carry out analysis and predictions in various fields. They discussed in details the process of data mining and its various stages such as feature selection and evaluation, classification and clustering and the method to carry out these stages with their merits and demerits. They suggested that existing traditional machine learning methods take much time in computation when data set volume increase. Similarly, in healthcare domain huge data is collected from different sources and researchers always try to make problem simpler to patient.

**Md Saiful Islam , Md Mahmudul Hasan, Xiaoyi Wang , Hayley D. Germack and Md Noor-E-Alam [23]** in their paper presented a review of the earlier studies made in the field healthcare and biomedical science using big data analytics. They followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) to conduct a database search in order to provide a systematic view of development in this field by selecting critical study elements like healthcare sub-areas, data mining techniques, types of analytics, big data and data sources. In their paper they suggested that there is a need of perspective analytics and integration of domain expert knowledge in the decision making as most of the existing literature mostly examines analytics in clinical and administrative decision making. In their paper they classified the literature by types of analytics, application areas and mining techniques and found that there is no comprehensive review available that can provide complete picture of data mining application in the healthcare industry. In their paper they highlighted some promising recommendation for future research including integration of domain expert knowledge, decrease prediction error and integration of predictive models in reality in healthcare data analytics in order to facilitate coordinated and well informed healthcare systems capable of ensuring maximum patient satisfaction.

**Karim Abouelmehdi, Abderrahim Beni-Hessane and Hayat Khaloufi [24]** studied about the privacy and security issues associated with using big healthcare data. In their paper they revealed that though big data has led healthcare industry to new heights by improving patient outcome, predict outbreak of epidemics, better decision making, reduce healthcare costs etc. there are many challenges associated in preserving security and patient's right to privacy that need to be taken care of. They surveyed and listed various security and privacy challenges in big data as applied to healthcare industry and discussed ways in which they can be addressed. They also listed a detailed description of various data protection laws in some countries along with their features. They presented a detailed summary of the ongoing research in big data privacy by analyzing various articles published all over the world. They presented privacy and security issues in each phase of big data life cycle along with advantages and flaws of existing technologies in the context of big healthcare data privacy and security.

**Blagoj Ristevski and Ming Chen [25]** in their paper presented a brief discussion on big data, its properties and usefulness of BDA in transforming the shape of human healthcare. They pointed out that BDA in healthcare is a promising process of integrating, exploring and analyzing large heterogeneous complex datasets consisting of biomedical data, clinical data, experimental data, EHR data etc. on which data mining techniques can be applied to reveal hidden patterns and novel knowledge that has the potential to play a significant role in enhancing the standards of healthcare policy, improve care delivery, expand access to healthcare, align payment with performance, and help curb the vexing growth of healthcare costs. In the end they also discussed various challenges and security and privacy issues associated with BDA in healthcare sciences that need to be resolved.

**Senthilkumar SA, Bharatendara K Rai, Amruta A Meshram, Angappa Gunasekaran,**

**Chandrakumarmangalam S [26]** in their paper presented a systematic literature review of papers published based on BDA in healthcare sciences and outlined various challenges such data security and leakage, data heterogeneity, velocity etc. that are being faced by researchers. They study revolved around major experiments on data governance in healthcare and suggested few new data visualization tools to the healthcare analyst to make effective decision making. They searched a total of 10496 published articles from the large publishers' viz. Science Direct, PubMed, Springer, Taylor, & Francis, Inderscience and other reports based on BDA in the field of healthcare sciences from year 2010-2015 and found that on 76 papers were relevant meeting all inclusion criteria and were retained. In their paper they also laid out architecture to implement process of big data analysis in healthcare management. They also listed and compared various available platforms/tools that have ability to analyze large healthcare datasets.

**Gemson Andrew Ebenezer J and Durga S [27]** in their paper based on survey of big data analytic in healthcare discussed how big data analytics is emerging as a prominent and promising field in medical sciences in enhancing the quality and standards by providing an insight into very large data sets available because of numerous technologies used in medical institutes; hence improving patient care and clinical decision support. Rate of electronic health records (EHR) in healthcare institutes are growing exponentially in both inpatient and outpatient aspects enabling resource for deriving

value that leads to enable advanced detection of powerful treatment, better clinical decision support and accurate predictions of who is likely to get sick. In their paper they also discussed and compared various big data platforms such as Google Cloud Services, Amazon S3, Microsoft Azure etc. and algorithmic techniques such as C4.5, k-means, Apriori, Support Vector Machines, Naïve Bayes, EM, CART, etc. currently available that are being used. They also pointed out various machine learning algorithms that are being employed by therapists to perform predictive analysis useful in identifying if someone may be at risk for readmission or is on a serious recession. In the end they discussed various challenges that are being faced in perform big data analytics in healthcare such as availability, ease of use, scalability, level of security and continuity, data heterogeneity, data incompleteness etc. that need to be addressed.

**Nivedita Das, Leena Das, Siddharth Swarup Rautaray and Manjusha Pandey [28]** in their paper discussed the area of big data analytics and its application in medical domain. Systematic analysis of extensive data can have to detect pattern so that the proper treatment for individuals and could track outcomes. In their research they pointed out that medical big data analytics has the potential to play a significant role in enhancing the standards of healthcare policy. They discussed various challenges being faced by the data scientists such as interoperability, manageability, security, development, reusability and maturity etc. that need to be addressed in order to utilize the full potential of big data analytics.

**Table 1.** Summarized research work done in Healthcare Big Data Analytics

S.NO.	Author	Problem Focused On	Work Done	Outcome
1.	Wullianallur Raghupathi and Viju Raghupathi	Promises and potential of Big Data Analytics in Healthcare.	Benefits, outlines an architectural framework and methodology, describes examples reported in the literature.	Framework for healthcare service.



2.	Fabricio F. Costa	Potentials and Challenges of Big Data in Biomedicine.	Platforms/tools available, combination of omics and clinical data, challenges associated.	Theoretical discussion on problems and consolidation in healthcare delivery
3.	Ruchita Gujarathi, Fabricio F. Costa	Impact of Online Networks and Big Data in Life Sciences.	Technology enhanced model for BDA analysis, Implications of Personalized medicine, issues associated.	With the future of IT and social and techniques discussion of big data, modeling, and machine learning and simulation
4.	S. M. Riazul Islam, Daehan Kwak, Md.Humaun Kabir, Mahmud Hossain, And Kyung-Sup Kwak	Comprehensive survey of IoT based technologies for healthcare.	Architectures/platforms, applications, and industrial trends in IoT-based health care solutions, IoT security and privacy.	eHealth and IoT policies, Secured access control in healthcare, address issues and challenges.
5.	Lidong Wang and Cheryl Ann Alexander	Big Data concept and characteristics, health care data and some major issues of Big Data.	Big data analytics applications and its issues.	Theoretical discussion on problems and consolidation in healthcare delivery
6.	J.Archenaa and E.A.Mary Anita	Comprehensive survey of Big data analytical technologies in healthcare and government.	Study on big data in healthcare, presented a multilevel scalable Hadoop framework for healthcare applications of big data.	Provides scope for the future work in big data healthcare.
7.	Mohammad Ahmad Alkhatib, Amir Talei-Khoei, Amir Hossein Ghapanchi	Literature review, BDA and data mining, Methodology, tools/platforms available.	Big data analytics applications and its issues.	Future scope, proposed methodology and techniques to enhance scope of BDA in healthcare.
8.	Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian	Big Data concept and characteristics, health care data and some major issues of Big Data analytics.	Major challenges with a focus on three upcoming and promising areas of medical research: image, signal, and genomics based analytics.	Promising scope of BDA in medical imaging, medical signal analytics and genomics.
9.	Liang Y and Kelemen A	Reviewing applications and challenges of big data in healthcare research.	Big data analytics applications and its issues.	Provides scope for the future work in big data healthcare.
10.	Vishnu S Basuthkar, Chetana Srinivas	Comprehensive Literature review, BDA and data mining, Methodology, tools/platforms available.	Discussions on previous research done, applications and issues of BDA in healthcare.	Theoretical discussion on problems and consolidation in healthcare delivery
11.	Jingwei Li, Wei Ding, Hsing Kenneth Cheng, Ping Chen, Dehai Di, Wei Huang	Comprehensive Literature review, BDA and data mining, Methodology, tools/platforms available.	Discussions on previous research done, applications and issues of BDA in healthcare.	Theoretical discussion on problems and consolidation in healthcare delivery
12.	Siuly Siuly, Yanchun	Case study of Medical big	Medical imaging, medical	Provides scope for the

	Zhang	data analytics in neurological diseases diagnosis.	signal analytics, Computer aided diagnosis and various methods available, challenges.	future work in big data healthcare.
13.	Mathieu Guillermin, Thierry Magnin	Big Data for Biomedical Research and Personalized Medicine, Epistemological and Ethical Cross-Analysis	Ethical issues associated with BDA in healthcare.	The impact of persuasion requirement is discussed.
14.	Muhammad Umer Sarwar, Muhammad Kashif Hanif, Ramzan Talib, Awais Mobeen, and Muhammad Aslam	Survey of big data analytics in healthcare sector.	Big data processing, analysis and prediction using machine learning approaches.	With the future of IT and social and techniques discussion of big data, modeling, and machine learning.
15.	Md Saiful Islam , Md Mahmudul Hasan, Xiaoyi Wang , Hayley D. Germack and Md Noor-E-Alam	Application and theoretical perspective of data mining and big data analytics in whole healthcare domain.	Healthcare sub-areas, data mining techniques, types of analytics, data, and data sources.	Theoretical discussion on problems and consolidation in healthcare delivery
16.	Karim Abouelmehdi, Abderrahim Beni-Hessane and Hayat Khaloufi	Security, privacy, performance issues in a healthcare system.	Security, privacy, performance issues in a healthcare system.	Security, privacy, performance issues in a healthcare system.
17.	Blagoj Ristevski and Ming Chen	Big Data concept and characteristics, health care data and some major issues of Big Data.	Big data analytics applications and its issues.	Theoretical discussion on problems and consolidation in healthcare delivery
18.	Senthilkumar SA, Bharatendara K Rai, Amruta A Meshram, Angappa Gunasekaran, Chandrakumarmangalam S	Literature review, BDA and data mining, Methodology, tools/platforms available.	Big data analytics applications and its issues.	Future scope, proposed methodology and techniques to enhance scope of BDA in healthcare.
19.	Gemson Andrew Ebenezer J and Durga S	Survey of big data analytics in healthcare sector.	Big data processing, analysis and prediction using machine learning approaches.	Provides scope for the future work in big data healthcare
20.	Nivedita Das, Leena Das, Siddharth Swarup Rautaray and Manjusha Pandey	Survey of big data analytics in healthcare sector.	Big data processing, analysis and prediction using BDA in healthcare.	Provides scope for the future work in big data healthcare.

### III. Architectural Framework and Methodology of BDA in Healthcare

Healthcare data is expected to grow exponentially in the coming years. Healthcare and biomedical sciences

have become data-intensive investigator generating and using large, complex, high dimensional and diverse domain specific datasets. Big data analytics is emerging as a prominent and promising field in medical sciences in enhancing the quality and standards by providing an insight into very large data

sets and hence improving outcomes and reducing cost.

The data generated in healthcare system comes from various sources that are internal (in-hospital devices) as well as external (in-home devices) which is often in multiple formats (text, .csv, flat files, images etc.) and stored at various location and from various types of applications. Data mining techniques can be used to extract useful information to gain insight for making better informed health related decisions.

Fig. 2 and Fig. 3 below show a general framework and methodology of big data and BDA in healthcare system.

After collection of data from various sources the data need to be pooled into various categories depending upon the type of characteristics it possess. After then this unstructured data needs to be converted into analysis-ready datasets, which include comprehensive workflows for of big data solutions. The transformed data is provided input to the next segment in which big data platform and tool selection is made. Finally in the last component the four commonly used applications of BDA in healthcare are shown.

The best possible platform for BDA in healthcare is Hadoop MapReduce framework which is capable of storing various formats of healthcare data types including medical images, genomic data, financial and claim data etc. and provides high scalability, reliability and availability than the traditional databases.

In fig. 3 the steps employed in methodology are addressed. In first step the problem is stated w.r.t the various trade-offs that will be faced while looking for the solution while implementing BDA. In next step based on concept statement various questions are addressed (importance and interests in problem). In

3<sup>rd</sup> step the problem is broken down into set of proposition. The data needed to find the solution is collected and processed, tool/platform selection, conceptual model employed and analytic technique to be used is made. In the last step the results obtained in 3<sup>rd</sup> step are used to gain insight for making better informed health related decisions.

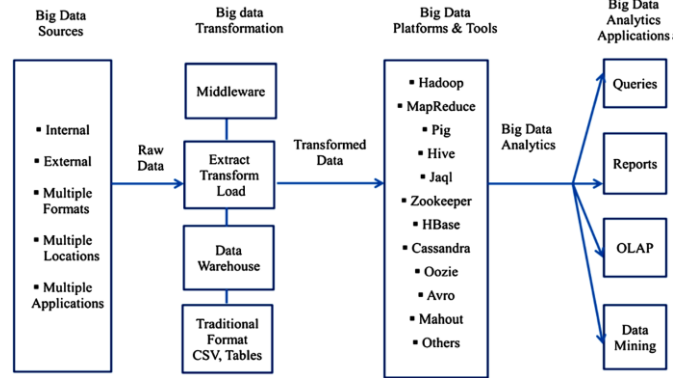


Fig. 2 An applied conceptual architecture of Big Data analytics (Raghupathi and Raghupathi, 2014) [9]

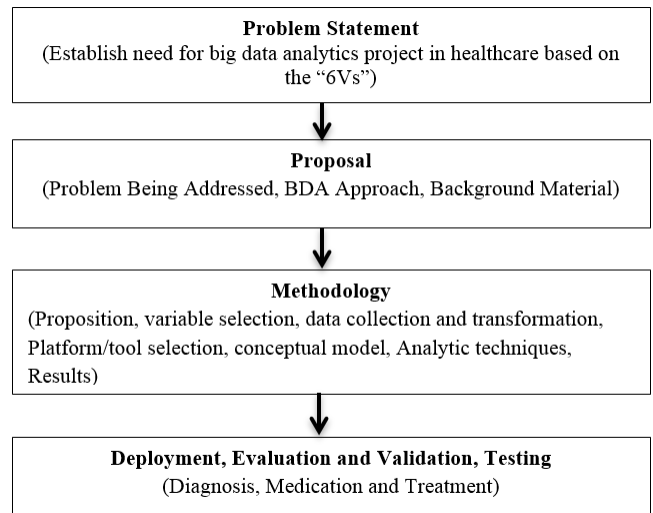


Fig. 3 Methodology of BDA in Healthcare

#### IV. Challenges of BDA in Medical Applications and Healthcare

The 6v's property possessed by the healthcare data is the minimum challenge for any BDA platform to resolve. The data generated has large volume, velocity and is highly heterogenic that creates problems in storage and processing. It is very difficult to aggregate and analyze unstructured data. Analyzing genomic and medical imaging data are

computationally intensive tasks which need highly complex algorithms. The platforms currently available are open source and hence the advantages and limitations also apply.

In healthcare diagnosis real time data is a key requirement and prerequisite in order to make the decisions more precise. The lag between data collection and processing needs to be addressed.

Another big issue is how medical data is widely spread at various locations. Integration of all such data sources need new infrastructure where all data providers can collaborate.

## V. CONCLUSION

Big data science has influenced every corner of life including health industry which has fully grasped its potential benefits. The implementation of BDA in healthcare has improved understanding of human life, ways of treatment, better diagnosis and précised decision making. It has reduced readmission, adverse events and treatment optimization for diseases affecting multiple organs. The emerging field of big data sciences is promising but is associated with many challenges that need to be addressed. These challenges include the aggregating and analyzing highly unstructured data, indexing and processing continuously streaming data, privacy, data leakage, information security and lack of infrastructure and unified standards, etc.

## VI. Acknowledgement

This paper is the outcome of study done during previous few years and will be helpful to students that are new to the field as this will provide them with nut shell information regarding this emerging field. It is also going to help with larger perspective to the people in the field of academia, research and healthcare to a greater extent.

## VII. REFERENCES

- [1]. IHTT: Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry;2013.<http://ihealthtran.com/wordpress/2013/03/iht%C2%B2-releases-big-data-research-reportdownload-today/>.
- [2]. Swarna Latha Garapati , Dr. Sridevi Garapati, Application of Big Data Analytics: An Innovation in Health Care International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 14, Number 1 (2018), pp. 15-27 © Research India Publications <http://www.ripublication.com>
- [3]. Burghard C: Big Data and Analytics Key to Accountable Care Success. IDC Health Insights; 2012.
- [4]. Dembosky A: "Data Prescription for Better Healthcare." Financial Times, December 12, 2012, p. 19; 2012. Available from: <http://www.ft.com/intl/cms/s/2/55cbca5a-4333-11e2-aa8f-00144feabdc0.html#axzz2W9cuwajK>.
- [5]. Feldman B, Martin EM, Skotnes T: "Big Data in Healthcare Hype and Hope." October 2012. Dr. Bonnie 360; 2012. <http://www.west-info.eu/files/big-data-inhealthcare.pdf>.
- [6]. Rabia Masood, S.A. Khan, M.N.A. Khan, "Plants Disease Segmentation using Image Processing", International Journal of Modern Education and Computer Science(IJMECS), Vol.8, No.1, pp.24-32, 2016.DOI: 10.5815/ijmeecs.2016.01.04.
- [7]. Michael J. Ward, Keith A. Marsolo, Craig M. Froehle, Applications of Business Analytics in Healthcare Bus Horiz. 2014; 57(5): 571–582. doi:10.1016/j.bushor.2014.06.003.
- [8]. Nivedita Das, Leena Das, siddharth Swarup Rautaray and Manjusha Pandey, Big data analytics for medical applications, I.J. Modern Education and Computer Science, 2018, 2, 35-42 Published Online February 2018 in MECS

- (<http://www.mecs-press.org/>) DOI: 10.5815/ijmecs.2018.02.04.
- [9]. Wullianallur Raghupathi and Viju Raghupathi, Big data analytics in healthcare: promise and potential, *Health Information Science and Systems* 2014, 2:3 <http://www.hissjournal.com/content/2/1/3>,
- [10]. Fabricio F. Costa, Big data in Biomedicine, *Drug Discovery Today* \_ Volume 19, Number 4 \_ April 2014, Elsevier.
- [11]. Ruchita Gujarathi, Fabricio F. Costa, The Impact of Online Networks and Big Data in Life Sciences, *Social Networking*, 2014, 3, 58-64 Published Online January 2014 (<http://www.scirp.org/journal/sn>)<http://dx.doi.org/10.4236/sn.2014.31007>.
- [12]. S. M. Riazul Islam, Daehan Kwak, Md.Humaun Kabir, Mahmud Hossain, And Kyung-Sup Kwak, The Internet of Things for Health Care: A Comprehensive Survey, 2169-3536 2015 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission.[http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information. VOLUME 3, 2015.
- [13]. Lidong Wang and Cheryl Ann Alexander , Big Data in Medical Applications and Health Care, *Current Research in Medicine* 2015, 6 (1): 1.8 DOI: 10.3844/amjsp.2015.1.8.
- [14]. J.Archenaa and E.A.Mary Anita, A Survey Of Big Data Analytics in Healthcare and Government, 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), Elsevier, *Procedia Computer Science* 50 ( 2015 ) 408 – 413, Available online at [www.sciencedirect.com](http://www.sciencedirect.com).
- [15]. Mohammad Ahmad Alkhatib, Amir Talaei-Khoei, Amir Hossein Ghapanchi, Analysis of Research in Healthcare Data Analytics, *Australasian Conference on Information Systems* Al Khatib et al 2015, Sydney, Healthcare Data Analytics.
- [16]. Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian, Big Data Analytics in Healthcare, *Hindawi Publishing Corporation BioMed Research International* Volume 2015, Article ID 370194, 16 pages <http://dx.doi.org/10.1155/2015/370194>.
- [17]. Liang Y and Kelemen A, Big Data Science and its Applications in Healthcare and Medical Research: Challenges and Opportunities, *Austin Biom and Biostat.* 2016; 3(1): 1030.
- [18]. Vishnu S Basuthkar, Chetana Srinivas, A Survey Of Cost-Effective Big Data In Healthcare Applications, *International Journal of Computer Applications (0975 – 8887) National Conference on Advances in Computing, Communication and Networking (ACCNNet – 2016)*.
- [19]. Jingwei Li, Wei Ding, Hsing Kenneth Cheng, Ping Chen, Dehai Di, Wei Huang, A Comprehensive Literature Review on Big Data in Healthcare, *Statistical Analysis and Review on Healthcare Big Data*, Twenty-second Americas Conference on Information Systems, San Diego, 2016.
- [20]. Siuly Siuly, Yanchun Zhang, Medical Big Data: Neurological Diseases Diagnosis Through Medical Data Analysis, *Data Sci. Eng.* (2016) 1(2):54–64, DOI 10.1007/s41019-016-0011-3, Received: 26 May 2016 / Accepted: 23 June 2016 / Published online: 27 July 2016 © The Author(s) 2016. This article is published with open access at [Springerlink.com](http://Springerlink.com).
- [21]. Mathieu Guillermin, Thierry Magnin, Big Data for Biomedical Research and Personalised Medicine: an Epistemological and Ethical Cross-Analysis *HSS*, vol. VI, no. 3(2017): 13-36.
- [22]. Muhammad Umer Sarwar, Muhammad Kashif Hanif, Ramzan Talib, Awais Mobeen, and Muhammad Aslam, A Survey of Big Data

Analytics in Healthcare, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.

- [23]. Md Saiful Islam , Md Mahmudul Hasan, Xiaoyi Wang , Hayley D. Germack and Md Noor-E-Alam, A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining, *Healthcare* 2018, 6, 54; doi:10.3390/healthcare6020054-[www.mdpi.com/journal/healthcare](http://www.mdpi.com/journal/healthcare).
- [24]. Karim Abouelmehdi, Abderrahim Beni-Hessane and Hayat Khaloufi, Big healthcare data: preserving security and privacy, Abouelmehdi et al. *J Big Data* (2018) 5:1 <https://doi.org/10.1186/s40537-017-0110-7>, Springer, *Journal of Big-Data*.
- [25]. Blagoj Ristevski and Ming Chen, Big Data Analytics in Medicine and Healthcare, DE GRUYTER, *Journal of Integrative Bioinformatics*. 2018; 20170030.
- [26]. Senthilkumar SA, Bharatendara K Rai, Amruta A Meshram, Angappa Gunasekaran, Chandrakumarmangalam S, Big Data in Healthcare Management: A Review of Literature, *American Journal of Theoretical and Applied Business* 2018; 4(2): 57-69 <http://www.sciencepublishinggroup.com/j/ajtab> , \doi: 10.11648/j.ajtab.20180402.14 ISSN: 2469-7834 (Print); ISSN: 2469-7842 (Online).
- [27]. Gemson Andrew Ebenezer J and Durga S, BIG DATA ANALYTICS IN HEALTHCARE: A SURVEY, VOL. 10, NO. 8, MAY 2015, ISSN 1819-6608, ARPN *Journal of Engineering and Applied Sciences* ©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.
- [28]. Nivedita Das, Leena Das, Siddharth Swarup Rautaray, Manjusha Pandey, " Big Data Analytics for Medical Applications", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.10, No.2, pp. 35-42, 2018.DOI: 10.5815/ijmeecs.2018.02.04.

**Cite this article as :**

Mubashir Hussain, Jatinder Manhas, "Applications of Big Data Analytics in the Field of Medical Diagnosis : A Review", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 5 Issue 3, pp. 625-638, May-June 2019. Journal URL : <http://ijsrcseit.com/CSEIT1953142>