Big Data: A Healthcare Revolution

Atmika Sharma

Bachelor of Technology Undergraduate Student, Dept. of Computer Science, Miranda House College, University of Delhi, New Delhi, India

---***---

Abstract - *Research identifies the recent surge in healthcare data as the key to improving health outcomes and reducing the ever increasing healthcare costs. Harnessing this data will be revolutionary for both payers and providers in the healthcare sector. Big data analytics will help payers make informed choices and offer incentives to providers for more effective care. This paper seeks to present big data as a concept and discuss the various advantages and challenges it brings along with itself. This paper also sheds light on Big Data Analytics implementations and platforms present in today's time that help deal with this flood of data.*

Key Words: Big Data, Healthcare, IT, Challenges, Benefits

1. INTRODUCTION

After decades of technological laggard, the field of medicine has begun to acclimatize to today's digital data age. New technologies make it possible to capture vast amounts of information about each individual patient over a large timescale. However, despite the advent of medical electronics, the data captured and gathered from these patients has remained vastly underutilized and thus wasted. Despite the inherent complexities of healthcare data, there is potential and benefit in developing and implementing big data solutions within this realm [9]. Data has the potential to transform healthcare - it is not just about processing huge datasets, but about making data 'hypothesis generating' rather than 'hypothesis driven' [11]. The power of Big Data is that it can be realized immediately and its effects can be felt from both micro and macro levels. Big Data analytics have the potential to help make sweeping improvements in organizations. Regulatory compliance is a top priority, but data analytics allow us to go beyond compliance and begin to fully understand the organization as a whole, including where to focus new investments without taking the emphasis off patient care quality [10].

1.1 What is Big Data?

Big data analytics (BDA) has emerged from two distinct concepts – big data and analytics. Together it represents a new information management approach that has been designed to derive previously untapped intelligence and insights from data to address many new and important questions [12]. Big Data refers to electronic datasets so large and complex that they are difficult (or impossible) to manage with traditional software and hardware [4]. A report delivered to the US Congress in August 2012 defines Big Data as "large volumes of high velocity, complex, and variable data that requires advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information" [1]. It can be classified into two types:

Structured data: For the most part, structured data refers to information with a high degree of organization, such that inclusion in a relational database is seamless and readily searchable by simple, straightforward search engine algorithms or other search operations [7].

Unstructured data: Unstructured data is a generic label for describing data that is not contained in a database or some other type of data structure. Unstructured data can be textual or non-textual. Textual unstructured data is generated in media like email messages, PowerPoint presentations, Word documents, collaboration software and instant messages. Non-textual unstructured data is generated in media like JPEG images, MP3 audio files and Flash video files [8].

1.2 The Shift in the Healthcare Sector

The healthcare model is undergoing an inversion. In the old model, facilities and other providers were given incentives to keep patients in treatment- that is, more inpatient days translated to more revenue. The trend with new models, including Accountable Care Organizations (ACO), is to provide incentives and compensate providers to keep patients healthy [3]. And, progressing across the analytics continuum towards being a data-driven organization involves a shift in the type of technologies and systems involved in working with the data, as well as an evolution in the types of business questions being asked [5].

Big data has many implications for patients, providers, researchers, payers and other healthcare constituents. It will impact how these players engage with the healthcare ecosystem, especially when external data, regionalization, globalization, mobility and social networking are involved. Patients especially will be an important element in keeping healthcare costs down and improving outcomes. Providing patients with accurate and up-to-date information and guidance rather than just data will help them make better decisions and better adhere to treatment programs [3]. In addition, if pharmaceutical developers could integrate population clinical data sets with genomics data, this development could facilitate those developers gaining approvals on more and better drug therapies more quickly than in the past and more importantly, expedite distribution to the right patients [1].

2. BIG DATA IN HEALTHCARE

The healthcare industry has managed to generate historically large amounts of data, driven mainly by record keeping, compliance and regulatory requirements and patient care [1]. In the global healthcare sector, there are three major types of digital data: clinical records, health research records, and business/organization operations records [4]. Rising rates of chronic diseases, aging populations, changing consumer expectations in terms of how they wish to receive care and increasing access to social media and mobile technologies are transforming the way healthcare is obtained and delivered. Payment models and healthcare systems in many countries are also evolving from a fee-for-service approach to an outcomes-based or accountable-care approach that requires access to more accurate data to document and track results. The healthcare industry has recently begun to turn to data and analytics in ways that are similar to other industries that rely on digital information to improve service with less expenditure [5].

Various techniques allow us to find correlations, patterns and structures in overwhelming volumes of data, giving them value. They reinforce the fact that data does not possess inherent value in the absence of a means to make sense of it. They are meaningless until analyzed for significance, visualized within a context or compared to other data. This means that the value of a dataset will vary according to its context [13].

While most data is stored in hard copy form, due to the overflow of data, the current trend is towards rapid digitization of these large amounts of data. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of 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 [1]. Healthcare organizations can leverage data to:

- gain better understanding of actual costs and outcomes
- identify areas where they have the most gains from improvements
- rigorously track performance over time [5]

By definition, big data in healthcare refers to electronic healthcare data sets so large and complex that they are difficult to manage with traditional software and/or hardware; nor can they be managed with traditional or common data management tools and methods [1]. The big data that healthcare organizations need to collect and analyze may come from hospitals, ambulatory care facilities, wellness centers, referral networks, labs and imaging centers, research and other non-traditional data sources [5]. The totality of data related to patient healthcare and well-being make up big data in the healthcare industry. It includes clinical data from CPOE and clinical decision support systems (physicians written notes and prescriptions, medical imaging, laboratory, pharmacy, insurance and other administrative data); patient data in electronic patient records (EPRs); machine generated/sensor data, such as monitoring vital signs; social media posts, including twitter feeds, blogs, status updates on Facebook and other platforms, and web pages; and less patient specific information, including emergency care data, news feeds and articles in medical journals [1]. Some other examples include case history, diet regime, list of medical health professionals in a healthcare organization, national health register data, medicine and surgical instruments expiry date identification based on RFID data [2].

Medical images, at one time, stored on film, are now created digitally and stored in Picture Archive and Communication Systems (PACS). They are the largest contributors to the expanding volume of Big Data in healthcare. And as medical imaging devices improve, in digital resolution capabilities, the data files dramatically increase in size. Information sensing wireless medical devices too have the potential for being major contributors to the flood of Big Data. With advances in sensor technology, there has been rapid growth in the number of wireless medical devices that continuously monitor patients and send reports to providers. Such devices have the potential to dramatically increase the amount of healthcare data, be it real-time or stored, captured for each patient. Also, health research- from drug development to biotechnology to public health- has long been data-intensive, and today its output is growing exponentially as well [4].

Thus, big data analytics applications in healthcare take advantage of the explosion in data to extract insights for making better informed decisions, and as a research category are referred to as, no surprise here, big data analytics in healthcare. By discovering associations and understanding patterns and trends within the data, big data analytics has the potential to improve care, save lives and lower costs [1].

2.1 What Big Data offers to Healthcare?

The potential for big data analytics in healthcare to lead to better outcomes exists across many scenarios, for example: by analyzing patient characteristics and the cost and outcomes of care to identify the most clinically and cost effective treatments and offer analysis and tools, thereby influencing provider behavior; applying advanced analytics to patient profiles (e.g. Segmentation and predictive modeling) to proactively identify individuals who would benefit from preventive changes or lifestyle changes; broad scale disease profiling to identify predictive events and support prevention initiatives; collecting and publishing data on medical procedures, thus assisting patients in determining the care protocols or regimens that offer the best value; identifying, predicting and minimizing fraud by implementing advanced analytic systems for fraud detection and checking accuracy and the consistency of claims; and, implementing much nearer to real-time, claim authorization; creating new revenue streams by aggregating and synthesizing patient clinical records and claims data sets to provide data and services to third parties, for example, licensing data to assist pharmaceutical companies in identifying patients for inclusion in clinical trials. Many payers are developing and deploying mobile apps that help patients manage their care, locate providers and improve their health overall. Via analytics, payers are able to monitor adherence to drug and treatment regimens and detect trends that lead to individual and population wellness benefits [1]. Some more areas in which enhanced data and analytics can yield great results include:

- Identifying treatments, programs and processes that are not delivering demonstrable benefits or are costing too much, and then determining how to replace them with more efficient and effective options

- Reducing readmissions by identifying environmental or lifestyle factors that increase risk or trigger adverse events, and adjusting treatment plans accordingly

- Building sustainability into a health system by bringing clinical, financial and operational data together to analyze resource utilization, productivity and throughput

- Attracting the best and brightest clinicians, who can help to build and maintain an organization's reputation by offering innovative health IT systems and mobile technologies that enable collaboration and easy, secure remote access to patient records [5]

All healthcare constituents- members, payers, providers, groups, researchers, governments, etc.- will be impacted by big data, which can predict how these players are likely to behave, encourage desirable behavior and minimize less desirable behavior [3].

2.2 The 4V's and their role in Healthcare

To realize these benefits, however, new approaches and technologies are required. Organizations need new analytics solutions and robust infrastructures that can handle big data and generate results rapidly [14]. Like big data in healthcare, the analytics associated with big data are described by three primary characteristics: volume, velocity and variety [1].

Volume: Overtime, health related data will be created and accumulated continuously, resulting in an incredible volume of data. The already daunting volume of existing healthcare data includes personal medical records, radiology images, clinical trial data, FDA submissions, human genetics and population data genomic sequences etc. Newer forms of big data, such as 3D imaging, genomics and biometric sensor reading, are also fuelling this exponential growth [1]. The benefit gained from the ability to process large amounts of information is the main attraction of big data analytics. Many organizations already have large amounts of archived data, perhaps in the form of logs, but not the capacity to process it [6].

Velocity: Data is accumulated in real time and at a rapid pace, or velocity. Velocity of mounting data increases with data that represents regular monitoring such as multiple daily diabetic glucose measurements, blood pressure readings, EKGs etc. [1]. While traditional data warehouse analytics tend to be based on periodic- daily, weekly or monthly- loads and updates of data, big data is processed and analyzed in real-time. This is important in healthcare for areas such as clinical decision support, where access to up-to-date information is vital for correct and timely decision-making and elimination of errors. Without current data, automated decisions cannot be trusted [3]. Specialized companies such as financial traders have already turned systems that cope with fast moving data to their advantage [6].

Variety: The ability to perform real-time analytics against high volume data in motion and across all specialties would revolutionize healthcare. Therein lays variety [1]. A common theme in big data systems is that the source data is diverse, and doesn't fall into neat relational structures. It could be text from social networks, image data, or raw feeds directly from a sensor source. None of these things come ready for integration into an application [6].

Big Data is overwhelming not only because of its volume, but also because of the diversity of data types and the speed in which it must be managed. Volume, velocity, and variety—often referred to as the three V's of Big Data— capture the true meaning of Big Data [4].

Veracity or 'data assurance' in big data refers to the analytics and outcomes being error-free and credible. Veracity is a goal, not yet a reality [1]. In healthcare, veracity of data is of paramount importance, requiring careful data curation and standardization efforts but at the same time seeming to be in opposition to the enforcement of privacy rights [13].

2.3 Big Data Life Cycle

The Big Data Life Cycle involves 5 steps:

1. *Data Collection:* It involves the collection of data from various sources and storing it in HDFS (Hadoop Distributed File System). Data can be anything such as case history, medical images, social logs, sensor data etc.

2. *Data Cleaning:* It involves the process of verifying whether there is any junk data or any data that has missed values. Such data needs to be removed

3. *Data Classification:* It involves the filtering of data based on the structure of that data. For example medical big data consists of mostly unstructured data such as hand written physician notes. Structured, semi-structured and unstructured data should be classified in order to perform meaningful analysis

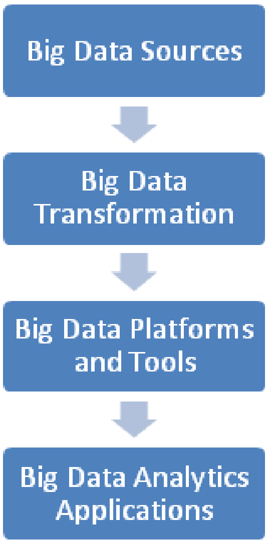
4. *Data Modeling:* It involves performing analysis on the classified data. For example an organization may require the list of malnourished children in a particular location. It has to classify the data based on the specific location, needs to trigger the health report of children, needs to identify the children whose family are under poverty line and then process that data

5. *Data Delivery:* It involves the generation of report based on the data modeling done. Taking the previous example into consideration, after the data is processed it will generate a report based on malnourished children in a particular location. This will help the organization take necessary measures to avoid any further complications [2].

At the all the stages of BDLC (Big Data Lifecycle) it requires data storage, data integrity and data access control [2].

2.4 Architectural Framework

The conceptual framework for a big data analytics project in healthcare is similar to that of a traditional health informatics or analytics project. They key difference is how the processing is executed. Because big data by definition is large, processing is broken down and executed across multiple nodes. In the second component, the data is in a raw state and needs to be processed or transformed at which point several options are available. A service oriented architectural approach combined with web services is one possibility. The data stays raw and services are used to call, retrieve and process the data. Another approach is data warehousing wherein data from various sources is aggregated and made ready for processing although the data is not available in real-time via the steps of extract, transform and load; data from diverse sources is cleansed and readied. In the next component several decisions are made regarding the data input approach, distributed design, tool selection and analytics models. Finally, the fours typical applications of big data analytics in healthcare come into the picture. These include queries, reports, OLAP and data mining. Visualization is an overarching theme across the four applications [1].



Google and Amazon implement MapReduce-based solutions to process huge datasets using a large number of computers-terabytes of data on thousands of computers. MapReduce algorithms take large problems and divide them into a set of discrete tasks that can then be distributed to a large number of computers for processing and the results can be combined into a problem solution [3]. The most significant platform for big data analytics is the open-source distributed data processing platform Hadoop (Apache platform). It belongs to the class "No SQL" technologies and has the potential to process extremely large amounts of data mainly by allocating partitioned data sets to numerous servers (nodes), each of which solves different parts of the larger problem and then integrates them for the final result [1].

2.5 Advantages

Potential benefits include detecting diseases at earlier stages when they can be treated more easily and effectively; managing specific individual and population health and detecting healthcare fraud more quickly and effectively. Certain developments or outcomes may be predicted and/or estimated based on vast amounts of historical data, such as length of stay; patients who will choose elective surgery; patients who likely will not benefit from surgery; patients at risk for medical complications; patients at risk for sepsis, MRSA, C. difficile or other hospital acquired illness; illness/disease progression; and possible co-morbid conditions [1].

McKinsey believes big data could help reduce waste and inefficiency in the following areas:

Clinical operations*:* Comparative effectiveness research to determine more clinically relevant and cost effective ways to diagnose patients.

Research and development*:* 1) Predictive modeling to lower attrition and produce a leaner, faster, more targeted R&D pipeline in drugs and devices; 2) statistical tools and algorithms to improve clinical trial design and patient recruitment to better match treatments to individual patients, thus reducing trial failures and speeding new treatments to market; and 3) analyzing clinical trials and patient records to identify follow-on indications and discover adverse effects before products reach the market [1]

Public Health*:* 1) Analyzing disease patterns and tracking disease outbreaks and transmission to improve public health surveillance and speed response [1]. This can be identified by analyzing the social logs of the patients suffering from a disease in a particular geo-location [2]; 2) faster development of more accurately targeted vaccines; and 3) turning large amounts of data into actionable information that can be used to identify needs, provide services, and predict and prevent crises, especially for the benefit of populations [1]

In addition, big data analytics can also contribute to:

Evidence-based medicine: combine and analyze a variety of structured and unstructured data-EMRs, financial and operational data, clinical data [1], minimizing drug doses to avoid side effect and providing efficient medicine based on genetic makeups. Performing an analysis of the data generated by the patients who have already suffered from the similar symptoms helps doctors provide effective medication to new patients. This helps in reducing readmission rates thereby reducing cost for the patients [2]

Genomic analytics*:* Execute gene sequencing more efficiently and cost effectively and make genomic analysis a part of the regular medical care decision process and the growing patient medical record [1]

Pre-adjunction fraud analysis*:* Rapidly analyze large numbers of claim requests to reduce fraud, waste and abuse [1]

Device/remote monitoring*:* New technology and methods of critical care and continuous care can change delivery and methods of delivery and ultimately reduce the cost of care through remote monitors feeding data in near real-time to electronic medical record databases [11]. Capture and analyze in real-time large volumes of fast-moving data from in-hospital and in-home devices, for safety monitoring and adverse event prediction [1]. In general, the use of data from remote monitoring systems can reduce patient in-hospital bed days, cut emergency department visits, and improve the focus of nursing [11].

Patient profile analytics: Apply advanced analytics to patient profiles to identify individuals who would benefit from proactive care or lifestyle changes [1].

Bringing together structured and unstructured data supports more insightful analysis that enables personalized and evidence-based medicine, more efficient processes and incentives that can improve patient behavior. Gaining access and applying clinical and advanced analytics to this valuable data enables organizations to improve insight into risk, outcomes, resources, referrals, performance and readmissions, and to take prescriptive action [5].

To help users quickly make sense of high-velocity data, organizations need analytics solutions that incorporate visualization capabilities. Visualization can help users identify patterns, make correlations among disparate data types, and explore data much more quickly and easily than when viewing a spreadsheet. Visual analytics tools that use familiar drag-and-drop functionality and enable access through mobile devices help bring analytics to a wide range of business users [14].

In addition to integrating data within the healthcare system, there are many potential benefits of integrating data from outside of the healthcare system. As an example, external data about healthy people holds immense potential value for research and the future delivery of healthcare. Typical healthcare data includes only people visiting doctors and hospitals, which biases that data toward people seeking treatment. Adding anonymous data from large numbers of healthy people could help establish baselines, draw correlations and help with understanding the nature of illnesses. More data, effectively used, leads to better information and decisions, and more meaningful efforts. External data will come from different medical systems in various regions and countries. Effectively working across these disparate data repositories can help identify local knowledge and best practices and leverage them regionally and globally. Aggregating data regionally and globally also provides healthcare researchers with larger populations for clinical studies, trending and disease monitoring for epidemics, as well as early detection and the potential for improved results [3].

2.6 Challenges

Volume is perhaps the most immediate challenge of big data, as it requires scalable storage and support for complex, distributed queries across multiple data sources [3]. Structured data is data that can be easily stored, queried, recalled, analyzed, and manipulated by machine. Historically in healthcare, structured and semi-structured data includes instrument readings and data generated by the on-going conversion of paper records to electronic health and medical records. Historically the point of generated unstructured data has been: office medical records, handwritten nurse and doctor notes, hospital admission and discharge records, paper prescriptions, radiograph films, MRI, CT and other images. Already, new data streams- structured and unstructured- are cascading into the healthcare realm from fitness devices, genetics and genomics, social media, research and other sources [1]. While standard techniques and technologies exist to deal with large volumes of structured data, it becomes a significant challenge to analyze and process a large amount of highly variable data and turn it into actionable information [3]. Algorithms capable of handling smaller datasets are insufficient for managing the volume of healthcare data- in the range of petabytes- now being produced. Healthcare applications too need more efficient ways to combine and convert varieties of data including automating conversion from structured to unstructured data [1].

Data timeliness is a challenge in various healthcare settings, such as clinical decision support, whether for making decisions or providing information that guides decisions. Big data can make decision support simpler, faster and ultimately more accurate because decisions are based on higher volumes of data that are more current and relevant. In some cases, there is a very limited window for clinical decision support- significantly smaller than the time it takes to run a report or analytic query [3].

Improving coordination of care, avoiding errors and reducing costs depend on high quality data, as do advances in drug safety and efficacy, diagnostic accuracy and more precise targeting of disease processes by treatments. But increased variety and high velocity hinder the ability to cleanse data before analyzing it and making decisions, magnifying the issue of data "trust". Quality issues are of acute concern in healthcare mainly for two reasons: life or death decisions depend on having accurate information, and the quality of healthcare data, especially unstructured data, is highly variable and all too often incorrect. The need to field code data at the point of care for electronics handling is a major barrier to acceptance of EMRs by physicians and nurses who lose the natural language ease of entry and understanding that hand-written notes provide [1].

Collecting, integrating and analyzing data can be a complex task because the data resides in many internal and external locations and its level of quality may be unknown [5]. As progress is made toward initiatives such as electronic health records (EHR), more and more external data will become available, and this will become an integration challenge. External sources include the National Health Information Network (NHIN), health information exchanges (HIE), health information organizations (HIO) and regional health information organizations (RHIO). As sources and volume of information increase, so will expectations [3].

There are also multiple interoperability challenges. Big data solution architectures have to be flexible enough to cope with not only the additional sources but also the evolution of schemas and structures used for transporting and storing data. To ensure analytics are meaningful, accurate and suitable, metadata and semantic layers are needed that accurately define the data and provide business context and guidance, including appropriate and inappropriate uses of the data [3].

Social networking data can act as a potentially important source of big data but poses challenges such as volume, lack of structure and velocity, as well as new challenges around integration and accuracy [3].

Healthcare is notoriously slow to redefine and redesign processes and tends to be a laggard in adopting technology that impacts the healthcare system, outside of some specific areas such as care delivery and research [3]. Virtualization and cloud computing are two cost-effective means of reducing complexity, but healthcare organizations have been slow to put them to good use, much slower than other industries, such as finance and retail largely because of traditional processes and concerns over the privacy of personal health information. This slow shift is because Big Data, in its volume, velocity, and variety, creates two major problems for healthcare organizations: potentially overwhelming complexity and increased IT expenditures [4]. In addition, keeping up with fast-changing and advancing Big Data technologies is a daunting task when time and resources are already constrained and competition for industry and technical talent remains fierce [10].

The cloud makes exposing and sharing big data easy and relatively inexpensive. However, significant security and privacy concerns exist, including the Health Insurance Portability and Accountability Act (HIPAA). A credentialing process could facilitate and automate this access, but there are complexities and challenges. Since providers, patients and other interested parties such as researchers need secure access; data access should be controlled by group, role and function. The biggest obstacle to effective use of big data is the nature of healthcare information. Payers, providers, research centers and other constituents all have their own silos of data. Even integration is difficult because of concerns about privacy and propriety, the complex and fragmented nature of the data, as well as the different schemas and standards underlying the data and lack of metadata within each silo. The security of the data once it leaves the cloud needs to be assured [3].

It may be that the more pragmatic approach is to legislate against misuse of data as opposed to prescribing allowable uses. The debate on data protection and open access should come to an ethically-based consensus agreement, allowing for the views of minorities to be respected, if the right of citizens to appropriate data protection is to be adequately balanced against their right to further improved healthcare based on patient data-facilitated clinical research. This balance is crucial if legislators wish to avoid overprotection of the rights of a minority becoming detrimental to the delivery of effective healthcare for the majority [13].

In addition to data that is readily available, such as demographics and medical history, another data source is information that patients divulge about themselves. When combined with outcomes, high-quality data provided by patients can become a valuable source of information for researchers and others looking to reduce costs, boost outcomes and improve treatment. Several challenges exist with self-reported data:

- Accuracy: People tend to understate their weight and the degree to which they engage in negative behaviors such as smoking; meanwhile, they tend to overstate positive behaviors, such as exercise. These inaccuracies can be accounted for by adjusting these biases and- through big data processing- improve accuracy time

- Privacy concerns: People are generally reluctant to divulge information about themselves because of privacy and other concerns. Creative ways need to be found to encourage and provide incentives for them to do so without adversely impacting data quality. Effective mechanisms and assurances must be put into place to ensure the privacy of the data that patients submit, including de-identification prior to external access

- Consistency: Standards need to be defined and implemented to promote consistency in self-reported data across the healthcare system to eliminate local discrepancies and increase the usefulness of data
- Facility: Mechanisms based on e-health and m-health- such as mobility and social networking-need to be creatively employed to ease members' ability to self-report. Providing access to some de-identified data can simultaneously improve levels of self-reporting as a community develops among members [3].

In the end, it is not just about images and large data files, but the increasing data sets being developed from many sources across the healthcare spectrum: medical records, data from remote monitoring and mHealth sources, genetic information, etc. As more and more data is being stored by healthcare organizations, more is being ignored as well. The challenge is not just in storage and access, but in making this data usable. Healthcare organizations are overwhelmed, and are not able to incorporate all this data for business intelligence and analytics [4]. The problem in healthcare isn't the lack of data but the lack of information that can be used to support decision-making, planning and strategy [3].

3. CONCLUSION

Big data analytics which leverages legions of disparate, structured, and unstructured data sources is going to play a vital role in how healthcare is practiced in the future. One can already see a spectrum of analytics being utilized, aiding in the decision making and performance of healthcare personnel and patients [9]. Big data analytics has the potential to revolutionize the healthcare industry by benefits for both patients and providers. It not only helps patients make informed decisions but also enables providers to optimize their business outputs. The surge of data has caused this shift in the healthcare sector and not only does it concern the healthcare industry but also brings growth potential for the IT sector. Even though several challenges and hurdles need be overcome, there is scope for massive advancement in healthcare and also for the tools and platforms of analytics.

ACKNOWLEDGEMENT

I would like to thank my advisor for this paper Mrs. Isha Pathak, Assistant Professor, Computer Science Department, Miranda House College for her valuable inputs and guidance throughout this paper.

REFERENCES

[1]: Wullianallur Raghupathi and Viju Raghupathi: *Big data analytics in healthcare: promise and potential. Health Information Science and Systems,* 2014 2:3 [1]

[2]: J.Archenaa and E.A.Mary Anita: *A Survey Of Big Data Analytics in Healthcare and Government,* ISBCC, 2015

[3]: Cognizant, Bill Hamilton: *Big Data is the Future of Healthcare*, White Paper, 2012

[4]: Frost and Sullivan: *Drowning in Big Data? Reducing Information Technology Complexities and Costs For Healthcare Organizations, A Frost and Sullivan White Paper,* http://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf

[5]: IBM: *Data-driven healthcare organizations use big data analytics for big gains,* http://www-03.ibm.com/industries/ca/en/healthcare/documents/Data_driven_healthcare_organizations_use_big_data_analytics_for_big_gains.pdf, 2013

[6]: Edd Dumbill: *What is big data? An introduction to the big data landscape, https://www.oreilly.com/ideas/what-is-big-data, 2012*

[7]: *Structured vs. Unstructured Data, White Paper, https://brightplanet.com/2012/06/structured-vs-unstructured-data/, 2012*

[8]: *http://searchbusinessanalytics.techtarget.com/definition/unstructured-data*

[9]: Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian: *Big Data Analytics in Healthcare,* 2015

[10]: Hewlett-Packard Development Company, *L.P: Big Data and Healthcare*, White Paper, http://h20195.www2.hp.com/v2/getpdf.aspx/4AA5-2847ENW.pdf?ver=1.0, 2014

[11]: Viswanathan Ganapathy: *New Horizons, New Hopes, Big Data in Healthcare*, White Paper, http://www.tcs.com/SiteCollectionDocuments/White%20 Papers/White-paper-Healthcare-Big-Data-1013-1.pdf

[12]: Canada Health Infoway: *Big Data Analytics in Health*, White Paper, https://www.infoway-inforoute.ca/en/component/edocman/1246-big-data-analytics-in-health-white-paper-full-report/view-document

[13]: Edwin Morley-Fletcher: *Big Data Healthcare: An overview of the challenges in data intensive healthcare*,

[14]: Intel*: Capitalizing on the Power of Big Data for Healthcare*, http://support.sas.com/resources/papers/proceedings14 /2522-2014.pdf

BIOGRAPHIES

Atmika Sharma is an undergraduate student at the Department of Computer Science, Miranda House College, University of Delhi. She is currently in the 4th year of her undergraduate studies.