© September 2015 | IJIRT | Volume 2 Issue 4 | ISSN: 2349-6002

Big Data Analytics in Health Care

S. G. Nandhini[1], V. Lavanya[2], K.Vasantha Kokilam[3]
[1]13mss032, [2]13mss025, III. M.Sc (software systems), SRI KRISHNA ARTS AND SCIENCE COLLEGE,
[3]Assistant Professor, SRI KRISHNA ARTS AND SCIENCE COLLEGE,
An autonomous college Affiliated to Bharathiar University
Accredited by NAAC with 'A' Grade, An ISO 9001:2008 Certified Institution
Coimbatore 641 008, Tamilnadu, INDIA.

***Abstract*- This paper provides a brief idea about additional value from health information used in health care centers using a new information management approach called as big data analytics .Including big data analytics in health sector provides stakeholders, the new insights that have the capacity to advance personalized care improve patient outcomes and avoid unnecessary costs. This paper describes big data analytics and its characteristics, advantages and challenges in health care.**

***Index Terms*- Big data, Analytics, Hadoop, Healthcare, Framework, Methodology, challenges; future applications**

I. INTRODUCTION

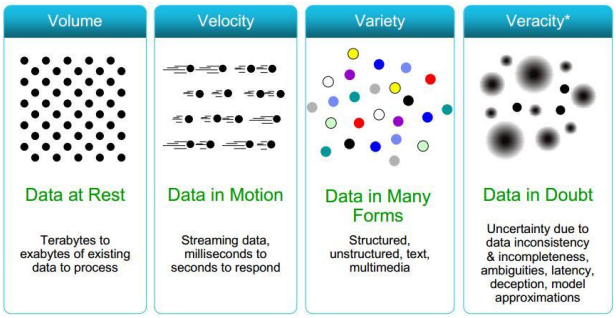
Big data is "a term that includes large volumes of complex, high velocity, and variable data that wants advanced techniques and technologies to enable the capture, memory, distribution, management
and analysis of the information".
The factors influencing big data are

• **Volume**: It is the amount of data produced by organizations or individuals. All organizations are searching for ways to handle the ever increasing data volume that's being created every day.

• **Velocity**: It is the speed and frequency at which data is produced, captured and shared. Consumers as well as businesses now generate lots of data and in most shorter cycles, from hours, minutes, seconds down to milliseconds.

• **Variety**: It is the proliferation of new data types including those from social, mobile and machine resources. New types include content, metrics, mobile, physical data points, process, location or geo-spatial, hardware data points, machine data, radio frequency identification (RFID), search, and web. It also includes unstructured data.

Veracity: It is defined as the accuracy of data. Incorrect data can cause a lot of problems for organizations. Hence, organizations need to ensure that the data is correct as well as the analyses performed on the data are correct. In automated decision-making, where no human is involved we need to be sure that both the data and the analyses are correct.



II. BIG DATA IN HEALTH CARE

Big data in healthcare can come from internal (e.g., electronic health records, clinical decision support systems, CPOE, etc.) and external sources (government sources, pharmacies, insurance companies etc.), often in multiple formats (flat files, relational tables, etc.) and residing at many locations (geographic as well as in different healthcare providers' sites) in numerous legacy and other applications (transaction processing.).

Resources and data types include:

1. **Web and social media data**: Clickstream and interaction data from social media such as Facebook, Twitter, and blogs. It can also

include medical health plan websites, Smartphone applications, etc.

2. **Machine-to-machine data**: Readings from meters, sensors, and other devices.

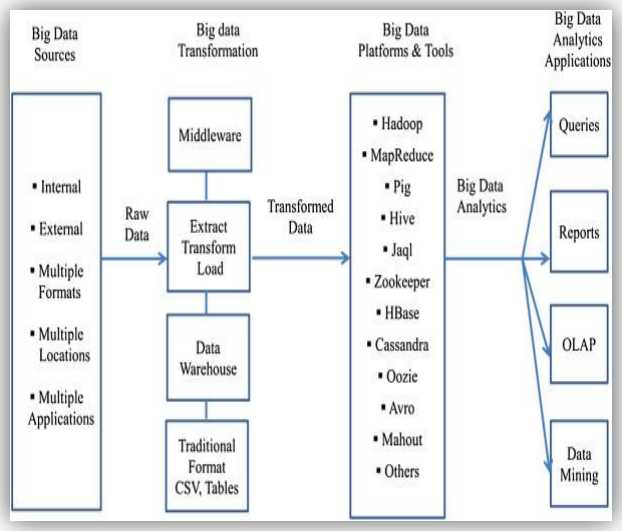
3. **Big transaction data**: Health care claims and other billing records increasingly available in semi-structured and unstructured formats.

4. **Biometric data:** Fingerprints, genetics, retinal scans, and similar to this types of data. This also includes X-rays and other medical images.

5. **Human-generated data**: Unstructured and semi-structured data such as electronic medical records (EMRs), physician's notes, email, and paper documents.

In recent years, BDA has become increasingly apparent that multiple streams of data like these can be leveraged with powerful new collection, aggregation, and analytics technologies and techniques to improve the delivery of health care at the level of individual patients as well as at the levels of disease- and condition-specific populations.

A conceptual architecture of big data analytics.



III. PLATFORMS & TOOLS FOR BIG DATA ANALYTICS IN HEALTHCARE

Numerous vendors including AWS, Cloudera, and MapR Technologies distribute open-source Hadoop platforms. Many proprietary options are also available like IBM's BigInsights. Since many of these platforms are cloud versions,these are widely available. Cassandra, HBase, and MongoDB, are widely used database component. The development costs may be lower since these tools are open source and free of charge, the drawback is the lack of technical support and minimal security. In the healthcare sector, these are notable drawbacks, and therefore the trade-offs must be represented. Also, these platforms/tools require a vast deal of programming skills the typical end-user in healthcare may not process. On considering the only recent emergence of big data analytics in healthcare, governance issues including ownership, security, and standards have yet to be represented. The next section offers an applied big data analytics in healthcare methodology to develop and implement a big data project for healthcare providers.

IV. METHODOLOGY

While many different methodologies are being developed in the rapidly emerging discipline, here the one that is practical and hands-on was discussed.

In *Step 1*, the big data analytics in healthcare team developed a 'concept statement'. This is a first step at establishing the need for a project. The next step is the description of the project's significance. The healthcare organization would note that there are trade-offs in terms of alternative options, scalability, cost, etc. After the concept statement is approved, the team will proceed to *Step 2*,

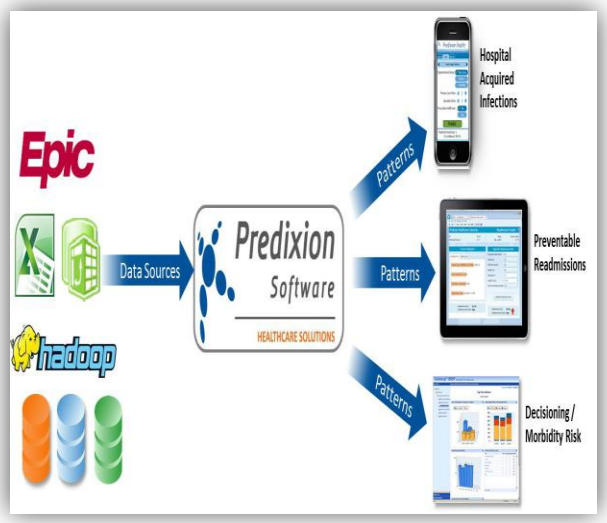
Step 2 the project development stage. Here, all details are filled in. On the basis of concept statement, several questions are queried: What problem is being addressed? Why it is interesting and important to the healthcare provider? What is the case for a 'big data' analytics to approach? (Because the cost and complexity of big data analytics are significantly higher compared to traditional analytics approaches, it is important to know their use). The project team also should provide information on the problem domain as well as earlier projects and research which was done in this domain.

Predixion Software: It uses cloud-based predictive analytic software to explain patterns in hospital datasets to reduce readmissions and prevent hospital-acquired conditions.

- Pulls data from a variety of resources, using machine learning, data mining, and mathematical algorithms to power predictions.
- Uses predictive analytics algorithm to risk score patients based on admission and throughout their hospital stay, to identify the

risk of readmission before they leave the hospital, with 86% accuracy.

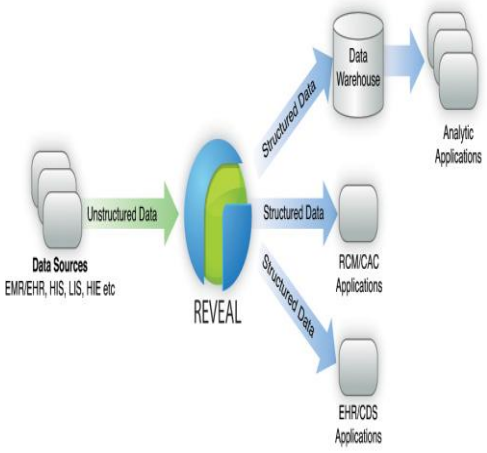
- Present project was applying analytics to prevent deaths and MRSA infections in the hospital setting.
- Current project helps to use predictive analytics as an aid for prevention of chronic disease – e.g., diabetes.



"

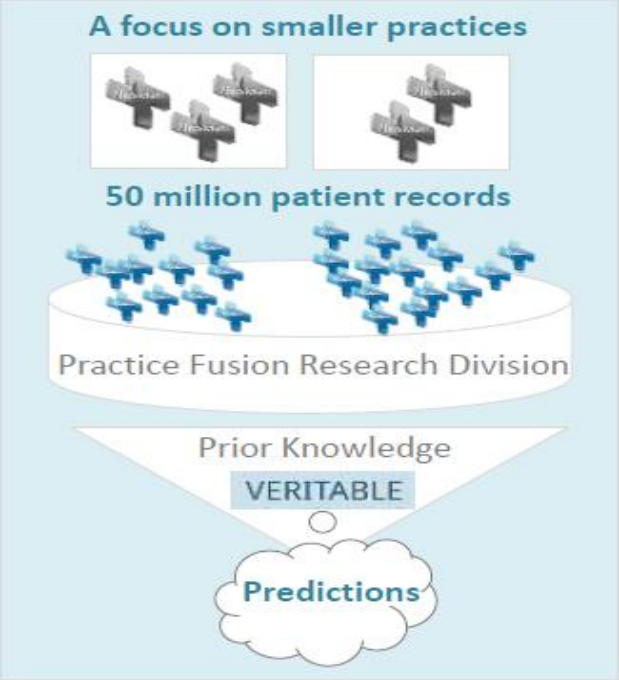
Health Fidelity: It is using NLP to convert unstructured data (e.g., narrative medical records) into structured data which is suitable for computer management, to represent needs in revenue cycle management, analytics and compliance.

- Health Fidelity's NLP technology converts specialized and complex medical cases and breaks out critical context to make it available in real time. It runs many data streams in multiple formats such as note types, jargon, domains, linguistic forms and grammatical relationships.
- This unique and complex process was first supported by the National Science Foundation and National Institute of Health.
- Clients include healthcare IT vendors that serve provider networks, medical practices, and large healthcare organizations.
- Early use cases focus on revenue cycle management, compliance, and analytics focused on quality improvement and cost reduction.



Practice Fusion: It is a free, cloud-based EMR platform for clinical practices that also collects data across multiple sources to improve public health analysis and medical research.

- Provides free EMR (Electronic Medical Record) platform for smaller activities like e-prescribing, labs, Meaningful Use, scheduling and charting.
- Analyzes data from the EMR system to monitor health on basis of detection of disease, and provide research-based insight (never raw data) to patient.
- It includes Collaborators such as Prior Knowledge and Stanford Center for Biomedical Informatics Research
- Present research is done on a disease called cancer, later on moves to heart disease.



V. OPPORTUNITIES OF BIG DATA ANALITICS IN HEALTH CARE

There is potential to layer BDA-type applications, in a privacy-protect manner, on basis of the foundational health IT architecture to achieve value that may not be found otherwise. The following are some of the innovative solutions and ideas.

• Clinical decision support: BDA technology that shift through larger amounts of data, understand, classify and learn from it, and then recommend another treatments to patients on the point of care.
• Personalized care: Predictive data mining or analytic solutions that can leverage personalized care in real time to highlight better practice treatment to patients (e.g., genomic DNA sequence for cancer care). These solutions may offer early detection and diagnosis of a disease.
• Public and population health: Big Data Analytics solutions focus on web-based and social media data to predict flu outbreaks based on patient's search, social content and question activity. Big Data Analytics solutions will also support doctors and epidemiologists performing analyses on patient strength and care to help identify disease status.
• Clinical operations: Big Data Analytics can support wait-time management, where it will focus large amounts of unstructured data for model various scenarios to identify events that may affect wait times before they actually occur.
• Financial and administrative: BDA supports decision makers by separating and analyzing data related to key performance indicators.

VI. BENEFITS TO HEALTH CARE

By effectively using big data, health care organizations ranging from single-physician offices to large hospital networks, organizations stand to realize significant benefits.

1. Health Care Quality and Efficiency

As per record on 2010, national health expenditures says that 17.9% of gross domestic product, up from 13.8% in 2000. But on other hand, the chronic diseases like diabetes are increasing and consumes a larger percentage of health care resources. Electronic health records (EHRs), paired with new analytics tools, and provides mining information for the most effective outcomes across maximum count. By using de-identified information, researchers provide assessments on the basis of true quality of care.

2. Earlier Disease Detection

Electronic sensors are employed in larger amount to monitor key biochemical markers, with the help of real-time analysis will work as the data streams from individual patients to compliant analysis systems. Analytics like these will alert specific patients and their providers to potentially adverse events, like the early development of infection, side effects to patients and allergic reactions.

3. Fraud Detection

Big data analytics is widely expected to transform medical claims payment systems, results in reduced submission of improper or fraudulent claims. Big data also used to improve population health management, the identification and measurement of most accurate quality measures, the management of captivated patients, and treatment protocols for chronic conditions such as diabetes and congestive heart failure (CHF). Through the

use of data mining technique, big data can be used to identify patients at risk for various diseases, adverse drug events, improve selection of candidates for patient centered interventions and identify costly procedures, waste and delays. The health care industry will achieve $300 billion in year basis by effectively influencing big data.

VII. BIG DATA CHALLENGES IN HEALTH CARE.

• Influencing the patient or data correlations in longitudinal records.
• Understanding unstructured medical notes in the right content.
• Efficiently handling large volumes of clinical imaging data and extracting potentially useful information and biomarkers.
• Analyzing genomic data is a computationally concentrated task and combining with standard medical data adds additional layers of complexity.
• Collecting patient's behavioral data with the help of several sensors; their various social interactions.

VIII. CONCLUSION AND FUTURE WORK

Big data analytics in healthcare will become a promising field for providing insight from very big data sets and improving outcomes on reducing costs. Its potential is so great; however there also remain challenges to overcome.

Big data analytics has the potential to transform the way healthcare providers use technologies to gain insight from their clinical and other data repositories and make informed decisions. In the future there will be the rapid, widespread implementation and use of big data analytics across the healthcare organization. To that end, the several challenges must be represented. As big data analytics becomes more important, issues such as providing privacy and security, establishing standards and governance, and continuous improvement of the tools and technologies will gain attention. Big data analytics and applications in healthcare are at a nascent stage of development, but future advances in platforms and tools can be their maturing process.

REFERENCES

- Raghupathi W: Data Mining in Health Care. In *Healthcare Informatics: Improving Efficiency and Productivity*. Edited by Kudyba S. Taylor & Francis; 2010:211-223.
- Burghard C: *Big Data and Analytics Key to Accountable Care Success*. IDC Health Insights.
- Dembosky A: *"Data Prescription for Better Healthcare." Financial Times*.
- Bian J, Topaloglu U, Yu F, Yu F: *Towards Large-scale Twitter Mining for Drug-related Adverse Events*. Maui, Hawaii: SHB; 2012.
- Raghupathi W, Raghupathi V: *AnOverview of Health Analytics*.