ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 22, December 2017

Potential of Big Data Analytics in Health Care- A Survey

K.Akila[1], J. Priya [2], Dr.D. Prabha [3]
Assistant Professor, IT, Meenakshi Sundararajan Engineering College, Chennai, India[1]
Assistant Professor, IT, Meenakshi Sundararajan Engineering College, Chennai, India[2]
Associate Professor, CSE, Sri Krishna College Of Engineering and Technology, Coimbatore, India[3]

Abstract**:** In an era of several torrents of data, big data plays a pivotal role in effectively processing the data. Big data analytics helps organizations harness their data and use it to identify new opportunities in various fields like meteorology, genomics, complex physics simulations, biological and environmental research, finance and business to healthcare. The healthcare sector has generated large volume of data with incredible velocity and diverse variety. Big Data in healthcare instigates from the large healthcare records that are very complex to manage with the traditional data analytics. Incorporation of Big Data in healthcare is an upsurge in the field of medical sciences. This paper targets the utilization of Big Data in various sectors of healthcare.

Keywords: ***Big data; Big data analytics; volume; velocity; variety; Health care***

I. INTRODUCTION

Data Analytics in health care also foresees outburst of epidemics, avoid diseases and can also increase and improve the quality of life in Big Data is a collection of enormous and complex data sets which are challenging to process using conventional data processing applications. The challenges comprise extracting, storing, searching, sharing & analysing data patterns. It is not just about size but also finds insights from noisy, heterogeneous complex and voluminous data. It answers questions that were previously unanswered. Big Data is mainly based on three main properties volume ,velocity and variety .In this digital world, we can acquire large volumes of complex unstructured data from multiple sources such as digital pictures, posts to social media, intelligent sensors, and transaction and healthcare records through interactive queries. The rate at which the data flows from different sources is the velocity. Variety refers to the diverse data types which includes audio, video images, text etc. Big data describes any large volume of structured, unstructured data and semi structured data that has the possibility to be mined for information. Existing conventional methods of data mining can hardly give worthy information. There is need for modifying existing machine learning methods for better data extraction and decision making

This paper mainly introduces the characteristics of Big Data, and it's prospective in the arena of health care. Big data has become the buzzword in today's research communities. The field of big data analytics has started to play a vital role in the evolution of healthcare practices and research. The major stakeholders of the healthcare sectors such as payers, providers, healthcare IT, and pharmaceutical companies are under increased pressure to revamp the quality of patient care at an affordable cost. Today health care industry generates large amount of data, more over government also generates petabytes of data every day. These data generated can be put to a good use when intervened with effective technology like big data. Big data analytics helps in discovering primal decisions by understanding the data patterns and its relationship between them with the help of machine learning algorithms.

The healthcare model is undergoing major changes. In the conventional model, health care providers considered that more inpatient days translated to more revenue. The trend with new models is to keep the patients' healthy and cure them as fast as possible. Patients are additionally an essential component in keeping human services costs down and enhancing results. Furnishing patients with correct and up-to-date information rather than just giving data will enable them to settle on better choices and cling to treatment programs.

II. CHALLENGES OF BIG DATA ANALYTICS IN HEALTH CARE

In the rapidly growing health care sector we get abundant data but the main challenge posed is the lack of information

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 22, December 2017

that can be used in planning, strategy and decision-making .For example a single inpatient stay generates large amount of records such as diagnoses report, medication, medical supplies, lab result and billing.

These need to be verified, processed and assimilated into a large data source to aid meaningful analysis and extract a pattern. When we start consolidating the records of all the patients and combine it into a framework the challenges in big data begin to emerge. This constitutes only a fraction of healthcare data landscape. Charted below are some of the difficulties of healthcare big data, including healthcare as a technology dawdler, fragmentation of data, security, standards and timeliness.

A. *Healthcare as a Technology dawdler*

The healthcare sector takes a long time to overhaul and redefine processes, which is a main hurdle to embrace technology that impacts healthcare system except for a few areas such as care delivery and research. Adding to the complexity healthcare includes vast range of legacy technology, which is difficult to handle.

B. *Fragmentation*

One of the major big data challenge in healthcare is that the data is fragmented and dispersed among various stakeholders, which includes payment gateway providers, labs, auxiliary and data vendors, organization and regulatory agencies. Big Data breaks the traditional model in which all data are loaded in a data ware house. Data federation emerges as a solution which collects metadata that is the original structure of the database rather than collecting all information from database .This in turn gives user the access through the layer that integrates data and analytics.

The principal obstacle for the effective use of big data is the characteristic of the healthcare information. Different sources have their own silos of data. It is very difficult to integrate the data from various sources because of the privacy invasion and propriety, the complexity of the fragmented nature of data as well as the diverse schemas and standards of data. Sometimes the lack of metadata also poses as a challenge in integrating the data. Even though the data is being shared integrating the data within a silo

The biggest obstacle to effective use of big data is the nature of healthcare information. Payers, providers, research centres and other constituents all have their own silos of data. These are fundamentally difficult to integrate 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. Regardless of the possibility that everybody shared their information, there would be sufficient difficulties incorporating it inside the storehouse, substantially less outside it.

C. *Security*

The field of healthcare would be tremendously instigated from the democratization of big data; for instance researchers can easily work together or involve in peer analysis and eradicate repetition of efforts. Researchers can also easily recognize prospective opportunities where they can collaborate and contribute.

The advent of big data in cloud makes the sharing of data relatively easy and economical. Yet, there is a significant privacy and security concern which includes the Health Insurance Portability and Accountability Act (HIPAA).

Though there are processes which could facilitate and automate this access, there are many challenges and complexities involved. As patients, providers, and other parties involved such as researchers need secure access, the access of data should be controlled by proper regulation. The regulations does not restrict itself only to data storage or data production but also includes ePHI(electronic protected health information), whether inside the production application or stored on a, mobile device, laptop or even a printer as the security of data once it leaves the cloud needs to be assured .

D. *Standards*

Managing the innumerous standards creates difficulties in interoperability, at least to certain extent. The architecture of Big data solution architectures should be supple enough to handle additional sources and also manage diverse schemas and structures which is used for storing and accessing data. To ensure meaningful analytics, accurate and suitable metadata and semantic layers needs to be incorporated so as to accurately describe the data and provide business context and guidance, which includes the proper use of data. With the progressing standards the data quality would eventually improve.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 22, December 2017

E. Timeliness

Healthcare data is dynamic, hence most elements require constant updates to remain in trend and relevant. In dynamic and rapidly evolving system data timeliness is one of the major challenges in numerous healthcare settings, such as clinical decision support, as in to make decisions or deliver information that guides decisions.

To sustain in the field of data management, timeliness and accuracy are fundamental requirements when it comes to data analysis. Failing in either of these dimensions can trade off the quality of the data. Decision support is made simpler, faster and ultimately more accurate by big data as decisions are based on higher volumes of current and relevant data. In some scenarios clinical decision support provides faster response than the time it takes to execute a report or analytic query. To obtain the best possible decision the data and query structure has to be carefully framed and at the same time it also has to meet the constraints of the processing windows also. In order to The main principle of big data is to have everything, but in some scenarios it is not feasible or useful — at times the hoarder reflex must be checked to make rational decisions.

III. NEED FOR BIG DATA ANALYTICS IN HEALTHCARE

Big Data is envisaging the field of healthcare .Big data can be used to foresee epidemics, give proper remedy for diseases, the life span and avoid preventable deaths.

A. Providing patient driven services

The main goal is to provide faster respite to the patients by providing evidence based medicine and detect diseases at the earlier stages based on the clinical data available, reduce drug doses to avoid side effect and provide efficient medicine based on genetic makeups(1). This helps in reducing repeated visits to hospitals and reduces the cost for patients.

B. Fraud Detection

Payers should have the capacity to identify fraud based on examination of irregularities in billing data, procedural patient records or benchmark data. Big Data analytics help payers assess data to detect abnormalities like overutilization of hospital's services in short time periods, or identical prescriptions for the same patient filled in multiple locations. Big Data uses anomaly detection to detect these incidents in real time and alert providers to investigate them before payment is made.

C. Assisted Diagnosis

With help of big data analytics the clinical researchers can regain access to wide-range of knowledge pools across numerous data sources to help in the accuracy of diagnosing patient conditions. Bringing together distinct data sets into a single big data repository and apply algorithms for predictive modelling provides deeper vision by identifying distinctions in subpopulations. These nuances are very rare that it is difficult to identify in smaller research samples, but by applying algorithms to these individual data sets, nuances can be made unmistakably detectable. Analysing on the already available data produced by the patients, who already suffered from the similar symptoms, helps doctor to provide effective medicines to new patients.

D. Detecting spreading diseases earlier

One of the major breakthrough in data analytics is that we can forsee outbursts in viral diseases earlier before spreading with the help of live analysis. known by scrutinising the social logs of the patients suffering from a disease in a particular geographic location. This helps the healthcare professionals to recommend the targets or victims by taking necessary preventive measures.

E. Personalized Treatment Planning

Personalized treatment planning is a patient centric treatment for a patient. Here the effects of medication are continuously monitored. The treatment to the patient can be changed based on how the drug works. The analysis can be tailored depending upon the individual needs of the patient. MapRa technology in big data provides real-time access, at both the detailed and summary level so as to take timely and effective decisions.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 22, December 2017

IV. BIG DATA, SOCIAL MEDIA AND HEALTHCARE

Social media helps to increase communication between providers, patients, and communities e.g., patients with same health conditions and providers with analogous specialties. This will not only help to democratize and globalize healthcare, but it is also a vital source of big data. There are many challenges posed by social networking data such as lack of structure and volume, velocity. It also paves way to new challenges around integration and accuracy.

For example, if a group of patients are discussing quality of care about a provider, there will never be complete agreement among them. Perception among each patient will be different, and there will be misunderstandings, prejudices based on accidents and other factors. The challenge is to create useful information out of this pool of data to deliver information such as improvement guidance and provider ratings.

V. TECHNOLOGY CHOICES IN BIG DATA

There are several technology solutions for dealing with big data, ranging from open source to proprietary and from cloud to on-site and. On-site options that can tame big data include Teradata, Vertica (HP) and Netezza (IBM). These solutions have low time to value and maintenance but have comparatively high total cost of ownership.

Cloud-hosted software as a service (SaaS) to an extent reduces the barriers of partaking in the big data arena. Map Reduce-based solutions to process huge datasets using a large number of computers — e.g., terabytes of data on thousands of computers are implemented by Google and Amazon. MapReduce algorithms take huge problems and split them into a set of distinct tasks that can then be disseminated to a large number of computers for processing and then the results are combined to form a solution. Other cloud-based solutions include Tableau, which supports visualization.

Most companies use Open-source Hadoop as a framework as it provides scalable, high-performance, low-cost option for dealing with big data. It is important to have training, professional services and support to efficiently deploy Hadoop solutions using the open source framework. Many vendors such as Microsoft, IBM Greenplum, and Oracle have commercialized Hadoop and have assimilated and aligned with the rest of their analytic offerings and database.

SaaS is one of the major technology players for democratizing the results of big data. SaaS-based solutions allow healthcare entities that control subsets of data to expose access through services that overcome the integration and aggregation challenges. Services that facilitate analytics which are both basic and advanced can also be made part of the overall offering.

VI. DATA ANALYTICS FOR HEALTHCARE

A. Information Extraction

Information extraction is one of the important phases in data analytics. The analytical platform should provide intelligent search and extract information for healthcare. During this process secure navigation, analysis, and discovery across diverse structured and unstructured data formats in claims databases, laboratory systems, electronic health records, digital radiology and imaging systems, and other information storage applications.

The framework acts a virtual data centre and helps to accessing different locations across multiple networks concurrently. The main role is to enable interoperability among existing healthcare information systems in multiple organizations and different data centre locations.

B. Feature selection

It is expected that the best way to take a decision is by considering all the factors that might impact the outcome and by weighing their relative importance. When we tend to overweigh peripheral variables at the expense of critical ones there are chances of taking all factors into account. Feature selection algorithm helps to minimize the risk of overweighing peripheral considerations and helps to identify the major variables in the data set .

C. Predictive Modeling

To understand the probability of predict outcomes the predictive modeling techniques are increasingly used by data scientists. However, for a high dimensional data model, a major hiccup is determining which features should be included in the models. In many cases the feature selection algorithms removes non-informative features from model but there are many different classes of feature selection algorithms. Predictive Modeling leads to important insights

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 22, December 2017

involving clinical researchers predicting patient outcomes from electronic medical records.

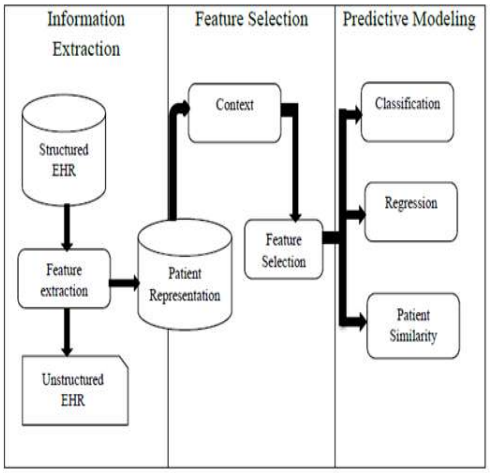

Fig. 1. Data Analytics for HealthCare

VII. BIG DATA ARCHITECTURE IN HEALTHCARE USING MAPR

The MapR Distribution with Hadoop brings the high volume of structured and unstructured healthcare data into a single storage that makes use of existing hardware and network components. With help of secure HIPAA-compliant Hadoop-enabled architecture multiple groups in healthcare organizations can store and access this data simultaneously at low cost. With the availability and easier access to wide-ranging patient data and medical research, doctors can spot and diagnose diseases in their early stages and assign more effective treatment based on a patient's genetics. They can also adjust drug doses thereby minimizing side effects and provide effective treatment. This comprehensive view also offers opportunities for improvement in care coordination, outcomes-based reimbursement, patient engagement and outreach and population health management. [4] discussed about Positioning Of a Vehicle in a Combined Indoor-Outdoor Scenario, The development in technology has given us all sophistications but equal amounts of threats too. This has brought us an urge to bring a complete security system that monitors an object continuously. Consider a situation where a cargo vehicle carrying valuable material is moving in an area using GPS (an outdoor sensor) we can monitor it but the actual problem arises when its movement involves both indoor (within the industry) and outdoor because GPS has its limitations in indoor environment. Hence it is essential to have an additional sensor that would enable us a continuous monitoring /tracking without cutoff of the signal. In this paper we bring out a solution by combining Ultra wide band (UWB) with GPS sensory information which eliminates the limitations of conventional tracking methods in mixed scenario(indoor and outdoor) The same method finds application in mobile robots, monitoring a person on grounds of security, etc.

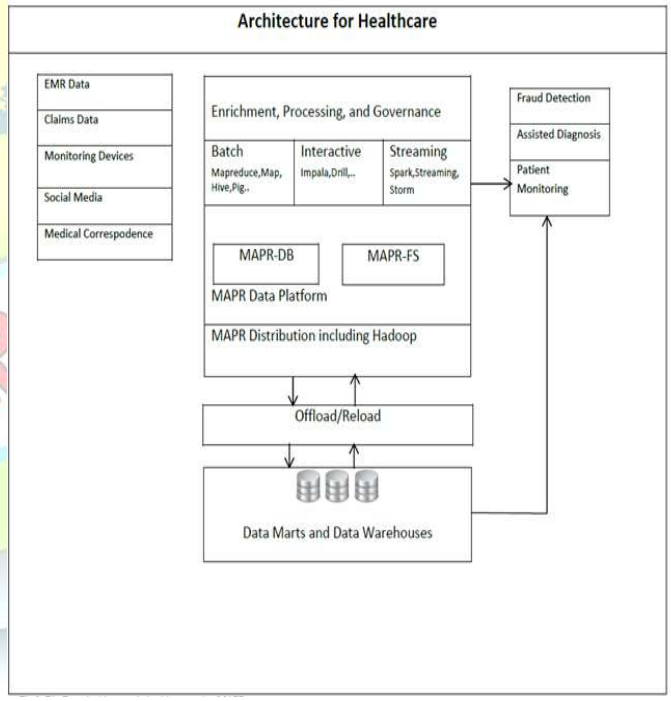

Fig2 Big Data Architecture using MapR

VIII. CONCLUSION

In the rapidly growing digital world we are moving towards evidence-based medicine, where we make use of the available clinical data and integrating it into clinical and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 22, December 2017

advanced analytics. Apprehending and coordinating all of the information about a patient together gives a more rounded view and provides insight into care coordination and outcomes-based reimbursement, population patient engagement and outreach health management. Attaining this 360-degree view of the patient helps to remove redundant and expensive testing make it more feasible for the patients, minimize errors in prescribing and administering drugs, and it also helps to avoid preventable deaths and increase the life span of mankind.

REFERENCES

[1] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," Health Information Science and Systems, vol. 2, no. 1, p. 1, 2014

[2] S. Chandra and D. Motwani, "An approach to enhance the performance of hadoop mapreduce framework for big data," in Micro-Electronics and Telecommunication Engineering (ICMETE), 2016 International Conference on. IEEE, 2016, pp. 178–182.

[3] J.Archenaa1 and E.A.Mary Anita2 "A Survey Of Big Data Analytics in Healthcare and Government", in 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

[4] Christo Ananth, S.Silvia Rachel, E.Edinda Christy, K.Mala, "Probabilistic Framework for the Positioning Of a Vehicle in a Combined Indoor-Outdoor Scenario", International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 2, Special Issue 13, March 2016, pp: 46-59

[5] White Paper by SAS, How Government are using the Power of High Performance Analytics, 2013.

[6] Michael Cooper &Peter Mell, *Tackling Big Data Problems*, NSIT Computer Security Workshop

[7] Jean Yean, *Big Data Bigger Opportunities*, White Paper, U.S.General Services Administration.