BIG DATA BENEFITS IN HEALTHCARE

Abstract:

The world is surrounded by data today. The quantity of data that we are using, and harvesting is increasing meteorically. Increasing use of the latest innovations and social media is generating huge data which can be of a great value if used properly. These huge datasets know as bigdata usually doesn’t fit in a limited storage due to its enormous size. The fields of science, engineering and technology are producing data at an exponential rate leading to Exabyte(s) of data every day. Many organizations are coming up with ways of storing this huge data which brought a great deal of attention towards big data analytics. In recent advances in healthcare during data collection during routine health care in the form of EHR (Electronic Health and Records), medical device data and insurance claims data created tremendous opportunities in health analysis.(1) In this paper, we review the background and the various methods of big data analytics in healthcare and we identify and present the elaborate view addressing the relevant challenges in two broad categories: data, accessibility.

Introduction:

The new advances in Information Technology (IT) guide to smooth creation of data. For instance, 72 hours of videos are uploaded to YouTube every minute. Especially, in the healthcare sector huge amount of data for maintaining health records and health care. Huge data is obtained from various sources like health checkup data which can be used to monitor and predict future health fluctuations for a patient. Due to this application personalized health monitoring can be done which can suggest the type of diet the patient to take to avoid certain problems. Storing the data in printed form is currently outdated and digitized records are used to store as the unlimited store in cloud. By description, big data in healthcare refers to electronic health datasets so large and complex that they are difficult to manage with traditional software, hardware, data management tools and methods.

Health care industry also provides data such as doctor’s notes, prescriptions, medical images like X-ray and other scans, insurance data, drug documents, medical journals and publications. By this data we can locate the causes of disease and the region it is first reported. By this we can act upon it and alert the remaining areas accordingly. Therefore, big data analytics applications in healthcare take advantage of extracting insights from data for better decisions making purpose. Analytics of big data is the process of inspecting enormous amount of data, from different data sources and in various formats, to deliver insights that can enable decision making in real time.(2)

Healthcare is a prime example of how the three of the V's of data, velocity (speed of generation of data), variety and volume , are an innate aspect of the data it produces. This data is spread among multiple healthcare systems, health insurers, researchers, government entities, etc. Furthermore, each of these data repositories is siloed and inherently incapable of providing a platform for global data transparency. To add to these three V's, the veracity of healthcare data is also critical for its meaningful use towards developing translational research. Moreover, due to the variation in data complexity and structures, unavailability of computational technologies, and concerns of sharing private patient data, few projects of large clinical data sets are made available to researchers in general. In the rest of this paper, firstly we introduce the common background, definitions and properties of big data. We will discuss in-depth in the domains of data quality, accessibility.

Related Work: (3)

Big data is mainly based on its core values which are V’s where some of them are Volume. Velocity, Variety and Veracity.

Volume is mainly based the data flow in the related field and the data which is used daily in social media It defies the storage and the level of research which should take place.

Velocity: This mainly deals with data which is transmitted or in motion and its speed . This helps in analysis of speed and how to increase the acceleration. It is a time sensitive entity which changes on the time.

Variety: This is basically the types of data in a field like structured like numerical data, text information, un structured like audio, images of scans, video and semi structured data. It also defines the complexity of the data coming in and the type of analysis used in producing accurate outputs.

Veracity: It mainly talks about the credibility of the information like the information we are taking in is real or face, does it have any credibility, is that masked. There are many strategies to tackle this problem because without that the conducted research is not worth the time to spend as the results won’t be as realistic as expected.

The obtained data is used in many analyses like exploratory analysis which shows the where about of the data and its basic features. To get a deeper and detailed view then we need to descriptive analysis which shows the patterns in the data. Then we can do predictive analysis , machine learning to use the use the trends and predict future results in the data.(4)(8)

This data can be used to predict when a person is going to be sick and the type of medication he should use. There are devices which monitor heartbeat and send the signals when there is any fluctuation in that. These can be used in data analysis as they are useful in real life.

Contribution :

Data :

Big data is not as big as it seems in healthcare

Collection of data and storing that staring before a decade in healthcare industry in healthcare industry by various federal agencies and private enterprises during their routine health check and insurance claims where as the data collection and usage in other fields began even before that. Everyone is not ready to distribute their own health records and their patients as they can be used in a wrong way. The publicly available datasets are mainly for research which is intensive care unit dataset and an important resource for clinical research. For example, a researcher is studying on drug interaction in humans then he would study multiple drug interaction. So, for this detailed analysis he needs the data of the patient condition before couple of weeks of the occurrence and couple of weeks after this which gives him clear effects of the multiple drug interactions. So, the subject wouldn’t be comfortable in sharing such a huge data with other person. There are the usual problems of data in health care industry.

Access :

Accessibility to scientific research and sharing health data needs validation due to patient privacy concerns and technology issues as it is easy to misuse the data with recent technology. This is also a huge issue right now as the data is not safe and the access to this data is not quite easy. There are many challenges like privacy, shareability, proprietary rights

Privacy:

Allotting access to health data is a vulnerability if it falls in wrong hands due to patient privacy considerations which are protected by government laws such as Health Insurance Profitability and Accountability Act of 1996.(5) There is always a fear of breach in patient’s perspective in the case of sharing data. For example, a company named protenus which is a data analytics company on health systems had a major breach in 2019 followed by a breach in 2018 which resulted in a loss of 45 Million Records in total.(6) These data breaches are occurred in health care and the data is quite important because of its privacy.

Shareability:

Sharing the data is also is also a compilated process in research works as maximum health data is masked to prevent the impact on companies in a case of data breach. If a type of data to be encoded/masked it need a huge process which the hospital needs to recruit professional in that field which is a huge decision for hospitals because they usually invest in the resources like MRI machine , equipment’s. This is the reason why sharing is a huge issue in health care. Even if the original data is shared for research purposes it is quite hard to share the code and flow of the analysis which is masked. Moreover, the data which is used in research is obtained from patients and in major cases the permits are not asked. So, there should be a law where the person should know when his data is used and for what purpose which make it transparent. There is few open source software like Git-hub where they follow disciplines and act upon it which made them success as they rely on the engagement of many talented scientist and engineering on their will to contribute for this developing community.

Preventing the spread of disease:

Healthcare data can be highly used in this area where the data is used to predict the outcomes. For example, if a place called City A is where a disease started and then it spread to city B which cause the spread to City C and City D. If we have the data on that from past then we can use that in future like if a new disease started spreading in City A then we can alert City B,C,D so on, and send the antidote to other city which decrease the chance of spreading the disease. Creating awareness on this can prevent major losses in the are but the data is quite sensitive where some places can be targeted. So, the data should be gathered by WHO and analyzed as well as stored properly.

Results:

The major factor in development of health care sector would be cloud computing as there is a huge need and demand for that. The data need to be stored in cloud as there is no way to store such huge data in offline storage units. For this we need to use trusted storages and online applications which can decrease the risk of breach in data like Amazon S3, Google Cloud. So, I think if we can gather the information regarding diseases and the areas, they’re spreading we can find a pattern which can result in the capability of stopping the future consequences. Moreover, Masking the original data would be a good idea which doesn’t expose the actual and original data which is supposed to be stored authentically.

Conclusion:

Big data in health care is increasing gradually in recent data due to its importance but is neglected till now. We need to create awareness to share the data which is used to stop the future issues. One of the essential budget-oriented talks is about its security and authenticity.(7) There is also a huge requirement in research in health care which needs to be implemented right now. “Big data” started with many believable promises in health care, but unfortunately, realistic health care is different from other disciplines with additional constraints of data quality, privacy, and regulatory policies. We can analyses the healthcare using geographical location where the whole data is available and the places where the victims are in huge number. This helps prioritizing in the case of distributing vaccinations as the place with most infected has more reach and more use. Moreover, the rate of electronic health record (EHR) adoption continues to climb in both inpatient and outpatient aspects. Those data could be an enabling resource for deriving insights for improving patient care and reducing waste. The analysis is helpful for advanced detection of powerful treatments and provide personalized health care to each and every person.

References:

1. <https://www.nih.gov/research-training/allofus-research-program>.
2. Amir Gandomi, Murtaza Haider, “Beyond the hype: Big data concepts, methods, and analytics,” International Journal of Information Management 35, pp. 137-144, 2015.
3. <https://www.ijstr.org/final-print/mar2017/Improving-Healthcare-Using-Big-Data-Analytics.pdf>
4. <https://www.researchgate.net/publication/279198958>
5. Health insurance portability and accountability act of 1996 [[PubMed](https://www.ncbi.nlm.nih.gov/pubmed/9180525)]
6. <https://www.fiercehealthcare.com/tech/32m-patient-records-breach-2019-double-all-2018-protenus-reports>
7. J Michael McGinnis, Leigh Stuckhardt, Robert Saunders, Mark Smith, et al. Best care at lower cost: the path to continuously learning health care in America. National Academies Press. 2013 [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=National+Academies+Press&title=Best+care+at+lower+cost:+the+path+to+continuously+learning+health+care+in+America&author=McGinnis+J+Michael&author=Stuckhardt+Leigh&author=Saunders+Robert&author=Smith+Mark&publication_year=2013&)]
8. <https://builtin.com/artificial-intelligence/machine-learning-healthcare>