Application of Big Data Analytics with Evidence Based Medicine

Aravind G[1], Varun K[2], Manjunath C R[3], Soumya K N[4]

[1,2,4]Department of Information Science & Engineering, [3]Department of Computer Science & Engineering
School of Engineering & Technology – Jain University
Bangalore, Karnataka, India

ABSTRACT

Numerous clinical practices used in certain organizations are said to be of an age old practice whose results are not without side effects. Evidence Based Medicine is the approach that enables an optimum decision making system with the acquisition of the best evidence based of a clinical research. The system improves the quality of healthcare and additionally provides certain standards that every organization can abide. The Indian Health Ministry laying down the standards for EHR (Electronic Health Records) only makes it eminent that data stored will be digitized. With Blockchain technology on the rise, organizations will be sharing patient data more conveniently and securely. Data collected from patients and also from clinical research, especially in India will be immense. As data grows in volume and complexity, the need for analytics arises. Big Data Analytics is the domain which deals with cleansing surmounts of complex data and gaining valuable insights from it using various computational techniques. Implementing big data analytics in an evidence based system will aid the health organization to visualize and find the best course for the patient's condition from the data provided in the patient's health record in accordance with the clinical research data.

Keywords: Evidence Based Medicine; Electronic Health Records; Big Data Analytics; Blockchain

Introduction

The current medicine system in India, provides no standards to the recommendation of a clinical practice for a patient. Not only are the standards absent, even the documentation of patient's health history is absent. Lack of medical history contributes to the improper judgement of the patient's condition, resulting in a medical error.Medical errors are being regarded as one of the most cause of deaths in the country. This may be due to improper analysis of the patient's health condition or lack of proper decision making mechanisms or maybe even outdation of the medical practice. Many patients do not get the treatment that is required for them to cure or alleviate their medical condition. This results in patients being treated in a way which produces side effects and consequences.Even if a medical study is published that is said to be acure, it takes a very long time for it to applied into practice on a large scale which could've saved many lives.

Big data is term that doesn't only imply the voluminous of data, but also the complexity and variety that comes with it. The need of analytics arose as data is growing in complexity and size. Complexity and the volume of the data that was collected made it impossible for humans to gain any sort of value of it. Advancement of computational techniques and hardware provided the right platform to perform these analytics on a computer. This led to the widespread usage of Big data analytics across various domains. Healthcare is also domain which can utilize big data analytics to its true potential.

Evidence Based Medicine is not a procedure that was developed recently. This procedure was stated many years and it's only recently that health organizations across some parts of the world have implemented it in practice. Numerous studies and researches are being

conducted on a daily basis and along with it, valuable evidences are being collected. This data satisfies the six V's of Big data that are Volume, Velocity, Variety, Veracity, Variability and Value. These studies are termed as external evidence. Similarly, the data collected from the patient's health records are called as internal evidence. In consensus with the patient, the best practice is conducted after the analytics of the external evidence in accordance with the internal evidence. Transparency, quality and a standardized approach is implemented with this system.

1. Background Work

The various implementations and reforms performed on Indian Healthcare in recent times.

i. The implementation of EHR has not been widespread across India. In fact, only few private Hospital groups have put it into use. Currently, the Apollo group have implemented in various hospitals situated in Chennai and Hyderabad. Sankara Nethralaya has partnered with TCS (Tata Consulting Services) for its version of EHRs. Despite having HER implemented in a few private organizations, not much data is being transferred as only few private hospitals have this system.

ii. A detailed assessment has been done for the proposed approach to fasten the process of a wide adoption of EBM in India through practice of Standard Treatment Guidelines (STGs). This initiative has been taken by Ministry of Health & Family Welfare, Government of India with the association of Federation of Indian Chambers of Commerce and Industry (FICCI).

iii. The Ministry of Health and Family Welfare has provided standards for the implementation of EHR (Electronic Health Records) in 2016. These standards include IT standards, Data ownership and privacy and Data security.

iv. India is known to have wide-range of medicine systems such as Ayurveda, Yoga and many more. Despite the variety, none of those systems provide organized and are not up to date with the current research studies.

v. A conference was held in 2013 which was themed after Evidence based medicine in Dermatology by the IADVL. Potential talks about the application of Evidence Based Medicine were made, hoping for a nationwide implementation of the procedure.

vi. In a paper published in 2014, author Bhushan Patwardhan, reviewed the integration of Ayurveda with Evidence based medicine. He goes on to talk about how documentation and proper methodology will be crucial to implement Evidence based Ayurveda. Ayurveda was not scientifically validated which meant the various methods and practices in Ayurveda were not evaluated or practiced. If it was validated, the unison of western and eastern medicines would reveal lots of new evidences and may give rise to new projects.

2. Comparative Study

The current guidelines in India were rated poor by The Journal of Evidence Based Medicine. These guidelines do not have any systematic assessment of evidence and did not follow any proper methodologies for providing recommendations, which evidently lead to poor scores in methodological domain. The current medical practices are very outdated with the research studies. This could save a lot of lives if it had the system been on par with the recent studies. Especially with the rise in number of publications of EBM on PubMed.
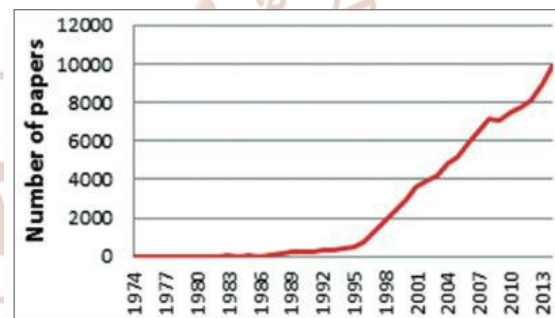

Fig 2.1: Publications of EBM papers on PubMed over the years

Upon searching for evidence based medicine articles on PubMed, out of the 89,749 articles only 512 articles are from India. This is a clear indication EBM is a known system in India but the implementation is lacking because of training and technology required to perform such analytics. Another issue with the current system is that, one person might get a more practically advanced medicine from the recent studies in a quality hospital. Whereas, another man with the same issue might be treated with an age old method which had

been proven to be ineffective. Records are rarely maintained in the current system. Potential inference/analytics can be gained from records, but due to the lack of records it isn't currently possible. This hinders the application of EBM in many health organizations. There are very few institutions which provide Evidence based education and one of them is Sanjay Gandhi Post Graduate Institute of Medical Sciences. Despite providing evidence based education, there has been no evidence of infrastructures or logistics for EBM implementation. The western countries such as US, UK and Germany are the pioneers of Evidence based medicine and are advanced in their usage of this method. This enables these countries to provide many articles/studies to the field of healthcare in the topic of evidence based medicine.

In a study that reviewed conventional education methods to evidence based methods, it was seen that evidence based medicine provided more critical appraisal skills among emergency medicine residents. Thus improving the quality of practice that is performed. With the help of Big data analytics and other techniques such as Artificial neural networks, it is possible to predict mortality of gastroenterology. Compared to the conventional system, evidence based medicine requires lots of investment of computational hardware and field experts who are proficient with the software. The requirement of clinical expertise is also required for EBM to be done effectively.

3. Analytical Approach to EBM

There are numerous analytical techniques available across the world. Many new techniques are invented almost every day. In this section we provide one of the analytical approaches to Evidence Based Medicine using the Hadoop distributed architecture. Even though India usage of Evidence based medicine is almost nil, the advancement in technology with the usage of EHR (Electronic Health Records) enable its viability to a certain degree of effectiveness. The pivotal component of this approach is EBM itself. EBM has five main steps which are listed below:

i. Problem definition or ask a clinical question
ii. Acquire best evidence from the resources.
iii. Evaluate or Assess the evidence.
iv. Apply the evidence into practice.
v. Evaluate or Assess the performance of the evidence in practice.

Firstly, we acquire the internal evidence. As mentioned before internal evidence is the clinical evidence acquired from a patient's health record. In this step we form a question or determine a problem definition best of the internal evidence. This step enables form the right question to which we can find suitable or apt answers for it. One of the formats used to determine the clinical question is the PICO (Problem Intervention Comparison Outcome) format. The PICO is more simplistic terms tells us about the patient's problem, the probable procedure or strategy that can be taken, comparing the strategy with other procedures and lastly what were the results and consequences by performing the procedure which we are interested in.

Secondly, acquiring the best evidence. This process focuses on acquiring the evidence which is best suited for the patient's current problems. The evidence acquired here is known as external evidence as it is derived from various studies and research. The ideal information is of high quality, easy to access, available easily and list all the harms and benefits of the evidence. The below image shows us how the quality of the evidence for different type of data sources.

Thirdly, assess the evidence. From the loads of research data which we acquire, there are multiple evidences we may encounter. Some of the evidence maybe bad evidences as well. For that reason, we must critically appraise the evidence to determine its validity for the current problem. To determine the validity of the evidence, a team of clinicians with several years of experience in assessment of evidence must be present to evaluate the external evidence in question. Only upon finding the best one we move on to the next procedure.

Fourthly, we apply the evidence found from the previous to the patient. Since not all clinicians will understand the value of adopting a new standard of care, it's essential to communicate the change and the benefit it adds to patient care. This is probably a crucial step in the process, if not the most complex. Conversion of the evidence into a consumable unit for the computer is done in this step so that data can visualized for more insights

Lastly, evaluate the performance of the evidence. Applying of the appraised evidence may be the best course to be taken, but it can provide new evidences or side effects and all kinds of evidences. During this

process it is important to assess whether certain evidence, which is applied to the patient, caused changes to better and that to the extent that it is confirmed by research. If the resultant data differ significantly from the expected data, it would be necessary to investigate why some patients did not respond to the procedure applied and what could be changes that are needed to be applied to get the expected results. It is important to keep on reassessing and revaluating evidences as they arise.

The concept of Big data analytics fits right within this process. While retrieving external evidence is the step that requires some of the concepts that are applied to big data. The sheer amounts of data collected is very large so we can say that data has a large volume. Many publications and studies are being posted on a daily basis. This indicates that data is growing on a daily basis with great velocity. Since these publications are based off real tests we can be sure of its veracity as well. We acquire data in the form of X-ray scan images, MRI scans, EKGs, blood pressure reading and many more. Each of these data can be structured or unstructured or even semi structured hence having a lot of variety. Finally, these data are required to provide some kind of insight or some value. The reason why Big Data Analytics is applicable here is because the data extracted during this process conforms to the six V's of Big data.

Data sources are main of two types namely, internal and external sources. Internal sources are provided from the clinic. These evidences are derived from decision support systems, CPOE, clinical trials, etc. Whereas the external evidences are pooled in from sources like government sources, laboratories, pharmacies, insurance companies & HMOs and many more. Data can be derived in multiple formats for instance flat files, .csv, relational tables, ASCII/text, etc. It may be present in a different place geographically from another health organization.

In the initial steps, data is pooled in and the data is raw which means that no amount of cleansing or pre-processing has been performed yet. Data sources may vary in the second component. It can be from a web service or a data warehouse. The process of Extract, Transform and Load(ETL) provides the facility to retrieve data from different sources and provide pre-processing like cleaning and transforming. Pre-processing is done so that it can be used and understood by a machine or a tool.

Upon transformation, the data is sent to some of the Big Data Analytics tools such as Hadoop, Spark, MapReduce, Mahout and many more which are selected to perform analytics. Hadoop is the most significant open-source distributed data processing platform used for analytics of big data. Some algorithms from the domains Natural Language Processing, Clustering and Statistical Machine Learning are used to extract the most significant and relevant data from the data pool.

Hadoop's distributed architecture enables it process very large amounts of data. In this architecture the data pool is separated into small clusters which represent the smaller problem instances of the whole problem. These smaller problems are solved within their own separate servers and then are combined to provide the overall result. The file system used to store these clusters are called HDFS (Hadoop Distributed File System). To perform such tasks, it utilizes MapReduce. MapReduce collects the results from each node after processing and integrated it into a unified final result. Despite sounding simple, there are various tools listed above that also support Hadoop to perform these analytics. Hadoop uses a NoSQL type of database, that is the data is not stored in row or column like any generic SQL database. An example for such a database would be MongoDB. Hadoop enables data to produce significant and relevant data that are consumable and in an understandable format. This step requires massive processing capabilities. This means that computational hardware required are of high standards and efficiency. Hardware is expensive and an alternative would be to use cloud service providers such as AWS Hadoop, Cloudera, and IBM BigInsights.

In the final stage we have applications such as OLAP (Online Analytical Processing) which query the data processed by Hadoop. The data maybe of a single value or maybe in the form of rows and columns. Either way the data queried is very understandable and has significance. This data can be visualised using libraries such as Seaborn and Tableau. Visualizing data provide many valuable data which couldn't be deduced when the data was in its raw form. This is how big data analytics fit right into the EBM process. The best evidence is acquired through all of these processes and it is put into practice. The evidence after implementation is evaluated and checked if it works like in the research study or provides more

evidence. If more evidences are present, then the process is iterated.

4. Conclusion

Big data analytics not only facilitates the implementation of EBM but also makes it reach it true potential. The implementation of EBM reduces the prescription cost, for instance if the evidence recommends two probable therapies who bear similar consequences and benefits, but there is major difference in cost; then the physician would recommend the cost effective one.

Despite reducing cost during clinical trial, it does come with a high budget installation and maintenance. Hadoop is said to be a complex framework and would require some kind of expertise to work with it efficiently. With the increased usage of other frameworks such as Spark along with Machine Learning algorithms and Deep learning libraries, the possibilities of performing analytics seem to be endless . Artificial Intelligence is on the horizon as well. It won't be too long before machines become a physician. Despite providing a better decision making system, it can never replace human intuition.

It is hard to make a change with the current status in Indian Healthcare but with the reputation it has at the moment, a change is required. We are at least around 5 years behind in technology in healthcare with only private hospital organizations providing some state of the art healthcare. EBM provides the right platform for Indian Healthcare to be more organized and put the current trends in technology to effective use.

Incorporating analytical approach mentioned before to EBM would benefit healthcare with a more optimized decision making system. This approach also utilizes various trending technological advancements such as Hadoop, Spark along with other machine learning algorithms in an effective manner. The data that will be acquired will be large, complex and of variety. Gaining insights from these data grows in difficulty which puts analytics as a necessity. The mentioned approach describes one of the many different analytical approaches that can be taken to gain value from data, but the distributed architecture of Hadoop gives it the edge as processing of large amounts of data and also processing of variety of data types is done more efficiently in this approach.

5. References

1) Omar El-Gayar and Prem Timsina, " *Opportunities for Business Intelligence and Big Data Analytics in Evidence Based Medicine*", System Sciences (HICSS), 2014 47th Hawaii International Conference,9 January 2014

2) Wullianallur Raghupathi and Viju Raghupathi, " *Big data analytics in healthcare: promise and potential*", Health Inf Sci Syst. 2014,7 February 2014.

3) Sunil Kumar Srivastava, "*Adoption of Electronic Health Records: A Roadmap for India*", Healthc Inform Res. 2016;31 October 2016

4) Kameshwar Prasad, "*Evidence-based medicine in India*", Journal of Clinical Epidemiology, 8 November 2012

5) Soumyadeep Bhaumik,Soushieta Jagadesh,May Ellatar,Neeraj Kohli, and Muhammad Riedha, " *Clinical practice guidelines in India: Quality appraisal and the use of evidence in their development*", Journal of Evidence Based Medicine 2018, 11 January 2018.

6) Ravindra Bhaskar Ghooi and Shailesh Deshpande "Evidence-*based medicine: A non-starter in India?*" Indian Journal of Health Sciences and Biomedical Research, vol. pp 121-126,January 2016.

7) Rebecca Hermon and Patricia A H Williams, *"Big data in Healthcare: What is it used for?"*, Australian eHealth Informatics and Security Conference 2014,3 December, 2014.

8) Sriganesh Kamath and Gordon Guyatt, " *Importance of evidence-based medicine on research and practice*", Indian Journal of Anesthesia 2016,vol. 60, pp.622-625, 15 September 2016.

9) Bhushan Patwardhan, "*Bridging Ayurveda with evidence-based scientific approaches in medicine* ", EPMA Journal, November 2014.

10) Mostafa Alavi-Moghaddam, Shahram Yazdani, Fathie Mortazavi, Samira Chichi, and Seyed-Mostafa Hosseini-Zijoud, "*Evidence-based Medicine versus the Conventional Approach to Journal Club Sessions: Which One Is More Successful in Teaching Critical Appraisal Skills?* ", Chonnam Medical Journal,May 2016

11) Jaya R Mallidi, "*The Challenges in Practice of Evidence Based Medicine*", iMedPub Journals, vol.1, September 18, 2017

12) I. K. Moppett and R. M. Pearse,"*Evidence-based medicine: the clue is in the name*", British Journal of Anaesthesia, 4 October 2017