Volume-6, Issue-1, January - 2017 • ISSN No 2277 - 8160 IF : 3.62 | IC Value 80.26

Original Research Paper Engineering

Real Time Predictive Analytics in Healthcare Universal Electronic Healthcare Record

Pratik Nichat Computer Department, Vivekanand Education Society's Institute of Technology, Mumbai, India

Ankit Shaw Computer Department, Vivekanand Education Society's Institute of Technology, Mumbai, India

ABSTRACT *In India, patients visit several health providers, throughout their life span, right from visiting a sub-centre, community-centre or primary-health centre in rural setups, or a general practitioner in his local vicinity, to a government /private hospital or clinic at the district, city, state or central level. Health records get generated with every clinical encounter during these ambulatory, inpatient or emergency visits. However, most health records are either lost, or remain in the custody of health care providers and eventually get destroyed, as per the retention period of medical records generally followed by hospitals in the country. There may be negligible to no health records maintained by some private practitioners, at clinic setups and by rural healthcare setups. Also, important clinical data is not available for research and for reference to aide in clinical decision support. Study of disease trends and statistical analysis of clinical nature also suffers. So what we lack today is a good Electronic Health Record System in our country, a system through which even a patient can maintain his records and access them from anywhere. Proper Records and statistical analysis on them can pave way for better preventive healthcare to patient.*

KEYWORDS : Health records, Private Clinics, Rural setups, Clinical data, Statistical analysis, Elecronic Health Record (EHR)

I. INTRODUCTION

For a health record of an individual to be clinically meaningful it needs to be from conception or birth, at the very least. As one progresses through one's life, every record of every clinical encounter represents a health-related event in one's life. Today our reports from doctor and hospital visit are lying idle, which could have been a great source of information. One of the reasons is because of lack of proper Electronic Health Record System. There are some but limited to big hospitals and clinics. What if we can make such record system as easy to use as Facebook? And engage, encourage the patient to maintain their own E-Record.

One of the two purposes of this project is to build an easy way to use Electronic Health Record system, connecting different tiers of our Health sector. System patients themselves can update and maintain their own records. The patients will be able to access the data using an APP or a website. Second and important purpose of this research is to build and design predictive algorithms that will use the healthcare data to predict and analyze risk of diseases in real time, while the predicting model runs in the background of the APP.

So the aim is to build a validated system for monitoring electronic health records in real time and to develop sophisticated algorithms to successfully identify patients with potential for clinical deterioration, or at a risk of diseases, in real time by monitoring the EHR of the patients.

II. PATIENT CENTRIC ELECRONIC HEALTH RECORDS

The EHR available today are mainly for the care providers who maintain it and are usually found in big hospitals and clinics. They are not connected for other health care providers.

Our aim is to create a customer centric EHR system where patient will also have control over it - a system as easy to use, as posting and updating your Facebook profile. The data that different health care facilities provide, to a large extent will be uploaded directly to your record. Each user will have a unique care id (linked to aadhar id), which will be used. The patients can then confirm if the data uploaded is theirs. These providers will be able to access data of a patient only if it's generated by them, else they would need the user to give them the access. The Clinicians, Pathology Labs, Hospitals will be able to upload your data. Users will also be able to update their data at any time. Users will be able to use APP or website to add, update their data. While the health providers will have a special login option from where they can add, update records of their patients. Then the patient records will be analyzed in real time to give trends and insights to patients.

Also, (Abbott and Coenen 2008) believe that information systems and information technology occupy a high position in improving healthcare industries in general, and in Electronic Healthcare Record (EHR) in specific; for the reason that implementing such technologies can save costs and times associated with daily hospital data records, such as patients schedules and billing. This is in addition to improving healthcare performance and efficiency by eliminating manual data records and paper work, and alongside smooth and flexible tracking of patient details.

III. REAL TIME PREDICTIVE ANALYSIS

The main focus of the project is discovering hidden patterns, information from the data using data mining, Natural Processing Language, Deep learning, Predictive Analytics and Statistical Analysis.

Data Mining is described as a process by which data is gathered, analyzed and stored in order to produce useful and high quality information and knowledge. This term also includes the way of how this data is gathered, filtering and preparation of the data for use and finally the processing of data to support data analytics and predictive modeling (Russom 2011).

This will not only help in effective decision making but will also give a new dimension to preventive healthcare. This will make medical diagnostics faster, more accurate, and more accessible. So designing algorithms, learning models, finding patters will be a big part of the project.

Scope of this will be - clinical decision support, re-admission prevention, chronic disease management, give patients control over their health record. Objective of the research will be to provide accurate insights and predictions related to symptoms, diagnoses, procedures and medications.

Predictive analytics (PA) supports healthcare sectors to achieve a high level of effective overall care and preventive care, as predictive systems' results allow treatments and actions to be taken when all the risks are recognized in early stages, which aids for minimizing costs (Conley et al 2008).

Predictive analytics uses technology and statistical methods to search through massive amounts of information, analyzing it to predict outcomes for individual patients. That information can include data from past treatment outcomes as well as the latest medical research published in peer-reviewed journals and databases.

Prediction modeling uses techniques such as artificial intelligence to create a prediction profile (algorithm) from past individuals. The model is then "deployed" so that a new individual can get a prediction instantly for whatever the need is, whether a bank loan or an accurate diagnosis.

In medicine, predictions can range from responses to medications to hospital readmission rates. Examples are predicting infections from methods of suturing, determining the likelihood of disease, helping a physician with a diagnosis, and even predicting future wellness.

The statistical methods are called learning models because they can grow in precision with additional cases. So the learning algorithms will be running in the background and will predict and analyze in real time, the individual patient's records for future risk and also help in predicting outbreaks of diseases.

The technology era has added significant value to the healthcare decision support system, since decision making systems in healthcare care sectors can be enhanced by focusing on patient diagnoses, behavior, and prevention in order to reach a high level of care and improve healthcare economics (Cannon & Tanner 2007).

IV. TOOLS FOR PREDICTIVE ANALYSIS

The different tools and techniques that would improve healthcare data analytics in order to support descriptive, predictive and prescriptive healthcare data analytics are as follows:

- **Advanced Data Visualization (ADV)**

ADV is different from other standards bars and line chart, since it can scale its visualization for millions of data points, also can handle different data types. ADV is easy to use and supports analysts to explore data widely. ADV can reduce quality problems which can occur when retrieving medical data for extra analysis. Moreover, ADV can offer rich results and fluid interactions in order to reveal clinical hidden patterns in the data. (Powell 2014; Wongsuphasawat et al. 2011)

- **Presto**

Presto is a distributed SQL query engine used to analyze huge amount of data that collected every single day. There is nothing better for healthcare sectors to find such a product which can handle a large amount of data that will come into the system. Data can take many hours and even days to be analyzed, but with Presto data now can be analyzed in just seconds or minutes. (Wulff 2013)

- **Hive**

Hive is one of the programs developed in order to handle large amount of data, it's is not processing and analyzing data quickly as presto, however Hive does all excel tasks efficiently that don't need for real time performance, due to this companies can use both Presto and Hive for best performance, since presto can access data stored on Hive. (Capriolo et al.2012)

- **Vertica**

Vertica program is very similar to Presto, but less expensive for the reason that Vertica eliminated costly architecture that used to associate with large amount of data. Also, Vertica has the feature of scalability which means it can cover hospital's data and analytics no matter how that data is big. Vertica can improve healthcare by reducing operational costs, accelerating medical reports and analyzing patients' patterns. (Vertica 2010; HP 2013)

- **Key Performance Indicators (KPI)**

KPI is a strategy evaluates in how company is executing its strategic vision. KPI can improve quality of medical healthcare for patients who are susceptible to hospital conditions when KPI used to specify significant indicators to be monitored and corrected, as well as identifying weaknesses. Also, KPI can use electronic medical record data to identify human practice and interventions. (Al-Azzawi 2014)

- **Online Analytics Processing (OLAP)**

OLAP can improve healthcare system by performing statistical calculation very fast through hierarchal and multidimensional organized data, and can increase data integrity checking, quality control and reporting services. OLAP has the ability to improve healthcare decision making system by giving a better tracking of medical records and diagnoses. (Pesi et al. 2009)

- **Online Transaction Processing (OLTP)**

OLTP is similar to OLAP, but it is designated to process patient care operations, such as patient registration, hospital documents and results review. (Ledbetter and Morgan 2001)

- **The Hadoop Distributed File System (HDFS)**

HDFS enhances healthcare data analytics system by dividing large amount of data into smaller one and distributed it across the other systems. Eliminating data redundancy, since HDFS has such feature built into storage layer which makes professional to focus on other responsibilities. HDFS can add a value through helping medical purposes in order to personalized treatment planning, assisting diagnosis, monitoring patient's signs and fraud detections. (Shvachko et al. 2010; Datastax 2013; Nori 2014)

- **Casandra File System (CFS)**

CFS is also distributed system like HDFS, however CFS is a designated system to perform analytic operation with no single point of failure. (Datastax 2013; Lakshman and Malik 2010)

- **MapReducing System:**

MapReducing system breaks Task into subtasks and gathering its outputs, as well as it enables many of the most common of operational calculations to be performed efficiently in a large amount of data. MapReducing system keep tracking on each server when tasks is being performed. The key strength of using MapReducing is the high level of parallelism, since many tasks can have performed at the same time if it's not waiting for other tasks results. (Dean and Ghemawat 2008)

- **Complex Event Processing (CEP):**

CEP has come recently to the healthcare sectors, which means an event of changing in state, for instance supposes a patient gained more weight and moved from obesity to morbidity obesity. Now complex patient event processing will detect this new pattern and add it to the patient's events and relate it with being diabetic, which means that complex event processing is relating and linking events to the real time, as well as that will enhance EHR systems. (Webster 2011)

- **Text Mining:**

Text Mining tools can be used and add a value in healthcare in terms of analyzing clinical records from the hospital emergency departments of physician response on call, as a similar complaints called the emergency department and were treated differently depending on the person who answered the phone. Such matter can effect in the quality of healthcare, as well as costs. Therefore, text mining can offer a treatment plan which will develop some standards and protocols to understand this matter. (Raja et al. 2014)

- **Cloud Computing:**

Cloud computing has increased hospital flexibility in order to respond for dynamic changes and latest medical updates, in addition to demonstrate a great healthcare value by reducing costs, increasing productivity and security and improve data analysis with

minimal management effort or service provider interaction. Cloud computing reduce strain which caused by huge amount of clinical data. One of the cloud innovations is Phillips Healthsuite platform that manages healthcare data and support doctors and patients. Phillips Healthsuite platform stores a huge amount of clinical and patient data which can be used directly in the future as an actionable data, a source of diagnosis analysis and disease prediction and prevention to increase patient care. (IBM 2011; Philips 2015)

- **Mahout:**

Mahout is an apache project aims to generate applications that supports healthcare data analytics on Hadoop systems. (Hortonworks 2015)

- **JAQL:**

JAQL is a functional query language aims to process large sets of data. JAQL facilitates parallel processing by converting high level queries into low level ones. JAQL assists and works well with MapReducing. (Beyer et al. 2011)

V. CONCLUSION

Using the proposed system will promise to leverage large amount of healthcare data properly, since doctors and nurses will be able to determine diseases and risks easily like some certain types of cancer, diabetes and blood pressure, as well as provide needed treatment in the right time. It will also enhance doctor's decision making process by defining better care, developing drugs and vaccines along with a better treatment plan in order to reach patient satisfaction. Moreover, proposed system will add a benefit of identifying risks early and mitigate it as much as possible. However, this study will need to push both doctors and patients to adopt new technique and collaborate together to reach high level of connection between both medical staff and patients in order to keep the system up to date and gather high quality of data.

This project will provide a way to securely store the medical records in a EHR database and help patients and the health sector with better preventive healthcare. It will give patients control over their records. Hospitals, pharmaceutical companies and insurance providers will see changes as well.

So overall this system will provide real time analysis of patient record to provide better care, directly to patients and a patient centric EHR system.

REFERENCES

1. Abbott, PA & Coenen, A 2008, 'Globalization and advances in information and communication technologies: The impact on nursing and health', Nursing Outlook, vol. 56, no. 5, pp 238 -246
2. Al - Azzawi, H. 2014. "Caradigm healthcare analytics." http://www.caradigm.com/media/68911/Caradigm - WP - Healthcare –Analytics – Jan – 2014 – US - EN.PDF Retrieved 09 August, 2015.
3. Beyer, K., Ercegovac, V., Gemulla, R., Balmin, A., Eltabakh, M., Ozcan, F. and Shekita, E. 2011. "Jaql: A Scripting Language for Large Scale Semi-Structured Data Analysis." http://web.cs.wpi.edu/~meltabakh/Publications/Jaql-PVLDB2011.pdf Retrieved 09 August, 2015.
4. Cannon, M., & Tanner, M. (2007), Healthy competition: What's holding back healthcare and how to free it Cato Institute, Washington, D.C.
5. Capriolo, E., Wampler, D. and Rutberglen, J. 2012. "Programming hive." http://www.reedbushey.com/99Programming%20Hive.pdf Retrieved 09 August, 2015
6. Conley, E., Owens, D., Luzio, S., Subramanian, M., Ali, A., Hardisty, A., & Rana, O. (2008), 'Simultaneous trend analysis for evaluating outcomes in patient - centred health monitoring services', Health Care Management Science, vol. 11, no. 2, pp. 152-166.
7. Datastax. 2013. "Comparing the hadoop distributed file system (HDFS) with the cassandra file system (CFS)." http://www.datastax.com/wp-content/uploads/2012/09/WP - DataStax - HDFSvsCFS.pdf Retrieved 09 August, 2015.
8. Dean, J. and Ghemawat, S., "MapReduce: simplified data processing on large clusters", Communication of The ACM -50th Anniversary Issue: 1958 –2008, 51, 1, January, 2008, pp 107-113.
9. Hortonworks 2015. "Apache mahout." http://hortonworks.com/hadoop/mahout/ Retrieved 09 August, 2015.
10. IBM. 2011. "Cloud computing: building a new foundation for healthcare." https://www-05.ibm.com/de/healthcare/literature/cloud-new-foundation-for-hv.pdf Retrieved 09 August, 2015.
11. Lakshman, A. and Malik, P., "Cassandra: a decentralized structured storage system", Operating Systems Review, 44, 2, April, 2010, pp 35-40.
12. Ledbetter, Craig S. and Morgan, Matthew W., "Toward best practice: leveraging the electronic patient record as a clinical data warehouse", Journal of Healthcare Information Management, 15, 2, Summer, 2001, pp 119-131.
13. Nori, S. 2014. "5 ways hadoop can help healthcare organizations and you" http://www.smartdatacollective.com/sameernori/282021/5-Ways-Hadoop-Can-Help-Healthcare-Organizations-And-You Retrieved 09 August, 2015.
14. Peši ,S. Stankovi, T. and Jankovi, D. , "Benefits of using OLAP versus RDBMS for data analyses in health care information systems", Electronics, 13, 2, December, 2009, pp 56-60.
15. Philips. 2015. "Open a world of cloud-based collaborative care." http://www.usa.philips.com/healthcare-innovation/about-health-Suite Retrieved 09 August, 2015
16. Raja, U., Mitchell, T., Day, T. and Michael Hardin, J. 2014. "Text mining in healthcare. Applications and opportunities." http://www.researchgate.net/publication/24182770_Text_mining_in_healthcare._Applications_and_opportunities Retrieved 09 August, 2015.
17. Russom, P. 2011, 'Executive summary big data analytics', Tdwi best practice report, Renton, Washington.
18. Shvachko, K., Kuang, H., Radia, S, and Chansler, R. 2010. "The hadoop distributed file system" http://zoo.cs.yale.edu/classes/cs422/2014fa/readings/papers/shvachko10hdfs.pdf Retrieved 09 August, 2015.
19. Vertica. 2010. "The vertica analytic database technical overview white paper." http://www.vertica.com/wpcontent/uploads/2011/01/VerticaArchitectureWhitePaper.pdf Retrieved 09 August, 2015.
20. Webster, C. 2011. "Clinical intelligence, complex event processing and process mining in process-Aware EMR / EHR BPM systems." http://chuckwebster.com/2011/07/clinical-intelligence/clinical-intelligence-complex-event-processing-process-mining-process-aware -emr-ehr-bpm-Systems Retrieved 09 August, 2015.
21. Wongsuphasawat, K., Gomez, J. A. G., Plaisant, C., Wang, T. D., Shneiderman, B., and Taieb - Maimon, M., "LifeFlow: visualizing an overview of event sequences", Proc. ACM SIGCHI Conference, ACM Press, New York, May 2011, pp 1747-1756.
22. Wulff, F. 2013. "Presto." https://prestodb.io/ Retrieved 09 August, 2015.