The Role of Big Data Mining in Healthcare Applications

Laura Elezabeth[1], Ved P. Mishra[2], Joanita Dsouza[3]

Amity University Dubai, UAE lauraeliza888@gmail.comjoany.dsouza98@gmail.com mishra.ved@gmail.com

Abstract: **As technology is growing every day, the need for the technology is also becoming essential in every field. The amount of data generated by the healthcare industry is becoming tough to manage and to examine it in efficient manner for future use. In the healthcare field, massive amount of data is generated, from individual patient's information to health history, clinical data and genetic data. The analysis of patient's data is becoming more important, to evaluate the medical condition of patient and to prevent and take precautions for future. With the help of technology and computerized automation of machines, data can be analyzed in more efficient manner. Managing the huge volume of data has many problems interrelated to data security, data integrity and inconsistency. Process mining and data mining techniques have opened a new access for diagnosis of disease. Similarly, to provide effective treatment for a disease's triennial prevention, data mining can be used. Big data mining can aid in analyzing medical operation indicators of hospitals for a period to help hospital administrators provide data support for medical decision-making. In this manuscript, the various applications of big data mining techniques have been analyzed to improve the healthcare systems.**

Keywords: **Big Data; Security; Healthcare; Big data mining; Medical big data.**

I. INTRODUCTION

In the medical field, huge amount of data is generated, from patient's personal information to medical history, from genetic data to clinical data. This medical big data generated is to be stored, not only for the sake of storing, but contains valuable information. This information, when broke down and methodized appropriately, can help in comprehension the 'ideas' of sickness and wellbeing and consequently achieve significant leaps forward in the restorative field particularly in the regions of illness analysis and aversion. With the aid of computers and technology, this medical data can be examined quicker in a less lumbering way. With enormous measure of information like this, we trust that we can make significant and dependable inferences in regard to wellbeing of a man. On the off chance that this supposition is valid and the outcomes are solid, this could be the start of the outright aversion or even annihilation of ailments. Medical big data can likewise be utilized to medicinal services and medication regulatory purposes. Dealing with huge amount of data has many issues related to data integrity, security and inconsistency [10 - 11]. Data mining methods and their applications in the medical field is a new concept although data mining methods have been applied in other fields for quite a while. Therefore, we face issues of practicality. Also, if the medical assumptions deduced from the data is wrong, all the work would be futile. Therefore, science and technology must go hand in hand.

II. MEDICAL DATA AS BIG DATA

A. How Medical Data Considered as Big

Big data is defined by its large volume, velocity, variety and veracity. The data recorded in the medical field is large in terms of volume as in large amount of patient details stored, high velocity in terms of large amount of data coming in high speeds

such as constant monitoring of patient's condition, big variety in terms of large number of varying datasets such as medical data of different age groups, high veracity in terms of incomplete patient records etc.

Figure 1 shows the composition of medical big data.

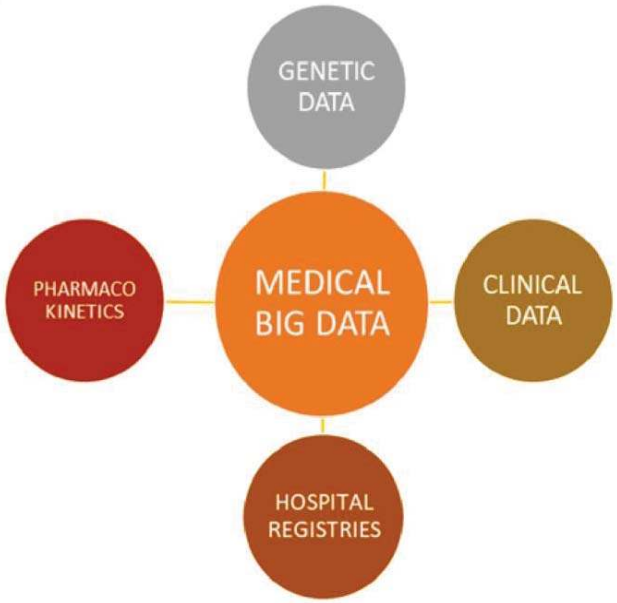

Fig. 1. Composition of medical big data.

B. How is Medical Big Data Different

But medical big data has slightly different features compared to big data of other fields. Medical big data are hard to frequently access, medical big data is somewhat structured in

978-1-5386-4692-2/18/$31.00 ©2018 IEEE

comparison and the use of medical big data has legal complications and issues associated with it [1].

Basically, this medical big data comprises of data on human genetics, medical imaging, pathogen genomics, routine clinical documentation, pharmacokinetics, digital epidemiology, course assessment etc.

III. BENEFITSOFMEDICALBIGDATA

Medical big data can be used to improve healthcare quality, predict epidemics, increasing analytical abilities, cure disease, build better health profiles, improve quality of life, improving outcome, avoid preventable deaths, build better predictive models and reducing resource wastage [12-16]. Big data can be used in understanding the biology of a disease by integrating the available large volumes of data to build meaningful relational models. Medicinal big data drives the progressions behind models of treatment. So with this kind of technology, we can understand so much about a patient, information as early in their life as possible, collecting warning signs of serious diseases at an early stage for a faster and cheaper treatment. Big data analysis in the medical field will ensure that even the smallest of detail will be taken into consideration. Following are the top benefits of big data in healthcare field [8- 9]:

- Facility performance optimization.
- Energy cost reduction.
- Ease of accessibility of information.
- Real time update and alerts.
- Proactive maintenance of equipment's.
- Reducing the healthcare delivery cost.
- Reducing the costs of research and growth.

IV. APPLICATIONOFBIGDATAMININGINMEDICAL FIELD

In the field of business and marketing, the application of data mining has been implemented and may be ahead of healthcare. But this is not the case now. Effective mining applications have been actualized in the medical field, some of which are depicted beneath.

A. Identifying Health Risks.

With the assistance of medical big data and robust mining methods and model building solution, we can identify patients with high risk health condition. This information can be bridled by doctors and medical staffs to identify the condition, so they can take steps to improve quality of healthcare and to prevent health problems in the future [2]. For example, Cancer is a serious illness which can be prevented and cured with the help of big data analytics. Cancer is quickly devastating individuals over the world. Big data can battle disease all the more viably. Healthcare suppliers will have upgraded capacity to recognize and analyze infections in their beginning periods, relegating more adequate treatments in view of a patient's genetic makeup, and direct medication measurements to limit symptoms and enhance viability. It will likewise fantastic help to parallelization and help in mapping the 3 billion DNA base sets [3].

Figure 2 shows the steps of identifying health risk using big data mining.

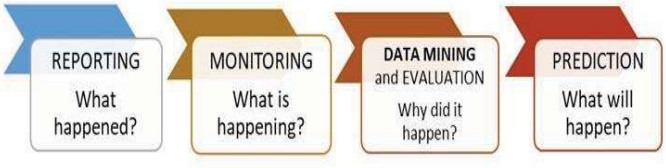

Fig. 2. Steps of identifying health risk using big data mining.(source: McKinsey Big Data Value Demonstration team)

B. Outbreaks and Early Recognition of Epidemics Require Constant Surveillance

To identify high risk patients, possible cases and deviation detection in the happening of predefined events, we can use the aid of computer-assisted surveillance research. Using the medical big data already in our hands, we can use powerful mining tools to deduce patterns and correlations to understand the health behaviour of an area. The surveillance system that uses data mining techniques to identify new and interesting patterns in infection control data. The system also associatively use rules culture and patient care data obtained from the information in the management systems of the lab and generates monthly patterns that are reviewed by an expert in infection control. Developers of the system conclude enhancing infection control with the data mining system is more sensitive than traditional infection control surveillance, and significantly more specific. This has been developed and implemented in the University of Alabama. [2].

Figure 3 shows the method of surveillance.

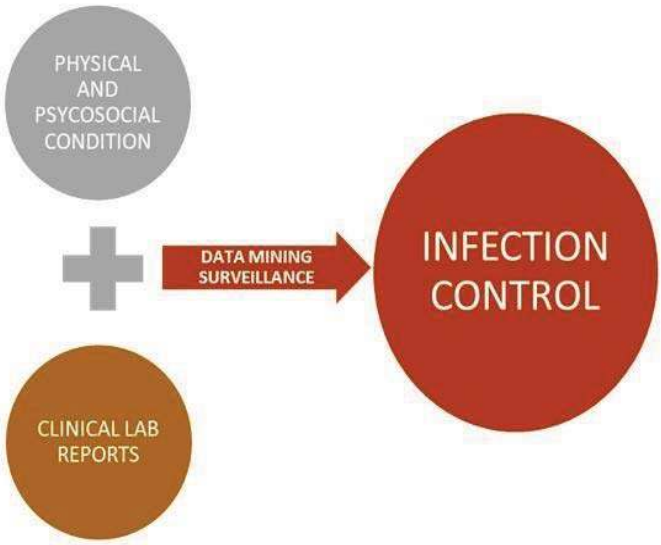

Fig. 3. Method of surveillance

C. Patient Monitoring

The utilization of big data makes the work lesser for doctor's and staff and enables them to work more proficiently. Sensors are utilized and set up near the beds of patients to persistently screen pulse, blood pressure and respiratory rate. Any change is detected immediately, and the staffs are alerted. One way in which big data can be used to aid in monitoring patient's vitals is: Electronic Health Records (EHR)- Big data analytics is widely used in EHR. Significant patient information like clinical history, lab reports and other relevant statistics among others. The information gathered in these electronic records, empower specialists to accurately deduce the patient history and encourage better and convenient conveyance of the human services administrations. Another significant preferred standpoint of using big data is that now patients can, without much of a stretch, monitor their medicines and lab tests. In this way, by coordinating the EHRs crosswise over different restorative offices, patients can reduce the frequency of hospital visits significantly [4].

D. Telemedicine

Digitization, cell phones, remote gadgets, and online video gatherings have set the ball moving for conveyance of remote clinical administrations. Telemedicine is of great importance and use today, all because of these mechanical advancements. The information gathered from these gadgets can be effortlessly shared which makes diagnosis a considerable measure simpler. Aside from remote patient checking, big data additionally helps in foreseeing intense medicinal conditions and forestall disintegration of patient's conditions. Another significant manner by which big data has changed telemedicine is by giving ongoing information which operates from remote areas utilizing robots [5].

E. Informed Strategic Planning and Predictive Analytics

Quick and smart choices should be taken with regards to treating patients experiencing various complex conditions. The part that big data plays here turns out to be significantly more essential in this specific circumstance. Big data gives bits of knowledge which enable specialists to settle on educated choices and enhance the general treatment process. These bits of knowledge additionally help in prescient examination, as it ends up plainly simpler to anticipate which patient is at the danger of diabetes or cardiovascular illnesses.

F. Informed Strategic Planning and Predictive Analytics

Healthcare fraud is a national issue, predominant in government and state, and private protection programs. While the most recent decade, human services misrepresentation has soar with billions of dollars being paid on improper claims. The National Healthcare Anti-Fraud Association moderately assesses that three percent of all human services spending, or $60 billion, is lost to medicinal services misrepresentation. Different appraisals put this number nearer to $200 billion. [5].

V. CHALLENGES

Although big data applications are a major break-through in the medical field, but there are a few challenges that need to be overcome.

A. Issues Related to Data Structure

Big data applications ought to be easy to understand, straightforward, and menu-driven. The greater part of information in medicinal services is unstructured, for example, from natural language processing, it is frequently divided, scattered, and very little standardized. The EHRs don't share well crosswise over authoritative lines, yet with unstructured information, even inside a similar association, unstructured information is hard to total and investigate. Medical big data is more heterogeneous compared to other big data of other fields. Big data will also need to solve metadata transparency issues [6].

B. Issues Related to Security

There are extensive security concerns in regard to the utilization of big data utilization, particularly in medicinal services given the institution of Health Insurance Portability and Accountability Act (HIPAA) enactment. Information that is made accessible on open source is uninhibitedly accessible and, thus, very less secure. Further, because of the affectability of medical data, there are noteworthy concerns identified with discretion. Additionally, this data is centralized, and in that capacity, it is very defenseless against attacks. Hence, empowering protection and security is critical [6].

Fig 4 illustrates the most frequent types of cybercrime in healthcare sector [15].

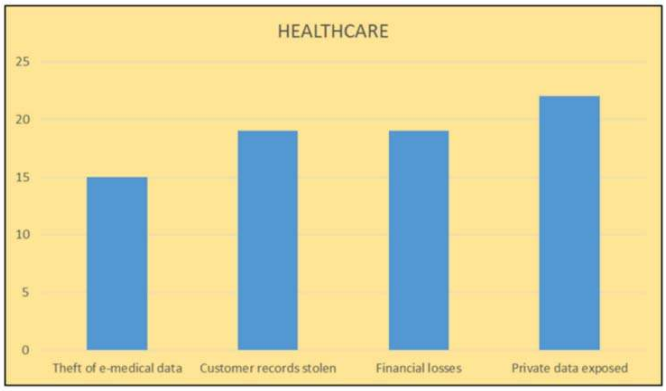

Fig. 4. The most frequent types of cybercrime in healthcare sector

In spite of the fact that assailants will dependably look for approaches to utilize recently created developments and set up stages against healthcare industry to impact people and associations from turning into the cause all their own problems.

The following table shows the top ten healthcare data breaches as given by the U.S. Department of Health and Human Services Office for Civil Rights [18].

1.	Anthem Blue Cross	78.8 Million Affected	January 2015
2.	Premera Blue Cross	11+ Million Affected	January 2015
3.	Excellus BlueCross BlueShield	10+ Million Affected	September 2015
4.	TRICARE	4.9 Million Affected	September 2011
5.	University of California, Los Angeles Health	4.5 Million	July 2015
6.	Community Health Systems	4.5 Million Affected	April-June 2014
7.	Advocate Health Care	4.03 Million Affected	August 2013
8	Medical Informatics Engineering	3.9 Million Affected	July 2015
9.	Banner Health	3.62 Million Affected	August 2016
10.	NewKirk Products	3.47 Million Affected	August 2016

C. Issues Related to Standardisation of Data

Although the EHRs share information inside a similar association, intra-authoritative, EHR stages are divided, best case scenario. Data is put away in groups that are not compatible with all applications and innovations. Issues are caused in exchange of that data, in addition due to the absence of standardization of data. It confuses data obtaining and purging. Restricted interoperability represents a huge challenge for big data due to rare standardization of data. This leaves big data to confront issues identified with the procurement and purging of information into an institutionalized organization to empower examination and worldwide sharing. With globalization of information, big data should manage an assortment of guidelines, hindrances of dialect, and distinctive wordings.

D. Issues Related to Storage and Transfer

When data is created, the expenses related with securing and storing them is high in contrast to the cost of generation of data. Expenses are additionally brought about with exchanging data starting with one place then onto the next and in addition breaking down it. Research has shown we have possessed the capacity to join the topics of Data structure and Storage and exchanges when they represent how organized information can be effortlessly put away, questioned, broke down, etcetera, however unstructured information isn't as effectively controlled. Cloud-based medical data innovation has the extra layer of security related with the extraction, change, and stacking of patient-related information. The utilization of big data should deliver issues identified with expanded uses and in addition the transmittance of secure or unreliable data.

Figure 5 summarizes the challenges.

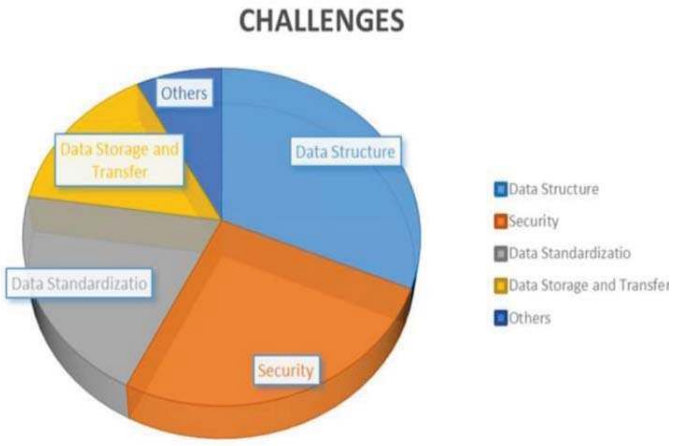

Fig. 5. Challenges to big data

VI. FUTUREAPPLICATIONS

By utilizing big data appropriately and conveying it in setting of clients work processes, specialists will have the capacity to, for instance, figure out who is in danger for illnesses like diabetes or certain kinds of disease and give preventive care.

There is additionally the possibility to fundamentally change:

- Insurance reimbursement models – what are the right approaches to reasonably repay doctors. [17].
- Patient satisfaction metrics – what characterizes better care?

Using the vast medical data and suitable data mining tools and methods, we can better the overall performance of health care services. [17].

- Risk identification and mitigation – how to bring down settlement rates
- Staffing and arranging
- Medication and antibody advancement

Organizations, showcasing, and even political crusades have been utilizing scientific devices to take a gander at this immense measure of data in new and profitable routes, running from identifying extortion to all the more adequately spending publicizing dollars. Medicinal services has customarily fallen behind these different parts because of worries around understanding classification, however that may change soon.

In the medical field, this could mean better alternatives to treat, forestall, and foresee ailments or better distinguish patients in danger. For instance, unique medications may work with changing levels of viability on various populaces of patients. Giving access to these instruments that use more than the information and are implanted inside a doctor's work process can help encourage better choices. A considerable measure of these achievements are as of now happening, yet it is extremely just be starting. What's more, soon enough big data will simply be as common as normal data.

VII. CONCLUSION

In the medical field, enormous quantity of data is created, from patient's personal data to health history, from hereditary information to clinical information. This medical big data contains profitable data. This information, when investigated and methodized legitimately, can help in the comprehension of the concepts of ailment and wellbeing and hence realize significant leaps forward in the medical field particularly in the territories of diagnosis and disease prevention. There are many useful applications that have been implemented and many more potential applications in this field like disease diagnosis, disease detection, infection control, telemedicine, fraud prevention etc. But also, there are hindrances that need to be overcome such as the issues related to data structure, security, data standardization, data storage and transfer. With the use of technology and potential data mining tools, big data can be investigated quicker in a less bulky way. With colossal measure of data like this, we trust that we can reach important and solid inferences in regard to wellbeing of a man. On the off chance that this presumption is valid, and the outcomes are dependable, this could be the start of the total aversion or even annihilation of illnesses.

REFERENCES

[1] Lee, H, C & Yoon, Hyung-Jin., Medical big data: promise and challenges. Retrieved from World Wide Web, 2017.

[2] Greely, T, H &Kulynych, J., Clinical genomics, big data, and electronic medical records: reconciling patient rights with research when privacy and science collide, Retrieved from World Wide Web, 2017.

[3] Obenshain, K, M., Application of Data Mining Techniques to Healthcare Data, Retrieved from World Wide Web, 2011.

[4] Wang, L & Alexander, C, A., Big Data in Medical Applications and Health Care, Retrieved from World Wide Web, 2015.

[5] Eysenbach, G., Challenges and Opportunities of Big Data in Health Care: A Systematic Review, Retrieved from World Wide Web, 2016.

[6] Mirdamadi, A., The Future of Big Data in Healthcare, Retrieved from World Wide Web, 2015.

[7] Schaeffer, C., Haque, A., Booton, L., Halleck, J. &Coustasse, A., "Big Data Management in United States Hospitals: Benefits and Barriers." In J. Sanchez (Ed.), Proceedings of the Business and Health Administration Association Annual Conference, Chicago, IL, 2016, April.

[8] Coustasse, A., Tomblin, S. and Slack, C., Impact of Radio-Frequency Identification (RFID) Technologies on the Hospital Supply Chain: A Literature Review. Perspectives in Health Information Management, 10(Fall): 1d, 1- 16, 2013.

[9] Jee, K., & Kim, G. H., Potentiality of Big Data in The Medical Sector: Focus on How to Reshape the Healthcare System. Healthcare Informatics Research, 19(2), 79-85, 2013.

[10] Wills, M. J., Decisions through Data: Analytics in Healthcare. Journal of Healthcare Management/American College of Healthcare Executives, 59(4), 254-262, 2014.

[11] Coustasse, A., Tomblin, S. and Slack, C., Impact of RadioFrequency Identification (RFID) Technologies on the Hospital Supply Chain: A Literature Review. Perspectives in Health Information Management, 10(Fall): 1d, 1- 16, 2013.

[12] Mishra, V.P., Shukla, B. & Bansal, A. Analysis of alarms to prevent the organizations network in real-time using process mining approach, Cluster Computing pp 1-8, https://doi.org/10.1007/s10586-018-2064-, 2018.

[13] Mishra, V.P., Shukla, B. & Bansal, A. Intelligent Intrusion Detection System with Innovative Data Cleaning Algorithm and Efficient Unique User Identification Algorithm, Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 01-Special Issue, 2018, pp 398-412.

[14] Ved Prakash Mishra &Balvinder Shukla, "Process Mining in Intrusion Detection – The need of current digital world", Springer Nature Singapore Pte Ltd. 2017: CCIS 712, pp. 238–246, 2017.

[15] V P Mishra, Yogesh Waran, Subheshree, "Detecting Attacks Using Big Data with Process Mining", International Journal of System Modeling and Simulation, v. 2, n. 2, June 2017, p. 5-7, 2017.

[16] Mirdamadi, A., The Future of Big Data in Healthcare, Retrieved from World Wide Web, 2015.

[17] Lord, N., Top 10 Biggest healthcare dat breaches of all time.Retrieved from World Wide Web, 2017.