Effect of Machine Learning in Healthcare Industry with reference to Artificial Intelligence

Prof. Mallieswari R, Tejaswini N, Sai Sowmya Karanam, Vasavi Pai P

Abstract –The Healthcare industry has seen immense amount of growth in recent times. It can be either in introducing new technology to detect diseases or new and efficient ways of treating a disease. Although the developments in this field differs, one parameter can be found in common viz. data. This data in healthcare typically is heterogenous, diverse, redundant and incomplete.

The rapidly growing field of bigdata analytics has helped in understanding and improving the healthcare system. This opportunity of the majority of data being present in the healthcare sector can be tapped using both prescriptive and predictive analytics. Implementation of these areas of study can give us more insights on the industry, and understand its needs. This will also aid us in improving the steps that are carried out while processing the query of a patient.

The output generated, by this process using the available or new data will assist the specialists or the patients in making decisions. Also, the output data can be used as an input to the 'intelligent' systems, leading to make the systems more capable of decision making. The system can become more efficient in terms of extracting right information about the patients at the right time. Also, this process has less room for human error due to which high precision can be expected. The analysis of the data provides us with more time and resources in hand for decision making. The decision made by the experts again converts itself as input for the system and starts the process of Machine Learning.

In this study we focus on how Artificial Intelligence,Machine Learning in particular can assist in improving the Healthcare Industry.

Keywords: Artificial Intelligence, Healthcare, Information,Machine Learning models.

I. INTRODUCTION

Healthcare systems are one of the complex systems to consider.This system is highly dynamic as there are new advent of diseases or technologies to tackle them.

Drug discoveries have reached new heights due to the accessibility to data, either of new formed genes that can combat diseases or to the technology that can aid the experts in the field of medicine to implement it in their practice. Both ways there have been major breakthroughs in the field of medicine.

One of the most extensively used concept in the field of medicine in recent times is Artificial Intelligence. In particular Machine Learning is a more advanced technique that can be used in this industry. Particularly because of the data which is present in the system.

Healthcare does consist of huge amounts of data. But the problem lies in the quality of data. The data available in this sector is varied, diverse, redundant and incomplete. Decisions become difficult in such cases. The main challenge is to obtain the right data, in right form to take decisions or to move forward.

This can be facilitated by Machine Learning. Although queries or diseases and their treatments are different, there is always a common ground where an "intelligent" system can take up the work. For example, lab reports. It is always advised that lab reports have to be consulted with a doctor. This is the second stage of the patient's service. If this part can be taken up by a system that can detect deviation of a particular parameter from the normal. This will aid in two aspects. The doctor utilization is reduced for the same patient and also a database is created for the diagnosis that is provided.

In a similar way, there are innumerable small instances in healthcare that can be transferred to an intelligent system. This will make the diagnosis much easier. This will create a database required for future reference. This will also aid in situations where there is a change in hospital or the specialist.

Most of the components of the healthcare industry are expensive and resource intensive.From the repots generated from the labs to the Operating Rooms or Intensive Care Units. Therefore, if these resources are not used in an effective manner, the capability of the system is not utilized completely.

This opportunity can be exploited with the use of Machine Learning. There are various parameters or Key Performance Indicators(KPIs) by which the implementation of machine learning in the system can be measured. Some of them are[3]:

- Prediction of re-admission risk
- Treatment efficacy
- Monitoring Adverse Drug Events (ADEs)
- Risk Assessment
- Tracking hospital operations

- Prediction of various diseases or an epidemic breakout
- Absenteeism and its cost to the hospitals
- Tailoring programs

Monitoring these parameters will help the sector improve on its service. This can be done by making an effective use of data.

This study has been broken down into four elements:

- Reasons for machine learning to be implemented in healthcare industry.
- Methodologies used
- Challenges faced by the industry
- Recommendations

II.   REASONS FOR MACHINE LEARNING TO BE IMPLEMENTED IN HEALTHCARE INDUSTRY

Several issues can be identified in the healthcare system that leads to inefficiency. There are enormous amounts of opportunities that can be tapped by the use of Artificial Intelligence, Machine Learning in particular. Some of the reasons are stated below.

A. Complex systems with unique diseases

Healthcare systems are characteristic of various diseases with abundant symptoms, many of which are common to a lot of diseases. This makes the system complex and therefore its diagnosis. Hence while implementing an intelligent system, this parameter has to be taken care of.

The data will have extremes and can be categorized as "outliers" that will provide a miscalculation in the system.

B. Detection of a disease is patient dependent

The initial step of detection has always been from the side of the patient. There exists no system that provides a tracking mechanism to alert the patient of the discrepancies in their health. Ignorance of the patient leads to severity of the condition. Only at this time the patient calls for a consultation.

There is no system that tracks or monitors the health of the patient and indicates the errors. This can either be before hospitalization or after the discharge.

C. Patient monitoring

Hospitals or doctors, after the discharge of a patient, have no track of the patient's well-being. This can increase the risk of re-admission of the patient. Hence the patient monitoring is highly important.

D. Adverse Drug Events (ADEs)[4]

Adverse Drug Events refer to the medication errors that occurs due to lack of information. About 5% of the readmission is due to ADEs. This means that there is no right information about the patient or about the disease. As we come across highly skilled doctors, the mistake lies with the patients. They miss out on symptoms that they think is insignificant but are actually crucial to the query.

ADEs include, incorrect dosage of medications, allergic reactions and adverse drug reactions. This information has to be tracked, either from database or through the patient's behavior.

E. No consistent patient database[1]

There is no presence of a universal database of a particular patient. This is due to theshifting behaviour of the patients. Also, one of the major factors affecting the non-availability of data is the habit of self-medication. This, prevailing significantly among patients, results in information not being available about the previous medication or dosage.

These reasons give rise to ineffective treatments. Tracking and monitoring is one of the significant challenges in healthcare due to data unavailability. This can be tackled in the following methods.

III.   METHODOLOGY

The basic methodology that can be used to implement Machine Learning in healthcare begins with the formation and maintenance of the right database. This forms the framework for all other activities.

The activities can either be carried out in prescriptive or predictive manner based on the situation.

A. Prescriptive Analytics

Prescriptive analytics can be carried out during the analysis of the course of action during a checkup. This requires the specialist to look for previous history of a particular patient or a disease and then determine the right course of action for the same.

The database access for the specialists as well as for the patients helps in providing the right service for the patients. The databases act as the input to various calculation engines that matches the results with the standard and provides an output to the data users.

This process can act as an input to the machine learning algorithms that helps the intelligent system to learn on its own.

In later stages the results obtained will be highly accurate and precise for the patients or doctors to act upon. This process can be represented as follows:
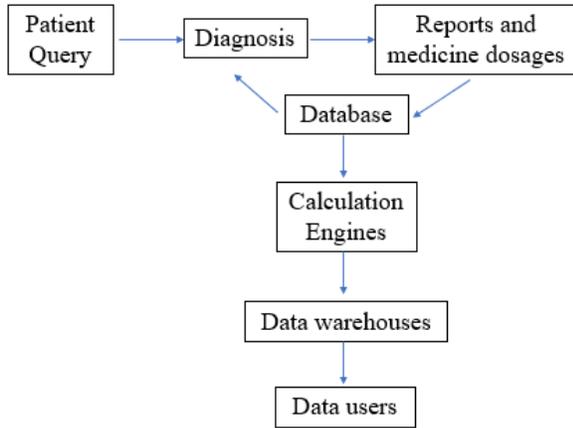
B. Predictive Analytics

This consists of predicting a risk. It can be either a risk of admission of a patient or a risk of re-admission after discharge.

Readmission of a patient takes place due to a fault in either the specialist's end or by the patient's end. From the point of view of a specialist, more probable case is the Adverse Drug Events (ADEs). This refers to the wrong dosage or the medicine prescribed.

Also, machine learning aids in understanding the traits that can cause an epidemic breakout. It can monitor the health of a person and mention the symptoms that the patient considers irrelevant. This will provide the doctor with a better understanding of the problem and hence giving a better diagnosis.Again, this will act as an input to machine learning algorithms.

Here, there is a model created for the patient's profile and then applied to all other cases to predict a particular condition. This is represented as follows:
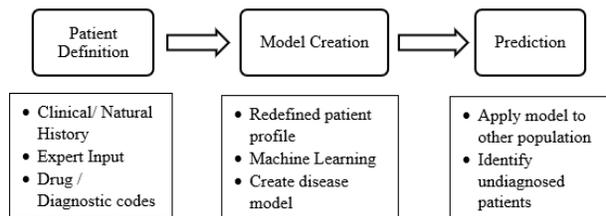
IV. CHALLENGES FACED

The area in which significant challenges are faced during the implementation of Machine Learning in Healthcare industry is in the data infrastructure. Data present in healthcare is characteristic of highly fragmented, noisy, high dimensional and heterogenous. Major obstacles to the fulfillment of this implementation are the noise, missing data and the nomenclature. Some of the common challenges faced are mentioned below.

A. Data volume

The data obtained in healthcare is highly diverse and complex. This makes the modelling very difficult. Also, since most of the people do not have an access to primary healthcare, availability of data also becomes difficult. Hence, we will not get as many patients as we want for the machine learning process. This can hinder the process.

B. Data Quality

The data available in healthcare is completely unstructured, heterogenous, ambiguous and incomplete. This provides us with problems such as data sparsity, redundancy and missing values.

C. Temporality

The diseases are always changing their form over time. This can be due to the resistance the disease develops or due to the advancement in medical sciences. Hence the filed is completely volatile. This changing behaviour becomes difficult to model. This will require a tremendous amount of inputs to create a machine learning algorithm[7].

D. Noise

Noise is the unrelated data that is present in the database or the warehouse. It can be incorrect data collection process, incorrect nomenclature, data duplication or conflicting records. It can cause the formation of "outliers" due to the presence of natural variations of the individuals[7]. This can cause an issue in modelling the data.

E. Missing data

Missing data is one of the biggest unavoidable problem. This can either be in data collection or measurement stage. In any stage of a health query there exists ambiguity about the condition of the patient, resulting in a missing data[6]. Further, non-existent data can also occur due to the fragmentation of data collection sources. Healthcare data is available not only on one hospital or one center, but in varied sources like super specialty hospitals, day care clinics etc. Not tracking this data will lead to missing values.

F. Medical Nomenclature

Another challenge the implementation possess is not from the quality but the representation. Referring to the nomenclature of the data input present, there has been a lot of coding schemes that have been developed. This implies that same attributes have been mentioned in a different way in different situations. This will cause a

12

data to be taken as heterogenous when actually it is the same data entered[6].

V. CONCLUSION

This study focuses on the implementation of Machine Learning in Healthcare, the methodology, its challenges faced during implementation. This conceptual framework can act as basis for the implementation.

This concept involves the merger of two completely different fields, technology and medicine. Thereby, the presence of the necessary resources is completely essential. Focusing on the data infrastructure, the human skills and capabilities and governance systems for sustenance of the implementation in unavoidable.

VI. RECOMMENDATIONS

Increasing the resource utilization is definitely feasible by the implementation of Machine Learning in healthcare industry. The main changes have to be brought about by considering three aspects in the system i.e. Data Infrastructure, Human skills and capabilities requirements and Governance.

Data Infrastructure consists of the architecture by which the data moves from various databases to the users. Traditional method consists of maintaining the database for each function. This will provide less accessibility to the patients' information. Instead using a pooled data called a Data Lake or a Metadata Dictionary will help in having a better access to the patients' information[6].

Medical Experts will need aid in coding the software with Machine Learning for their use. Hence, on the people level there are three profiles of people that are additionally required. A Data Scientist, who can build the mathematical algorithms, an Analytics Engineer to make the software more efficient and finally an analytics translator that can understand a business problem and translate it into a technical language and vice versa[6].

Governance systems refer to the check placed on the manner in which the data moves from the database to the data users. Data from various data bases can be pooled in a data lake. From this Metadata Dictionary, the data can move to various calculation engines according to the requirement. Later, to the warehouse for the basis for Machine Learning. This eliminates the need for separate databases for various functions and hence reducing redundancy[4].

Application of these three stages transforms the complete healthcare system into more user friendly and data driven system. This reduces a major portion of idle time for all resources as well as the patients, making the system highly productive. Higher number of cases can be considered under this situation.

Further points to be considered during the implementation are –

1. Healthcare surveillance should be conducted to understand the needs of the system
2. Maintaining a transparency of the medical data provides higher accessibility for the users, both the patients and the end users.
3. Formation of online platforms and communities enables the exchange of various information and also helps the healthcare industry to track and monitor the patients.

REFERENCES

[1] Keith Feldman, Louis Faust, Xian Wu, Chao Huang and Nitesh V. Chawla, "The impact of complex Healthcare data on the Machine Learning Pipeline", (2017), Journal of Medicine Systems.
[2] David A. Kvancz, Marcus N. Sredzinski, and Celynda G. Tadlock, "Predictive Analytics: A Case Study in Machine-Learning and Claims Databases", (December 2016), The American Journal of Pharmacy Benefits.
[3] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang and Joel T. Dudley, "Deep learning for healthcare: review, opportunities and challenges", (2017), Briefings in Bioinformatics.
[4] Rubenfeld GD, Angus DC, Pinsky, "Outcomes research in critical care" American Journal of Respiratory and Critical Care Medicine.
[5] Hogan WR &Wagner MM, "Accuracy of data in computer-based patient records", Journal of the American Medical Informatics Association.
[6] Liao, Lee, "Appropriate medical data categorization for data mining classificationtechniques". MED. INFORM., 2002.
[7] Sitar-Taut, "Using machine learning algorithms in cardiovascular disease risk evaluation." Journal of Applied Computer Science & Mathematics, 2009.

Prof. Mallieswari R,MSc., MPhil, MBA from Bharathiar University, Coimbatore is a Gold Medalist in Applied Statistics at the Master level. Over 14 Years of teaching experience in the area of Applied Mathematics and Statistics at PG Level – Management area. Research interests are in the field of Statistical Quality Control and Operations Research. Current research focuses on Advance Acceptance Sampling (SQC) in Industrial Operations.
Tejaswini N, Student, Ramaiah Institute of Management
Sai Sowmya Karanam,Student, Ramaiah Institute of Management
Vasavi Pai P,Student, Ramaiah Institute of Management