

A Survey on Applications of Big Data and Machine Learning in Healthcare Analytics

M.Janardhan¹, R.Manvitereddy²

^{1,2}Dept Of Computer Science And Engineering

^{1,2}G.Pullaiah College Of Engineering And Technology

Abstract- In healthcare sector a massive amount of structured or unstructured data is generated from doctor's notes, reports and treatments. This generated data from healthcare management system is analyzed using cloud enabled platform which uses big data analytics. Various data is collected and correlation analysis is performed on this data. The other advantage of using big data in the field of medicine is that we can detect the diseases beforehand that allow doctors to give treatment proactively. Fraud rate can also be reduced. The main purpose of this work is to describe the importance of data analytics and machine learning techniques to analyze the diseases and the patients are treated more effectively.

I. INTRODUCTION

Big data is a procedure of studying large amount of variety of the data that helps organizations to make more informed decisions. The way we analyze and leverage the data in an industry has been changed using big data. Big data can be used in healthcare analytics which can reduce treatment cost, estimate the outbreak of epidemics, and avoid preventable diseases. As the world population is increasing day by day the human lifespan is also increasing which introduces new challenges to treatment methods. In this paper we would like to address a use of big data analytics in healthcare: how can big data help? This section will answer such a question, and it clearly describes the part of big data in future health care.

Machine learning is a category of AI which plays a major role in predicting outcomes by automatically learning and improving from experience without programming explicitly. This technique can be applied in various ways like health care by developing algorithms which can identify complex patterns. The intelligent data-driven decisions can be made by this technique. Machine learning algorithms can be implemented either by supervised or unsupervised algorithms. Supervised learning algorithm use a set of trained examples that consists a pair of input object and desired output value. This algorithm produces a function which can map new examples. Supervised machine learning techniques are more convenient for medical data classification [1]. Unsupervised

algorithm uses the data set that is neither classified nor labeled, and generates a hypothesis that predicts the outcome.

II. RELATED WORK

Healthcare analytics collect massive amount of data and use the best strategies in treatment of people. A set of clinical studies are given as examples, machine learning is used to produce description for these cases. The large volumes of data in varied forms are digitized to manage the huge data in healthcare industry. The IBM Senior Manager of Intelligent Information systems research, John Smith has said that one of the applications is used for detecting melanoma. Breast cancer can be predicted from a data set utilizing the data analytic techniques of Big Data along with Map Reduce.

The computer is fed with different images of diseases and teaches the system how to identify the disease by recognizing the features and also give the doctor with treatment procedures. A research done by Berkman and Kawachi in 2000 has also documented associations between cultural factors, social factors and health [2].

III. FEATURES OF BIG DATA IN HEALTHCARE

Doug Laney in 2001 has defined big data as "the 3V's" which are Volume (large amounts of historical and new data), variety (text or multimedia, structured or unstructured data) and Velocity (how fast data can be used). In addition to that other analysts also defined big data as Veracity, Variability and also Complexity. The increase in velocity and volume is known by most of us. But the crucial one is specialization. The opportunity to use the common hardware is achieved by big data technology.

IV. KINDS OF BIG DATA

Big data analytics can be classified as following four types which should be considered at big data analysis.

4.1 PRESCRIPTIVE

Prescriptive analysis is a analysis which displays the action to be done and also give recommendations for further steps, and allowing us to answer business kind of questions. This prescriptive analysis can decrease time for marketing the new medicines and can detect the right patients for clinical treatments. This also helps for drug development.

4.2 PREDICTIVE

Predictive analysis examines and reads the past data patterns and predict the outcomes for present problem or situation. By analyzing the past and historical data, predictive analysis can foresee what happens in the future. Sentiment analysis is one of the examples for this analysis which predicts the sentiment score based on the examined data. This analysis is useful in developing effective customer relationship management.

4.3 DIAGNOSTIC

This is a kind of analysis which is mainly applicable in businesses. It completely analyses the root cause and reveal the patterns in their business. The factors that affect the business directly and indirectly are identified. The decisions must be made smarter at the results of this analysis for better growth in the business.

4.4 DESCRIPTIVE

Descriptive analysis studies what happens in the real time by considering the present data. This analysis is referred as simplest type, as it converts big data to bit sized nuggets that are useful [3].

V. USING BIG DATA ANALYTICS IN HEALTHCARE SECTORS

The ways of organizing their healthcare clinical facilities can be potentially transformed by big data analytics. These techniques can enhance care and save patient's life, and provide health facilities at lower charges. Data analytics can improve the standards of healthcare in the following areas.

5.1 PUBLIC HEALTH

By studying the patterns of different diseases and recording their outcomes, the public health can be progressed by using analytics technique. Massive data is used to identify the requirements, predict, prevent and provide the required services. This is also helpful in predicting the future and taking precautions against the diseases.

5.2 EMR (ELECTRONIC MEDICAL RECORD)

An EMR consists of records of patient's medical history. The structured and unstructured medical data is examined by any of the data analytics methods and provide right treatments for the patients at right time.

5.3 PATIENT PROFILE ANALYTICS

By observing the lifestyle changes of the patient's, proactive analytics can be used for better care and treatment of the patient.

5.4 FRAUD DETECTION

This method analyzes large number of claim requests to decrease the fraud rate. The abuse, fraud and waste can be deducted by effective use of analytics methods.

5.5 SAFETY MONITORING

For analysis of real time huge amounts of fast coming data in hospitals we use the above discussed data analytic methods. This approach helps in negative event prediction and for safety monitoring [4].

5.6 PET and MRI

PET is abbreviated as Positron Emission Tomography. It can estimate both the amyloid and cerebral metabolism. The early detection of diseases is done easily by using algorithms that identify the specific elements of images for medical purposes.

5.7 PATIENT CARE IN RADIOLOGY

Big data aids radiology in decision support systems. This helps the radiologists to help them in studying images for decision support.

5.8 PEDIATRIC CARDIOLOGY

The Pediatric Cardiac Critical Care Consortium uses big data to enhance the quality of healthcare by gathering the data by clinical reports and outputs from each patient's medical record and analyze the data to provide doctors with a performance feedback [6].

5.9 VIRTUAL CARE AND WEARABLE HEALTHCARE TECHNOLOGIES

It raises the healthcare quality. This technology helps people to access virtually their health data, but a question rises. How secure is the data which is unprotected?

5.10 PROJECT DATASPHERE

Statistical Analysis System is a major part of developing transparency of industry wide pharmaceutical data. The aim of this is to provide globally accessible secure data and an environment for analysis where different organizations can share clinical trial information. This improves the scientists to quickly enhance in care [5].

VI. ADVANTAGES OF MACHINE LEARNING IN MEDICINE

Machine learning is helping doctors in identification and diagnosis of diseases. An algorithm to identify cancerous tumors was developed by Google using Machine learning and AI. Machine learning algorithms consider and study the data and help the physician in healthcare decision making [9]. These techniques are more useful in the fields with huge image datasets, like radiology, pathology and cardiology. The algorithms monitor the images and find out the irregularities where more attention is required. This increases the efficiency and accuracy.

6.1 TYPES OF MACHINE LEARNING ALGORITHMS

Based on the output from the algorithm or the type of input variable given for training the machine, machine learning algorithms can be categorized into four types. They are

- Supervised learning
- Unsupervised learning
- Semi Supervised learning
- Reinforcement learning

6.2 APPLICATIONS OF MACHINE LEARNING IN HEALTHCARE

6.2.1 MEDICAL IMAGE DIAGNOSIS

Machine learning uses computer vision to diagnosis the image. In the year 2010, Microsoft's team, INNER EYE worked on image diagnosis tools for analyzing the image using Deep learning. Diagnosis is a intricate procedure that collect various factors like color of patient's eye and food habits of patient, which cannot be examined by the machine completely. These machines can help the doctors and provide more scientific knowledge to make correct decisions.

Oncology department of Memorial Sloan Kettering (MSK) is focusing on partnership with IBM Watson. It has lumps of information regarding the treatments used for cancer patients. So it helps physicians by implying treatments which worked best in the past. This kind of tool is already used in so many places.

6.2.2 MEDICAL DATA COLLECTION

or more live health data we need to merge data from various devices. The treatment for Parkinson's and Asperger's syndrome can be given through mobile apps where users can evaluate their health condition by facial recognition using machine learning. IBM is integrating with Medtronic to make sense of diabetes and insulin data in real time. In Spite of having enormous amounts of data provided by IOT, the machine learning techniques must use in information to make real time changes to the treatment. By pooling of large data is aids the patients and doctors to deal with tough and rare cases.

6.2.3 DISCOVERY OF DRUG

Application of machine learning in healthcare is preliminary in drug discovery which has various future uses. IBM uses this technique in many of its health applications. Google has also joined with many companies in discovering the drug using machine learning. For multifactorial diseases, we need to find mechanisms and alternative ways for treatment. This is called as Precision medicine. This involves unsupervised learning, identifying the patterns in data rather than prediction [10].

MIT Clinical Machine Learning Group research on precision medicine is to explore algorithms which understand the disease and provide a powerful treatment for Type 2 diseases. Machine learning techniques are implemented by Microsoft's Project Hanover in different ways and it also integrated along with the Knight Cancer Institute which treats cancer with Precision medicine using Artificial Intelligence.

MICROSOFT'S PROJECT HANOVER
The experimentation or manufacturing data is possible to help pharmaceutical manufacturers to decrease the time required for manufacturing the drug and also increase duplication and lower the costs.

6.2.4 ROBOTIC SURGERY

In Robotic surgery technique, da Vinci Robot got a good attention. This device can handle dexterous robotic limbs to carry out surgeries with less vibration and with more efficiency. Some of the robotic surgeries use computer vision

in support with machine learning to recognize any part of body. Machine learning can also be used to fix the motion of robotic limbs [8].

Examples of Machine learning in Robotic Surgery

Most of the applications fall into these categories

- Automation of Suturing
- Machine learning for evaluating Surgical skills
- Machine learning to enhance Surgical Robot materials
- Surgical Workflow Modelling using machine learning

6.2.5 RADIOLOGY AND RADIOTHERAPY

Google's DeepMind Health along with UCLH (University College London Hospital) is working on machine learning algorithms that distinguish between good tissues and harmful tissues that cause cancer. UCLH and Deep Mind are ensuring to increase the accuracy in radiotherapy planning and to increase the speed of segmentation process by applying Machine learning techniques [10].

VII. FUTURESCOPE

AI will be broadly adopted by health sector by 2050's. The first use will be in radiology and pathology departments for image analysis. Over 50% of users trust that Machine learning with AI will be all over in the healthcare by 2050. There is a crucial need for healthcare case studies in machine learning and AI. Fifty percent of companies in US say that healthcare industries need to be convinced of ROI from Machine learning investments. Patients care is increased by Decision Support systems which use ML/AI. Hospital purchases would be decreased and the patients can be charged less.

The government of US has also decided to make huge volumes of data public using big data analytics. Using big data analytics in healthcare will improve efficiency in treating people and also reduces costs and fraud rate. New platforms and algorithms are being developed to help healthcare sectors to leverage the expanding set of data.

REFERENCES

- [1] S. Anto, Dr.S.Chandramathi, Supervised Machine Learning Approaches for Medical Data Set Classification - A Rev IJCST
Vol. 2, Issue 4, Oct . - Dec. 2011
<https://www.ncbi.nlm.nih.gov/books/NBK19924/>
- [2] <http://www.smartshiftech.com/big-data-and-analytics-which-type-analytics-does-your-business-need>

- [3] <http://www.builtinla.com/blog/significant-benefits-big-dataanalytics-healthcare-industry>
- [4] <http://www.informationmanagement.com/news/big-dataanalytics/how-big-data-keep-united-healthcare-nimble-10026746-1.html?>
- [5] <http://www.scimage.com/the-future-of-big>
- [6] [-data/ \[7\]https://www.techemergence.com/machine-learning-in-healthcare-executive-consensus/](https://www.techemergence.com/machine-learning-in-healthcare-executive-consensus/)
- [8] <https://www.techemergence.com/machine-learning-in-surgical-robotics-4-applications/>
- [9] <https://www.healthcatalyst.com/clinical-applications-of-machine-learning-in-healthcare>
- [10] <https://www.techemergence.com/machine-learning-health-care-applications/>