

MACHINE LEARNING AND BIGDATA ANALYTICS IN HEALTH CARE SYSTEM - A SURVEY

J. Arul Valan¹, Dr. E. BabuRaj²,

¹Assistant Professor, ²Professor,
Department of CSE,

¹National Institute of Technology,

²Marian Engineering College,

¹Nagaland, ²Trivandrum, India.

valanmspt@yahoo.co.in,

alanchybabu@gmail.com

May 25, 2018

Abstract

Big Data is now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. With advance research in health sector, there is variety of perishable data available in health care domain. The advances in healthcare system will rapidly enlarge the size of the health records that are accessible electronically. More demand for innovated and intelligent systems in healthcare industry leads to a need for the introduction of automation of processes. All aspects of human life can be rehabilitated by big data and machine learning. In past few years, there has been significant developments in how machine learning can be used in health care sector. The performance of machine learning algorithms has been investigated and to support the problem in health care services. With advance research in health sector, there is variety of perishable data available in health care domain. Big data

analytics refers to the process of collecting, organizing and analyzing large sets of variety of data to discover hidden knowledge and useful information. This paper discusses the potential of utilizing Machine Learning Technologies along with Big Data Tools in Health care systems.

Key Words: Big Data, Machine Learning, Health Care, Electronic Health Record, Challenges, Applications, Big Data Tools.

1 INTRODUCTION

In this modern world, sources of data, such as IoT, social networks, social media and electronic devices harvests huge amount of unstructured data. A data revolution in health industry signifies a radical and exciting shift. Health Care industry includes nursing homes, hospitals, medicals, charitable organizations and health centers. Nowadays health industry facing lots of hurdles because of the inadequacy of medical facilities and with new kinds of diseases. In a health sector, data is generated from diverse sources: wearable devices, annotations, prescriptions, symptoms, food habits, lab reports, health insurance and policy, doctors and nurses information, Electronic Health Records (HER), hospital details and so on. All these data are increasing day by day in the order of Peta bytes [1]. It is impossible for human beings to analyze the enormous amount of data generated in health sector This immense growth of data in health sector has given lot of challenges like data gathering, storage, transmission, processing, analysis and decision making [2]. This enormous amount of unstructured data cannot be processed efficiently by traditional database technology. A key solution to this problem is Big Data. Big data analytics refers to the process of gathering, organizing and analyzing large sets of variety of data to discover hidden knowledge and useful information. Big data tools can handle structured, unstructured and semi-structured data. Machine learning offers a solution to find patterns and association from the data, which enables healthcare professionals to take precise decisions. Hence, if health care is combined efficiently with big data and machine learning, it results in effective storage, processing and analysis which can help to improve decision making process and finding better treatment solutions for present diseases

[3].

2 OVERVIEW OF MACHINE LEARNING IN HEALTH CARE

Machine learning is one of the emerging area because it provides the ability to automatically attain deep insights, identify hidden patterns, and create predictive models from the preprocessed data, all without requiring explicit programming instructions and human interventions. Machine learning algorithms are classified as Supervised, Unsupervised and Reinforcement techniques. Deep learning also one of the machine learning algorithm and it is one of the significant algorithm used in health care sector. Successful application of machine learning in health care depends with intelligent algorithms and rich data sets. Self-driving car and websites that recommend items based on the purchasing decisions of other people are all examples of machine learning being used in the real world [12]. Machine learning is one the valuable and necessary tool for the e-health care system that is used to combine and make sense of health care data. Perfect accurate diagnosis is achieved by Machine learning along with domain experts. The analysis of data is classified into descriptive, diagnostic, predictive, prescriptive and decisive categories. After the preprocessing of data, the analysis may be descriptive or diagnostic. But while we are analyzing present and historical data, the analysis will be predictive, prescriptive and decisive. Preprocessing plays a major role in the machine learning process. Preprocessed data is given as an input to the machine learning algorithm. Recently lots of innovative research going on in the medical field along with big data and machine learning. One phase of machine learning consists of data preprocessing plus learning. Efficient classification and prediction can be attained by proper machine learning algorithms. Models used in machine learning are Classification, Clustering and Regression [12].

3 ROLE OF BIGDATA IN HEALTH SECTOR

Wearable and IoT enabled devices provide us an enormous stream of data about human life and behavior. Along with existing health data, the behavioral data obtainable from these devices may greatly enhance opportunities to predict long-term health conditions. Within the domain of health care, technological advances such as on-line health forums and digital data opening up new opportunities, including: new approaches to diagnosis diseases and patient- and treatment-monitoring, improved surveillance of disease and risk factors, and improved health investigation and disease control. The data about human behavior and belief, together with Electronic Health Records (EHRs) and genomics information, can potentially provide us clear picture about human health. Big data analytics plays a major role in a e-health services [13].

4 BIGDATA TOOLS AND TECHNIQUES

Big Data analytics is bagged with several tools and techniques. The following are the some of the important tools that support health care data storage and processing. **Hadoop:** It is a framework for parallel processing in a distributed environment. HDFS and Map Reduce are the main components of Hadoop. Hadoop Distributed File System (HDFS) is used for storage and parallel processing is taken care by Map-Reduce, which is a programming model. [16].

Spark: Spark is a fast cluster computing technology. It well supports in-memory computing and also offers scalable data analytics platform. It supports open source environment. Spark can handle applications like machine learning algorithms and natural language processing [2].

Storm: It is an open source distributed real-time computation system. It can process unbounded streams of data. It is simple, can be embed with any programming language.

MOA (Massive On-Line Analysis): It is a framework for data stream mining.

Mahout: It is an open source software and explicitly for machine learning and data mining. It can manage and process huge

amount of data.

Flume: It is a tool for data gathering from different web servers to Hadoop. It can handle un-structured data.

Sqoop: It is a tool to exchange data between Hadoop and relational databases. It handles structured data.

Hbase: It is a NoSQL open source database. It offers real-time read/write access to large datasets.

S4: It is a platform for processing continuous stream of data.

5 APPLICATION OF BIGDATA AND MACHINE LEARNING IN HEALTH CARE

E-healthcare industry is being transformed by the advancements in machine learning and Big data analytics. Machine learning and big data are now being used in healthcare to provide superior patient care and has resulted in improved predicted outcomes. Nowadays, machine learning supports personalized health care through improved diagnostics and predictive prescriptive healthcare analytics. Machine learning algorithms process huge structured and unstructured datasets (big data) and provide useful perceptions that allow effective health care services and it can make accurate decisions, significantly improve operating efficiencies, and reducing the unwanted cost. **Monitoring patient vitals:** Sensors are being used in the patient beds to continuously monitor blood pressure, heart-beat and respiratory rate. Any deviations in pattern is immediately reported to doctors and healthcare administrators [7].

Healthcare Intelligence: Machine learning and Big Data are being used in healthcare Intelligence applications [7].

Fraud Prevention and Detection: Machine learning algorithms help to avoid a wide range of human errors on the side of health care administrators in the form of dosage level, medicines, and other kinds of treatments. It will also be used to detect and prevent fraudulent claims of insurance from the insurance agencies [7].

Real-time Alerting: Wearable IoT enabled devices will collect patients health data continuously and send this data to the cloud

setup. This information will be accessed from the cloud databases on the state of health of the general public, which will allow doctors to compare this data along with existing patient record. For example, if patients blood pressure increases, the system will immediately send an alert to the doctor who will then take care the immediate action [8].

Predictive Analytics in Healthcare: One of the key functionality of machine learning algorithm is Predictive Analytics. In health industry, it will save human lives. It will help doctors to make decisions within seconds and improve patients treatment. This is particularly useful in case of patients with complex medical histories, suffering from multiple conditions [8].

Evidence-Based Medicine: It is one of the remarkable application of machine learning. Using evidence-based medicine, the doctor can compare symptoms of a patient to a larger patient database in order to come to an accurate diagnosis faster and more efficiently. Here big data role is assimilating information from different sources and normalizing the data in an accurate manner [9]

Analyzing Hospital Networks: With the help of the recent IoT enabled cloud based technologies we can easily tracking and monitoring the facility, availability of specialist doctors, nearby hospitals and locations of the hospitals [9].

6 CHALLENGES

Data Structure Issues - Different healthcare organizations use different formats of data for storage and processing [10].

Security Issues - One of the important security issue in big data analytics is privacy and attack. Data is collected from diverse of sources and that is made available on open source hence, highly vulnerable to attacks [10].

Data Standardization Issues - Lack of standard form of data for storage [10].

Data sharing - Another important challenge that creates problem. Both the medical organizations and the patients are not ready to share their private data.

Lack of Appropriate Skills - Nowadays, there is an immense growth in the IT sector. So, it is important and necessary that

health care team members should update their knowledge regularly in accordance with their working field.

Inaccuracies (Veracity) - Data is gathered from different sources. Keeping up-to-date information is another challenge in the health sector [17].

Real-Time Analytics (Velocity) - In the health sector continuous flow of stream data is tremendous. It is very different to handle and process the cloud based stream data in real-time. The use of cloud based technologies leads to issues of security and privacy [17].

7 CONCLUSION

IT technologies plays a tremendous part in all the walk of our daily life. The main concern of the human's life is the good health, and proper care during the crucial illness time. The advent of the IoT enabled cloud based technologies made this process very transparent and effective. In today's E-healthcare environment, more than 75% of patient data is unstructured and this data comes from diverse digital sources such as sensors, emails, Electronic Health Records, doctors' and nurses' notes and third-party sources from outside of the hospital. Through these new bid of big data technologies with machine learning, health care providers and administrators are making great jumps in improving patient outcomes. A review of all the tools used in health care is provided in this paper. The applications and hurdles of machine algorithms with big data in the health care industry is briefed in this paper.

References

- [1] Dr.W.Liu and Dr.E.K.Park. Big Data as an E-Health Service. IEEE Electronic, 10 April 2014.
- [2] Prof.Jigna Ashish Patel and Dr.Priyanka Sharma. Big Data for Better Health Planning. IEEE International Conference on Advances in Engineering and Technology Research, August 2014.

- [3] E.P.Ephzibah and R. Sujatha. Big data management with machine learning inscribed by domain knowledge for health care. International Journal of Engineering Technology, 2017.
- [4] Andrew L. Beam and Isaac S. Kohane. Big Data and Machine Learning in Health Care. American Medical Association, 2018.
- [5] D. Wang, X. Liu, and M. Wang. A dt-svm strategy for stock futures prediction with big data. IEEE 16th International Conference on Computational Science and Engineering (CSE). IEEE, 2013.
- [6] S. Suthaharan. Support vector machine. Machine Learning Models and Algorithms for Big Data Classification, Springer, 2016.
- [7] <http://attunelive.com/big-data-applications-healthcare/>
- [8] <https://www.datapine.com/blog/big-data-examples-in-healthcare/>
- [9] <http://www.ingrammicroadvisor.com/data-center/7-big-data-use-cases-for-healthcare>
- [10] Kruse CS, Goswamy R, Rayal Y and Marawi S. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform, Nov 2016.
- [11] J. Han, J. Pei, and M. Kamber. Data mining: concepts and techniques. Elsevier, 2011.
- [12] <http://web.orionhealth.com/rs/981-HE035/images/>
- [13] <https://www.outsource2india.com/Healthcare/articles/machine-learning-big-data-to-change-healthcare.asp>
- [14] D. S. Deulkar and R. R. Deshmukh. Data Mining Classification. Imperial Journal of Interdisciplinary Research Vol. 2, 2016.
- [15] Victor, N., Lopez, D., Abawajy. Privacy models for big data: a survey. Int. J. Big Data Intell. 3, 2016.

- [16] <http://hadoop.apache.org> last access 17.01.2014
- [17] Jee K, Kim GH. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthc Inform Res*, Jun 2013.
- [18] Gunasekaran Manogaran, Chandu Thota, Daphne Lopez, V. Vijayakumar, and Kaja M. Abbas. *Big Data Knowledge System in Healthcare*. Springer International Publishing AG, 2017.
- [19]