Big Data Analysis and Mining for Health Care System

Anonymous

1. INTRODUCTION

With the rapid increase of the population around the world, the treatment models are growing fast and the data is used for some of the decisions for these fast changes models. One of the challenging tasks is the precise diagnosis and due to several factors for complicated disease diagnosis, it can be delayed. If accurate diagnosis can be made, there can be significant life-saving. Data analysis in the health care industry will provide detection of origin of disease, treatment of the disease at lower cost, evolving health profiles, fraud detection in health insurance.

The transformation of paper-based data for healthcare to electronic systems serves various benefits like better management, time-saving, easier access for monitoring data. The huge amount of data collected by health data provider organizations are being implemented using Electronic Health Record (EHR) or Electronic Medical Record (EMR). These datasets are not appropriate for analytical purposes and are not structured. Development in technology and health data sets using electronic are enormous and complex which include the data about patients, previous medical history, changing conditions, treatment and drug prescribed, health improvement, etc. To capture and analyze these unstructured data using the traditional way is difficult. The task of changing raw data into actionable information is complex when there is a huge volume of heterogeneous data and their formats.[1][2][3]

Big Data technique is used for the evidence-based system analysis. Hadoop is an open-source framework used for processing and analyzing of various parts of the huge data holds concurrently after splitting the data and distribute data into various parts. For processing the enormous volume of data in parallel, MapReduce can be used. [1][3] The two major categories for data mining are unsupervised and supervised learning. Predictive rules can be generated to classify the records based on the label. Clustering can be used to separate families of homogeneous data from heterogeneous data. [2]

Section 2 will provide background about each method. Section 3 will provide common similarities between the methods. Section 4 will give details about how each method approaches healthcare systems using different methods. Section 5 will provide summary about method which is not used in other methods. Section 6 will provide conclusion of all three referenced papers.

2. BACKGROUND

3. COMMON THEMES

4. DISCORDANT THEMES

5. NON-OVERLAPPING THEMES

6. CONCLUSION

7. REFERENCES

[1] S. Kumar and M. Singh. Big data analytics for healthcare industry: impact, applications, and tools. *Big Data Mining and Analytics*, 2(1):48–57, March 2019.

[2] Z. A. Mdaghri, M. El Yadari, A. Benyoussef, and A. El Kenz. Study and analysis of data mining for healthcare. In *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, pages 77–82, Oct 2016.

[3] C. Pasupathi and V. Kalavakonda. Evidence based health care system using big data for disease diagnosis. In *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, pages 743–747, Feb 2016.