IJMTST is UGC Approved Journal with Journal Id: 43137

International Journal for Modern Trends in Science and Technology, 8(S08):19-23, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST08S0803
Available online at: https://www.ijmtst.com/vol8si08.html

Big Data Analytics and Machine Learning for Biomedical Information

V Naveen Kumar | A Ravi Kishore

Computer Science and Engineering, Bapatla Engineering College, Bapatla, AP, India

To Cite this Article
V Naveen Kumar and A Ravi Kishore. Big Data Analytics and Machine Learning for Biomedical Information International Journal for Modern Trends in Science and Technology 2022, 8(S08), pp. 19-23. https://doi.org/10.46501/IJMTST08S0803

Article Info
Received: 26 May 2022; Accepted: 24 June 2022; Published: 28 June 2022.

ABSTRACT

Machine learning and big data may be used to healthcare data in the most prominent way. The use of big data and machine learning in healthcare is increasing because of the needs of the patients. Considering, the present situation, it is imperative that efforts be made to maximize the value of the data created by electronic health record apps. In the field of health care, there is a great deal of data, particularly in the brain and cardiac areas. Data in this area need a specific emphasis, and the designs that are now addressing them must include the most recent technological advancements to anticipate certain trends. Different health care architectures are discussed in this article, which utilizes real data from various sources throughout the world. To develop a prediction model and data handling strategies, this paper combines machine learning algorithms with the big data framework.

KEYWORDS: ***Clinical Data, Big Data, Healthcare, Machine Learning, E-Health Records.***

1. INTRODUCTION

The use of big data analytics is critical in the management of information obtained from a variety of sources security messes are made by a variety of sources architectural fails to handle real-time information. The There is a wide range of data models discussed. With the use of a prediction model, the medical field is covered in detail in this article The vast majority of the healthcare industry's output goes to many sub-concepts, as well as all the related areas, will Identify a frequent issue, such as obtaining an MRI or X-ray that demonstrates a patient's illness, and you'll have access to all the relevant data. be stored in a repository by the administrator to ensure the integrity of the data and the authenticity of the information. information gathered How the data is seen and accepted processed foods are the primary emphasis on which one should put one's attention. Implementing any form of prediction model. As to collect and manage data, one must be an expert in this field. sources and pre-processing of various kinds of the data and its subsequent use It's difficult to get data for various projects. task. There are two parts to the job on which this essay focuses. stages. The first step is to conduct a poll on a variety of topics. methods now used by the medical in the context of big data When it comes to the second, strategy, which concerned with elucidating the various big data architectures presently being used medical records are handled with care. In addition, pay attention to the new design, which is more effective, has been put forward. what happens when you handle data from many worldwide source's resources. Data mining is a critical component of modern business. Given the present

state of the world's data is measured in petabytes, of which 75% is generated each day. of the information is medically relevant. A wide range of medical issues and circumstances domains dominated by large amounts of data comprehended. Thefour Vs system of big data, pertaining to the medical field is described as
[1] follows.

(i) Volume

There is an enormous quantity of data generated in the medical field, much more than in any other field combined. Every second, new text and picture data is generated because of medical and genetic research [2]. It's important to understand how this much data is being generated in such a short period of time, as well as the sort of data that has to be taken into account. For a particular condition, such as COVID-19, there is no prior knowledge, thus the DNA samples from the positive instances must be taken, analysed, and compiled into a single pattern of data. A similar trend may be seen in the many DNA patterns and the various studies we do on a daily basis [3].

(ii) Velocity

Speed of data creation is the most important factor. vital to comprehend the structure to handle it ensure that the infrastructure is capable of handling such a massive volume in terms of data. Having a lot of data moving at once might be a problem. There's been some gridlock. Implementation resides in the data structure that is being used. management. Big data has a variety of technologies that can manage different types of data. a problem of this kind.

(iii) Variety

The most critical consideration is the nature of the data collected. The sources and types of data used to acquire this information are critical in this circumstance. Listed below are the three types of data. For instance, a pre-designated structure with rows and columns of data might be a systematic format, which is structured data. A lot of the data that may be utilized for prediction is in the form of language, and this data can be used to forecast anything related to the issue. In the form of multimedia, unstructured data may take on a variety of shapes and sizes, including the likes of films, audio files, and photographs. Recent studies have gathered these facts from a variety of sources. Unstructured data, such as multimedia, may be collected in a variety of ways, much as MRI scans for brain illnesses. Structured and unstructured information are both present in semi-structured data. XML and JSON are the primary data collection formats in this situation. The emphasis of this strategy is on how to collect several sorts of information at the same time without compromising the data pattern's architectural design.

(iv) Veracity

The reliability of data acquired from a variety of sources is a topic of conversation. For this kind of real-time data, it is impossible to test and refine hypotheses based on the information already acquired. Mismatches or missing data will invalidate the decision models, which will lead to poor results. The three kings of predictions, predictive modelling, description modelling, and prescriptive modelling, are used in a common study of healthcare applications in big data. Figure 1 depicts the survey's methodology [1]. Various sources are discussed in depth in the literature study, including the current systems. The article's lateral section covers a variety of topics relevant to big data, including a review of the literature, an analysis of the present system, domain modelling, a proposal for a new system, and a discussion of the most effective technique currently in use.

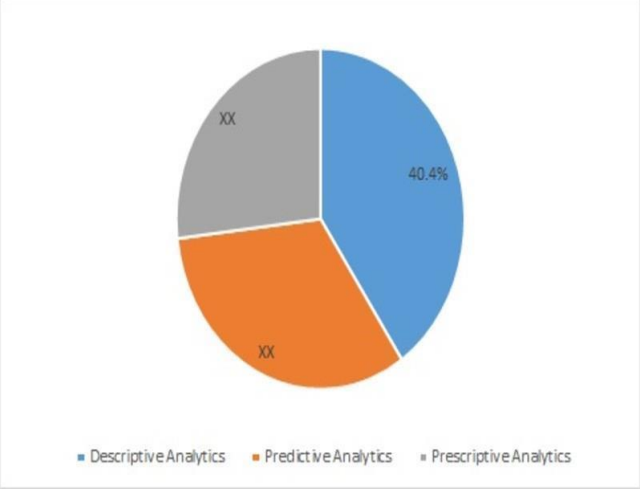

Fig.1 Analysis of the rise of big data in the medical field in India

2. RELATEDWORK

Personalized health care management with the use of big data is discussed in [1]. A virtual physiological human (VPH) was utilized to communicate with a doctor

through the technology that serves as our assistant in this architectural piece. The fifth V of big data was explored in this article. That is worth.

This article discusses the need of using big data in health care and providing patients with individualized information. Doctors and engineers were linked in this article to better execute tailored health care services, and the author described the approach they used to accomplish so.

[2] A tailored health application for watches or mobile devices is used to keep tabs on the patient's well-being, and the EHR implementation addresses issues such as how a patient's personal health information should be stored. In 2009, the World Health Organization (WHO) revisedits clinical guidelines, resulting in significant changes to the medical field's implementation, including alterations to the system's security and data manipulation. Figure 2 shows how to put this into action: [2] The internet of things (IoT) and the health care business are discussed in [3]. It is possible to identifylow, medium, and high-risk management in this paradigm. In contrast to rural regions, citiesget 80 percentof medical care. In this article, a design is shown that can determine if a disease has a low or high probability of spreading. It's now clear if the therapy is for cure or prevention. In Figure 3, the architecture is explained. To determine whether a patient is at low or high risk, a correlation will be established between the data collected by the IoT device and the risk prediction algorithms. As a result, it's possible to either avoid or treat the issue. [4] discusses many methods for doing a prediction analysis.

Analyzing data in three distinct ways: It is possible to predict the conclusion of a given issue using predictive analytics, which relies on prior experience with the algorithm to do so. Using the problem description as a starting point, descriptive analytics aims to provide an explanation of the issue and possible remedies. (3) Prescriptive analytics is based on experience in the same issue area and on the implementation of detail insight into the underlying cause. Figure 4 depicts the implementation paths for big data. Using big data, [5] identifies human activity patterns and these patterns define the process of comprehending human emotions and behaviorto detect human health trends. Live broadcasting and the establishment of a cloud domain are not being addressed by the current technology [6-10].

In terms of data storage and prediction model building, cloud services make use of several technologies like Spark with Mahout Machine Learning. However, the major issue of live streaming has not been addressed in any of the publications thus far. It has not yet been proposed or implemented in the medical field to use live streaming. Live stream data is not currently being managed by this present system, which has data management with a fixed size and weight. This section's lateral portion will focuson the live streaming mechanism in the medical field.

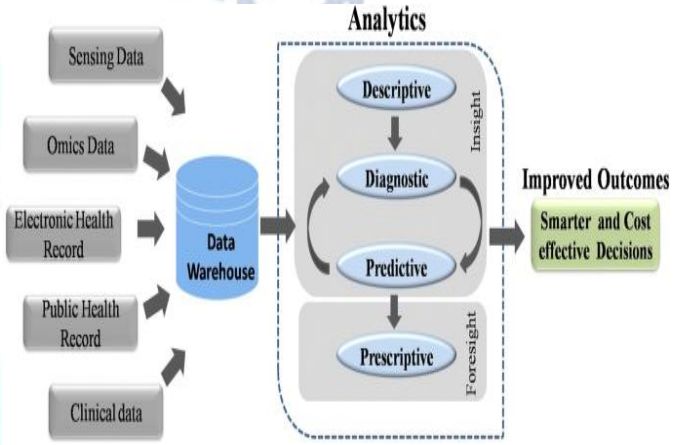

Fig.2 E-Health Records with Big Data Analytics

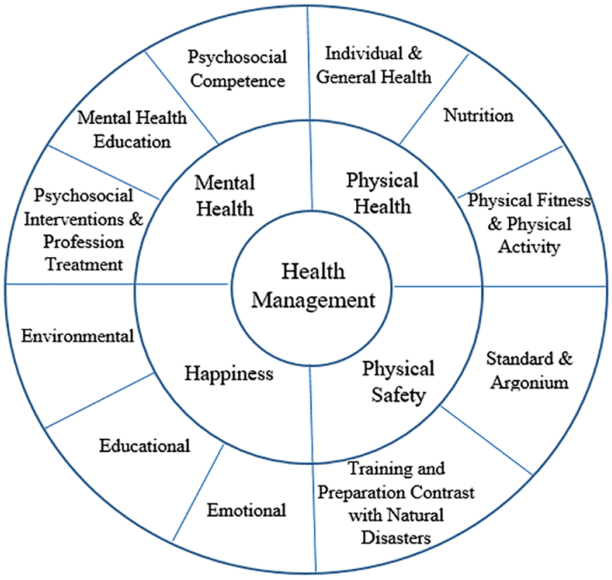

Fig.3 Conceptual Health Management system

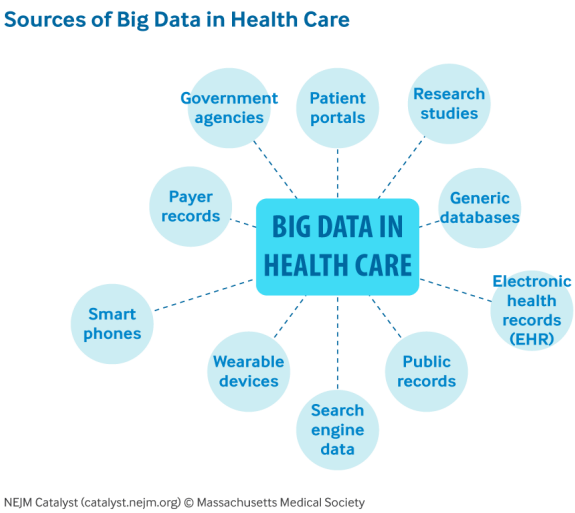

Fig.4 Pathway to Healthcare

3. PROPOSED SYSTEM ARCHITECTURE

There are two types of health care management that are discussed in the proposed system, both of which are designed to keep the system safe and secure. Both elements of the suggested strategy are listed below. The following are possible outcomes for the composition of this article. Building a single architecture does not reveal the best practices [11, 12]. Architectures are not meant to be used by everyone.

FIGURE 2 exemplifies the finest application to date. Examples of several ways to solving the issue [6-8] may be foundwithin this designframework.The data collecting and refining process is separated, as is the storage mechanism's architecture and method. There are a variety of ways to approach this storing technique. It is possible to employ cloud storage or any other kind of private storage method. Data may be used by separating it into pieces, which are then remapped to other feature types or categories using the MapReduce component. As a large data framework, Spark was used to build real-time datasets The processing of real-time data is one of the most critical aspects of big data, and this may be a highly effective application of the current approach. Kafka is the best way to handle live streaming data implementations [9-10]. The real-time broadcasting of financial and medical data.

As an example, flight data may also be live data that must be monitored. [13-14]. Figure 5 depicts how Spark streaming works. However, Kafka and flume are also trustworthy for data input mechanisms [15]. An IoT device that is connected to the cloud will take the patient's health information using the machine learning algorithm, and the information that was considered as input will be verified by the machine learning models as to whether it is genuine or not [16-17]. If the data is accurate, the information will be processed and the patient's information will be sent to the doctor. When it comes to this, it's data that's current. The machine learning models integrated into the gadget are based on real-time data and user experience. Some internal processes will be implemented, and the live data will be audited and handled. Scala is the primary framework, with Kafka serving as the data input platform for a large data system [18, 19].

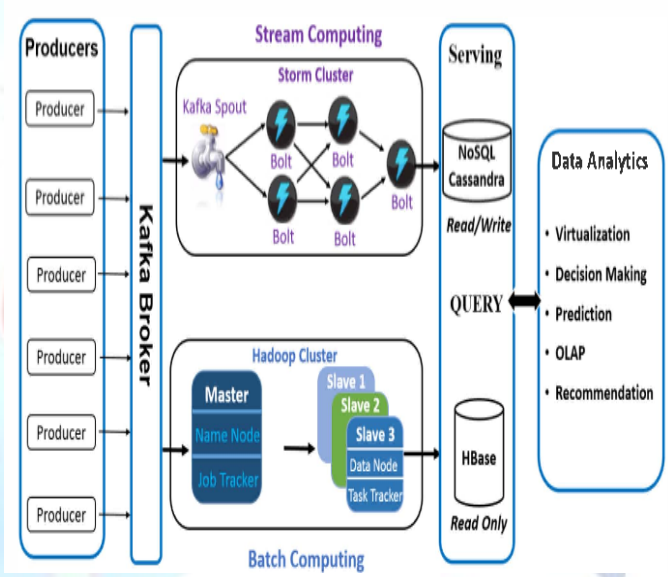

Fig.5 Suggested Framework

4. FUTURE SCOPE AND CONCLUSION

The most significant mechanism in recent years was Big Data, which was developed to handle data in many formats. Big Data is now. The data mining mechanisms will pre-process all the information received from various sources, and these data characteristics will be sent to the frameworks in accordance with the implementation and issues. This article explores the use of several big data platforms in the medical field. There must be no data loss in the information obtained from various sources and the prediction models will use that information. After the results of the manipulation were mapped to the repository, the feature developed on the big data platform utilizing any sort of cloud storage was required. There will be a period when the data will be collected from various repositories and stored according to the file format. The recorded data is subsequently used in accordance with the task at hand.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] K. Nti, J. A. Quarcoo, J. Aning and G. K. Fosu, "A mini-review of machine learning in big data analytics: Applications, challenges, and prospects," in Big Data Mining and Analytics, vol. 5, no. 2, pp. 81-97, June 2022, doi: 10.26599/BDMA.2021.9020028.

[2] M. Assefi, E. Behravesh, G. Liu and A. P. Tafti, "Big data machine learning using apache spark MLlib," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 3492-3498, doi: 10.1109/BigData.2017.8258338.

[3] J. McHugh, P. E. Cuddihy, J. W. Williams, K. S. Aggour, V. S. Kumar and V. Mulwad, "Integrated access to big data polystores through a knowledge-driven framework," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 1494-1503, doi: 10.1109/BigData.2017.8258083.

[4] S. Chunzi, W. Xuanren and L. Ling, "The Application of Big Data Analytics in Online Foreign Language Learning among College Students : Empirical Research on Monitoring the Learning Outcomes and Predicting Final Grades," 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), 2020, pp. 266-269, doi: 10.1109/MLBDBI51377.2020.00056.

[5] C. K. Leung, Y. Chen, C. S. H. Hoi, S. Shang and A. Cuzzocrea, "Machine Learning and OLAP on Big COVID-19 Data," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 5118-5127, doi: 10.1109/BigData50022.2020.9378407.

[6] Nurhayati, Busman and V. Amrizal, "Big Data Analysis Using Hadoop Framework and Machine Learning as Decision Support System (DSS) (Case Study: Knowledge of Islam Mindset)," 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018, pp. 1-6, doi: 10.1109/CITSM.2018.8674354.

[7] M. Muniswamaiah, T. Agerwala and C. C. Tappert, "Federated Query processing for Big Data in Data Science," 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 6145-6147, doi: 10.1109/BigData47090.2019.9005530.

[8] S. Sehgal and M. Agarwal, "Analogous Examination of Various Machine Learning Algorithm Applied to Big Data," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018, pp. 121-123, doi: 10.1109/ICACCCN.2018.8748855.

[9] S. Kobashi et al., "Prediction of post-operative implanted knee function using machine learning in clinical big data," 2016 International Conference on Machine Learning and Cybernetics (ICMLC), 2016, pp. 195-200, doi: 10.1109/ICMLC.2016.7860900.

[10] L. Liu, "Deception, Robustness and Trust in Big Data Fueled Deep Learning Systems," 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 3-3, doi: 10.1109/BigData47090.2019.9005597.

[11] H. H. Huang and H. Liu, "Big data machine learning and graph analytics: Current state and future challenges," 2014 IEEE International Conference on Big Data (Big Data), 2014, pp. 16-17, doi: 10.1109/BigData.2014.7004471.

[12] A. A. Akinyelu, "Hybrid Machine Learning-Based Intelligent Technique for Improved Big Data Analytics," 2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI), 2019, pp. 7-11, doi: 10.1109/ISCMI47871.2019.9004305.

[13] W. Zhong, N. Yu and C. Ai, "Applying big data based deep learning system to intrusion detection," in Big Data Mining and Analytics, vol. 3, no. 3, pp. 181-195, Sept. 2020, doi: 10.26599/BDMA.2020.9020003.

[14] K. R. Swetha, N. M, A. M. P and M. Y. M, "Prediction of Pneumonia Using Big Data, Deep Learning and Machine Learning Techniques," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1697-1700, doi: 10.1109/ICCES51350.2021.9489188.

[15] J. V. N. Lakshmi and A. Sheshasaayee, "Machine learning approaches on map reduce for Big Data analytics," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), 2015, pp. 480-484, doi: 10.1109/ICGCIoT.2015.7380512.

[16] Z. Peng, "Stocks Analysis and Prediction Using Big Data Analytics," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2019, pp. 309-312, doi: 10.1109/ICITBS.2019.00081.

[17] G. Sasubilli and A. Kumar, "Machine Learning and Big Data Implementation on Health Care data," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 859-864, doi: 10.1109/ICICCS48265.2020.9120906.

[18] I. Ha and B. Back, "Effective Garbage Data Filtering Algorithm for SNS Big Data Processing by Machine Learning," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 520-524, doi: 10.1109/ICAIIC48513.2020.9065276.

[19] X. Li, "Application exploration of Daqing smart tourism big data based on association rule mining and machine learning," 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), 2022, pp. 54-56, doi: 10.1109/EEBDA53927.2022.9745010.