A Novel IoT-Based Approach Towards Diabetes Prediction Using Big Data

Riya Biswas, Souvik Pal, Nguyen Ha Huy Cuong and Arindam Chakrabarty

Abstract Big data is a modern teamster of today's economical world. Data are being digitalized in today's world as imperative judgment is taken by Big data analytics. In our manuscript, we have discussed about Big data analytics in IoT ecosystems and its implications in healthcare. Healthcare is concerned now a days and big data is holding all the supportive hands in IoT-based healthcare systems. In healthcare, we have discussed about Diabetes Mellitus which is a non-communicable disease. This paper deals with the proposed system of diagnosis of diabetes. Hence it is assertive that we do some surveys on how we can manage to handle large data files, technologies are defined and also predictions of diabetes through IOT sensor and management have been discussed.

Keywords Big data · Hadoop · Map reduce · HDFS · Pig · HIVE · HBase · IoT

1 Introduction

In current surroundings data is generating from multiple origin. These data is of multiple diversity. This bulk of enormous data is considered as Big Data. Big data and IoT are the buzz words now days. It is used to express cumbrous bulk of structured and unstructured data. Some characteristics of Big data being discussed [1, 2].

R. Biswas (✉)
Department of Computer Science & Engineering, JIS College of Engineering, Kalyani, India
e-mail: sonariyabiswas@yahoo.in

S. Pal
Department of Computer Science & Engineering, Brainware University, Kolkata, India
e-mail: souvikpal22@gmail.com

N. H. H. Cuong
College of Information Technology, The University of Da Nang, Da Nang, Vietnam
e-mail: nguyenhahuycuong@gmail.com

A. Chakrabarty
Department of Management, Rajiv Gandhi University, Itanagar, India
e-mail: arindam.management@gmail.com

The amount of data can be blunder, malleable and scalable data which are processed through technologies used called Hadoop, Map Reduce, HIVE, Pig. The production of volume of data is rapidly enlarging every year due to occurrence of newly technologies, accessory and communication [3]. In the modern milestone of smart phones and wearable devices, endless sum of health data folder of patient from different challenges continue featured by healthcare industry [4]. Mostly confusion begin where system process through heterogeneous data sets [5, 6]. In healthcare sector a leading non-communicable disease (NCD) is Diabetes Mellitus. There are basically three stages of type 1 diabetes, type 2 Diabetes, Gestational diabetes. Diabetes is a scheme of metabolic diseases consist of high blood sugar levels concluded lengthy season a numerous operation supported on Internet Of Things been developed for management of diabetes and it composite of physical objects [7, 8]. IOT is mostly a model for interconnecting sensor which does tracking, sensing, processing and diagnosing, coming up with a enclosed device and detector which can link up and also exchange content beyond the internet [9, 10].

In this paper, we are going to discuss literature survey of the related work in the Sect. 2. Section 3 deals with the architecture of the proposed diabetes diagnosis system, proposed algorithm, and sequence diagram of the algorithm.

2 Literature Survey

In this section, we have discussed on literature survey of the background study. Chavan et al. [9] have discussed about Big data is a word which defines massive and convoluted set of data. Some technologies like Hadoop, HDFS, Map Reduce, Pig, Hive, HBase being used. Khan et al. [11] have expressed a proposed data life cycle which utilize employ the technologies and nomenclature of Big data management, investigating and scarceness. Nizam and Hassan [12] have discussed that it is tough to operate with Big data resolving management traditional dataset. Chen et al. [13] have discussed that Initially generic background of Big data is inform then study about the connected technologies i.e. cloud computing, Internet Of Things, data centers and Hadoop. Archenaa and Mary Anita [14] have deliberate about the approach of how we expose newly expose surplus value from the data autogenic by healthcare and government. Prasad et al. [15] have discussed that diabetes is one of the leading non-communicable disease Mellitus. This system will prophesy searching algorithm in Hadoop/Map Reduce. Huzooree et al. [16] has explains that Diabetes Mellitus (DM). The goal of this paper is to ecumenical review centering on recent glucose projection model is declared depending on the rating to performing data analytics in wireless body area in network system. Kumar and Pranavi [17] has discussed that the important function is providing dilution healthcare by modern application such as Big data and cloud. A ecumenical survey is made on diabetes dataset with random forest (RF), SVM, k-NN, CART and LDA algorithms. Joaheer and Nagowah [18] have explained Telemedicine, Electronic Health Records (EHR) and social media. This paper also describes the repung of Big data and also it proposed architecture for

diabetes Mellitus to predict patient with chronic disease in maturius. Saravana kumar et al. [19] have discussed that the unstructured nature of lifecycle from healthcare of Big data This paper analyzing algorithm in Hadoop/Map Reduce is used for prediction of diabetes type, hindrance. Al-Taee et al. [20] has discussed that the self-management of diabetes by IOT based podium. A completely practical model system is created, achieved, point-to-point function is approved successful.

3   Architecture of Proposed Diabetes Diagnosis System

This section describes the architecture of diabetes diagnosis system that analyzes the various Data and initially it accumulated data from numerous devices and it is initially stored in an unstructured or semi-structured format. Initially data should be digitized to stem EHR as well as data are smart devices, research and development SNM data repository which is begin captured by existent technologies and used to redirect those data to centralized database for anatomy. The data are gathered for processing in Hadoop data system then data will be accumulated by apache flume. Apache flume is used here which is a item of hadoop ecosystem. Then the data will be pushed to Hbase by agents for further processing (Fig. 1).

The outputted data moves to HIVE it is a business application running in SQL queries against a hadoop cluster. It uses then map reduce. Map reduce has two tasks
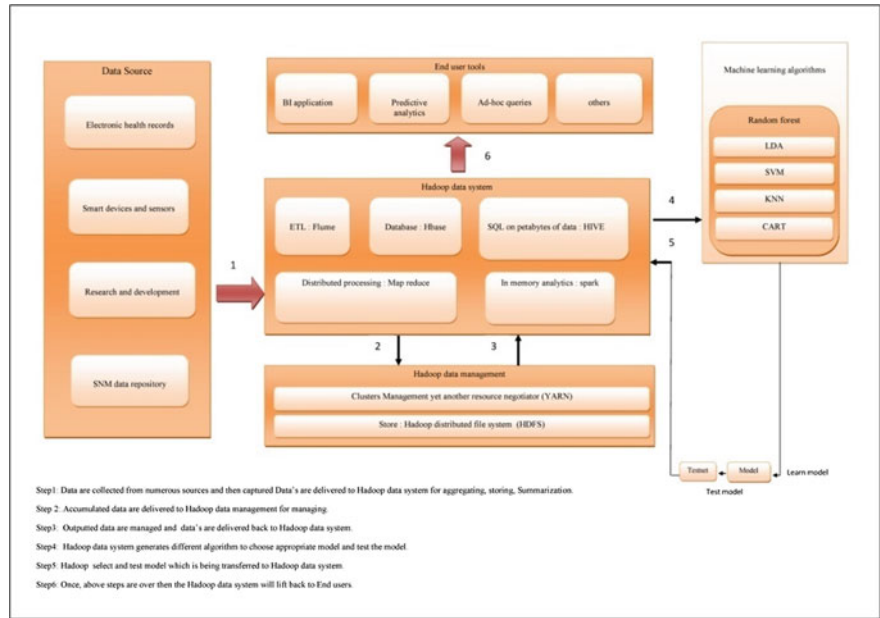

Fig. 1   Architecture for our algorithm

data is splitted and passed into mapping function for produce output values. Next data can use spark which is a frame work that is same as hadoop which provides open-source platforms and can be used by anyone. Then the outputted data are aggregated in hadoop data manager where YARN is used i.e. it is a centralized platform. Then implementing various machine learning algorithms such as RF, LDA, CART and K-NN for prediction it also learn specific data and the test absolute model which will be back to hadoop and hadoop data system will completely send it back to end user tool.

Here it initially identify the course of people tolerate from diabetes registry by working healthcare analytics to big data technology for identifying the diabetes.

3.1 Proposed Algorithm

This section deals with the proposed algorithm of the diabetes diagnosis system.

Step 1: *Initially data assembles from various sources like EHR, smart device and sensors devices and research and development and SNM data respitory.*

Step 2: *Hadoop is a framework which permits for distributed processing of enormous data set. It is a framework which has a capability for stocking and considering data which are prevailing in various machines. It also service map reduce which permits for diving the query into limited chunk and achieve them in co-ordinately.*

Step 3: *Initialized data's from various sources need to be delivered to hadoop data system to process the data where the data's are accumulated by using apache flume, then data moves to HIVE which run SQL queries then data's are place down to map Reduce for summarization. After that spark is used which furnish a open-source platform.*

Step 4: *The processed data's aggregated from hadoop data system need to be managed, so to manage the data are implemented in hadoop data manager where YRAN is used to add new features to the hadoop. It is a centralized platform used for Resource Management.*

Step 5: *The outputted values of managed data are a switched back to hadoop data system.*

Step 6: *The outputted data should be evaluated so Hadoop data system will generate machine learning algorithm.*

Step 7: *Hadoop ensures the appropriate algorithm for the data does evaluated according to their category, from the set of algorithm to determine the appropriate data pattern and lining the data for earning prediction.*

Step 8: *Outputted data tested by Hadoop and draws one specialized model of algorithm and learned the specific model.*

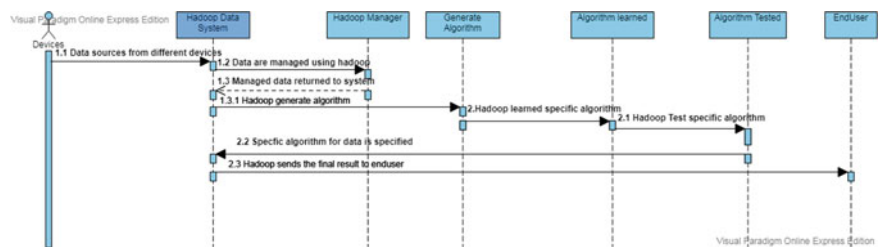Step 9: *Outputted data are for specific algorithm switch back to hadoop data system.*

Fig. 2 Sequence diagram for our proposed algorithm

Step 10: *Once, the above steps are done hadoop data system will switch over to end user tools.*

3.2 Sequence Diagram

Initially data are accumulated from multiple sources. Data are being processed by hadoop data system and managed by hadoop data manager and outputted managed data are swapped back to hadoop data system. The hadoop data system will generate machine learning algorithm. According to the kind of data it will choose specific model and the outputted model will reversed back to hadoop data system. After completion of above steps hadoop will swap back to end user tools. In this section, Fig. 2 is expressing that the data sources which are being assembled from various devices are processed and managed. It also analyzes various algorithm and choose appropriate algorithm to learned specific model to predict diabetes.

4 Analysis

In our work, we have built the need of predicting techniques to measures the diabetes unlike the traditional models which doesn't provide enough efficiency, accuracy and fastest delivery. This technique possesses several data from EHR, R&D and other sources like smart devices.

By using existing technologies, it is possible to capture and send to a centralized database for analysis. Also, unlike most of the other proposed works, we used to gathered data from various devices and processed the data in hadoop data system and then processed data are being managed by hadoop data management and additionally it also applied machine learning algorithm such as RF, LDA, CART and K-NN provoked by hadoop. The machine learning algorithm has main benefits over the most other techniques as it provides more accurate throughput to user and it gains the performance rate of the model.
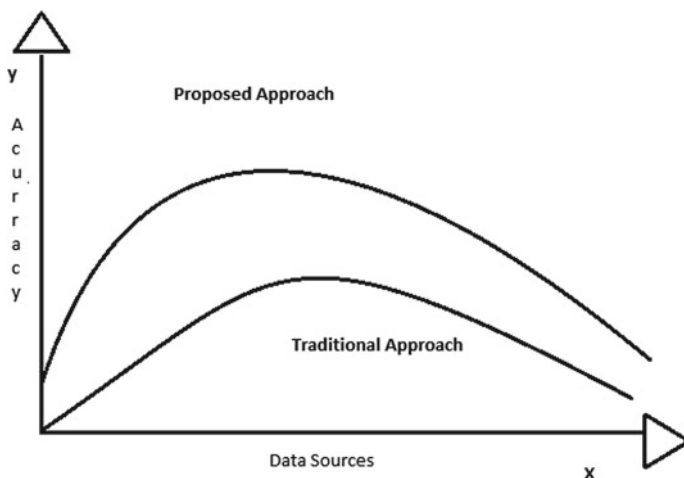
Fig. 3 Compares the traditional approach with the proposed approach shows accuracy will increase as data sources with increase

In our proposed work, considering all factors, we can say that the efficiency, accuracy and fastest delivery medical care at lower cost. This can be compared with the traditional approach by depicting them in the form of graphs for both the traditional approach and our proposed approach. It is depicted that data that we are getting from multiple sources are used to predict the diabetes. So, here accuracy perform a efficient role. Figure 3 compares the data sources with the accuracy as the data sources will increase accuracy throughput will also increase in proposed approach and decrease in traditional approach. Figure 4 is expressing the comparison of cost of traditional approach as compared to proposed approach the cost will decline in proposed approach.

5 Conclusion

Peoples are engaged in today's world in the feverish slots and not pickings any care of their own health, starring to difficulties of continuing disease such as diabetes. In this paper, a recent framework is suggested that utilize. This framework will analyze and predict diabetes Mellitus and providing way to improve healthcare complexity and delivering earliest potential working. As well as this framework is operate for self-treatment and also in future providing faster medical care within a chip costs. In this paper, it also provides many various machine learning algorithms such as RF, SVM, CART, LDA and K-NN to predicting data patterns. The frame work is working currently under development.
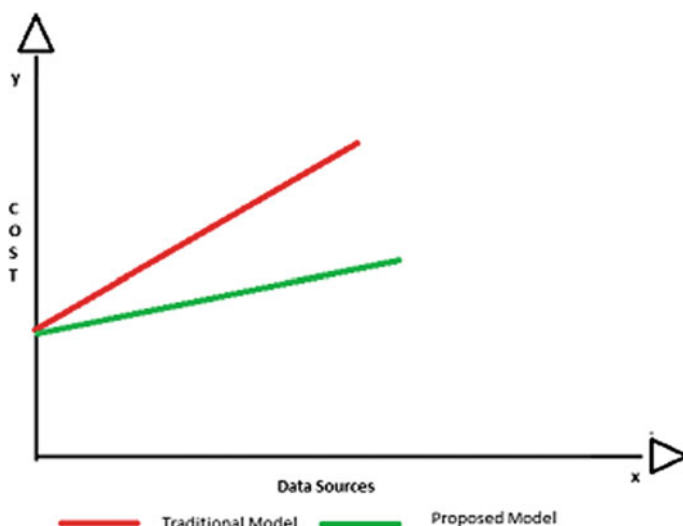
Fig. 4 Compares the traditional approach cost with the proposed approach and display the increase in cost in traditional approach and decrease in proposed approach

References

1. Joshitta S (2016) Applications of big data analytics for diagnosing diabetic mellitus: issues and challenges. Int J Recent Trends Eng Res 02:454–461
2. Emani CK, Cullot N, Nicolle C (2015) Understandable big data: a survey. LE2I UMR6306, CNRS, ENSAM, Univ. Bourgogne Franche-Comté, F-21000 Dijon, vol 17, pp 70–81
3. Honest N, Patel A (2016) A survey of big data analytics. Int J Inf Sci Tech (IJIST) 6(1/2)
4. Pooja M, Das D (2016) Comparative analysis of IoT based healthcare architectures. Int J Appl Eng Res 11(20):10216–10221. ISSN 0973-4562
5. Gaitanou P, Garoufallou E, Balatsoukas P (2014) The effectiveness of big data in health care: a systematic review. In: Closs S et al (Eds) MTSR 2014, CCIS 478, pp 141–153
6. Thara DK, Premasudha BG, Ram VR, Suma R (2016) Impact of big data in healthcare: a survey. In: 2016 2nd international conference on contemporary computing and informatics (IC3I), Noida, pp 729–735
7. Azzawi MA, Hassan R, Abu Bakar KA (2016) A review on internet of things (IoT) in healthcare. Int J Appl Eng Res 11(20):10216–10221. ISSN 0973-4562
8. Deshkar S, Thanseeh RA, Menon VG (2017) A review on IoT based m-health systems for diabetes. Int J Comput Sci Telecommun 8(1)
9. Chavan V, Rajesh, Phursule N (2014) Survey paper on big data. Int J Comput Sci Inf Technol (IJCSIT) 5(6):7932–7939
10. Alelyani S, Ibrahim A (2018) Internet-of-things in telemedicine for diabetes management. In: 2018 15th learning and technology conference (L&T), Jeddah, pp 20–23
11. Khan N, Yaqoob I, Hashem IAT, Inayat Z, Kamaleldin Mahmoud Ali W, Alam M, Shiraz M, Gani A (2014) Big data: survey, technologies, opportunities, and challenges. Sci World J (Hindawi Publishing Corporatione) 2014:18 pages
12. Nizam T, Hassan SI (2017) Big data: a survey paper on big data innovation and its technology. Int J Adv Res Comput Sci 8(5):2173–2177
13. Chen M, Mao S, Liu Y (2014) Big data: a survey. Mobile Netw Appl 19:171–209

14. Archenaa J, Mary Anita EA (2015) A survey of big data analytics in healthcare and government Proc Comput Sci 50:408–413
15. Prasad ST, Sangavi S, Deepa A, Sairabanu F, Ragasudha R (2017) Diabetic data analysis in big data with predictive method. In: 2017 international conference on algorithms, methodology, models and applications in emerging technologies (ICAMMAET), Chennai, pp 1–4
16. Huzooree G, Khedo KK, Joonas N (2017) Glucose prediction data analytics for diabetic patients monitoring. In: 2017 1st international conference on next generation computing applications (NextComp), Mauritius, pp 188–195
17. Kumar PS, Pranavi S (2017) Performance analysis of machine learning algorithms on diabetes dataset using big data analytics. In: 2017 international conference on Infocom technologies and unmanned systems (trends and future directions) (ICTUS), Dubai, pp 508–513
18. Joaheer R, Nagowah SD (2017) A big data framework for diabetes in Mauritius. In: 2017 international conference on Infocom technologies and unmanned systems (trends and future directions) (ICTUS), Dubai, pp 126–132
19. Saravana kumar NM, Eswari T, Sampath P, Lavanya S (2015) Predictive methodology for diabetic data analysis in big data. Published by Elsevier B.V. 50:203–208
20. Al-Taee MA, Al-Nuaimy W, Al-Ataby A, Muhsin ZJ, Abood SN (2015) Mobile health platform for diabetes management based on the Internet-of-Things. In: 2015 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT), Amman, pp 1–5