An application overview of IoT enabled-big data analytics in Health sector with special reference to Covid-19

R. Biswas*

Department of Physics, Tezpur University, Tezpur-784028, India

*Corresponding author: rajib@tezu.ernet.in

Abstract— Big Data analytics has come a long way since its inception. This field is growing day by day. With the advent of large handling capacity of computational analysis of modern computing systems as well as Internet of Things (IoT), this field has revolutionized the way we think about data. It has influenced the major domains such as healthcare, automobile, computing, climatology, and space communications. Of late, the health care sector has been largely influenced by this. This communication deals with the areas of healthcare where big data analytics has been   largely   influential. Encompassing the basics of Big Data Analytics (BDA) driven by IoT, the applications of it in healthcare sector are outlined, accompanied by future expectations. Additionally, it also presents a comprehensive analysis of recent application with special reference to Covid-19 in this sector.

Keywords: Bigdata; IoT, Big Data Analytics; Covid-19; healthcare

1. INTRODUCTION

Literally, big data refers to voluminous information. However, when, we incorporate the term Big data in healthcare, and then there arises diverse definitions in this sector. However, the suitable definition can be given as the enormous data entailing biological, clinical, and environmental as well as lifestyle information concerning large individuals, which corresponds to their health and wellbeing along a time span. In nonprofessional's word, it refers to dataset whose enormity in size cannot be handled by single database software   for   capturing,   storing and   subsequent   analysis. Accordingly, the stress goes to the parameters involving size and volume. Accordingly, it engages three V's. They are variety, veracity and velocity. The first term variety links to diverse types, sources and format. Again, veracity refers to quality and validity whereas velocity encapsulates availability in time. Moreover, reliability, data protection and privacy are also key points that have influence on big data. In order to give more impetus to the working of Big Data, Internet of Things (IoT) emerges out as a vital source. Of late, IoT has become a common technological term for big enterprises. With the advent of smart objects growing at a rate higher than the population of world, IoT has now become an indispensable part of the modern era. However, it goes with a caution of acceptability followed by adaptability [1-9]. In short, Big Data Analytics refers to effective integration and efficient analysis of various forms of data over a period of time, which has the ability to cater to some impending problems. This concise documentation will

deal with the areas of healthcare where Big data analytics (BDA) has been largely influential. Encompassing the basics of IoT driven Big data analytics, the applications of it in healthcare sector will be outlined, accompanied by future expectations. Additionally, this brief outline presents a comprehensive analysis of very recent applications in this sector with special reference to Covid-19.

2. IoT and its journey with BDA

IoT has evolved a long way. It is a common thread, which connects all assorted devices in a synergistic way. Modern house architecture is mostly IoT enabled. Starting form Thermistors, water heater, fridge and smart lighting system, these household applications are attached/connected to each other through IoT. With the advent of RFID technology [9, 10], IoT evolves dramatically engaging several aspects and stakeholders spanning academia and industry. The whole concept can be generalized as a fusion of three perspectives, viz.; orientation with things, orientation with internet and lastly orientation with semantics. Each of the perspective goes with its own agenda. Meanwhile, the fastest growing IoT spans wide range of voluminous data, which stem from strategically distributed sensors/smart objects. This in turn gives rise to assorted analytics. Yet, there is a possibility of data analytics stemming from IoT to become ineffective as well as costly; in case, we habitually costly if we always stick to cumulative transfer and subsequent handling of data in a central storage system. In order to tackle this issue, researchers develop an innovative strategy of micro service oriented platform, which will decentralize the data tree. This platform is driven by software-defined infrastructure (SDI) which will disrupt the environmental monolith of IoBDA (IoT oriented Big Data Analytics) implementation. Accordingly, SDI comprises of software-defined network (SDN) and software-defined storage (SDS)[10-18]. As such, the dissociation of data transmission from IoT nodes such as switches and routers is enabled by SDN. SDS that helps in decoupling management of data store form the entire unit further accompanies this. As a result, these two vital components emerge as an integral part in facilitating workload demands via OPI in case of heterogeneous hardware. Thus, microservice-oriented platform proves to be a boon in segregating business logics, which are domain specific, from resource control and management [10, 13, 15-18].

3. Objective of BDA and working principle

The objective of Big Data analytics is towards developing new technologies such as capturing IoT oriented devices, sensors, and mobile applications with the following outlines.

a. Collection of genomic information became cheaper

b. Increasing Patient social communications in digital forms

c. Accumulation of more medical knowledge/discoveries

The modus operandi of BDA has been illustrated in Figure 1. As can be seen, there are components spanning from lab to pharmacy entailing patient, physician and research & development. It is also accompanied by social networking data of the patient in order to build up a synergistic effort in resource optimization. The idea of BDA is to make healthcare accessible to all with the optimal output and simultaneously saving precious time and expenditure of patient. In terms of data analysis, IoTBDA uses Monte Carlo or Convergence Analytics as per expediency [9, 10].
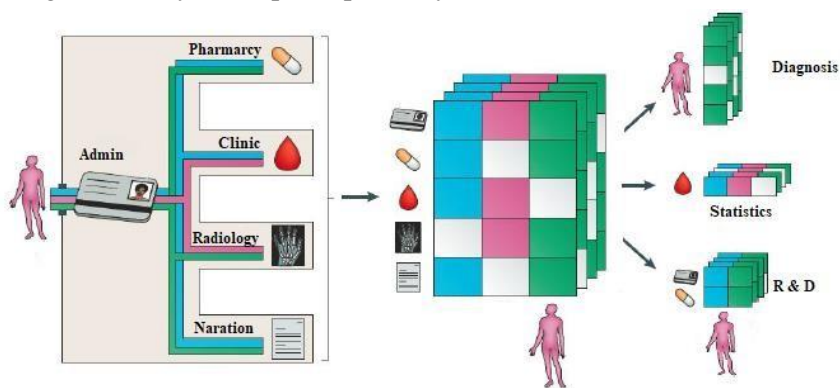

Fig. 1. Schematic of BDA in healthcare

Key elements for BDA

As the name goes, BDA in Healthcare is dependent on several key elements. For example, patient data is a vital element. More technically, we can term it as electronic health records. Again, when we look at this key aspect, then there arise two questions regarding data related to patients. Are they structured or unstructured? Accordingly, we have two sort of database. They are viz. Structured Health Records and Unstructured Health Records. Additionally, there is the necessity of a supporting system for handling these voluminous data. Nonetheless, the clinical data and medicinal data are also one of the prima facie of BDA. With these components, there is addition of another platform in the context of social networking data of stakeholders.

Beneficial effect of BDA

BDA is revolutionizing the health industry. There are several beneficial influence of BDA; few of which are enlisted as below.

a) Healthy living

b) Better understanding of triggers of chronic diseases for

effective early detection

c) Population health

d) Precision Medicine

e) Optimizing workflows in Healthcare

In the next section, all are elaborated.

　　a)　Healthy Living.

One of the beneficial effects of Big Data over healthcare is the ability to provide valuable tool which acts as a boon for behavioral change. As for example, we can cite mobile health (mHealth). This tool contemplates lifestyle data which is linked to nutrition, physical activity, sleeping habit of user and then integrates them with large factual reference data. Afterwards, it personalizes interventions of the user. Not only that, it furnishes valuable information enabling proper identification of proper medication, thus giving a detailed overview of progress and setbacks in therapy.

　　b)　Better understanding of triggers of chronic diseases for effective early detection

In addition to that, BD succors in early diagnosis of triggers of chronic diseases. It can propel ongoing research towards smooth understanding between social and other parameters. They comprise of physical behaviors, nutrition, genetic factors, environmental factors and development of mental/physical diseases. In general, the interaction between different systems controlling disease propagation still baffles researchers. Accordingly, we rely on BD which can provide us a solution by building an integrated view of health encompassing various biomarkers. These will eventually (i.e. omics, quantified self-data) can help to improve early detection of diseases and long term management of adverse health factors thereby reducing costs.

　　c)　Population health

Public health policy is influenced by region and socio-economic status. They in turn define societal action whose main goal is to improve health outcomes. Through specific interventions, BDA can act as guidance towards addressing policies pertinent to a certain population. These policies are again dependent on the quality of research and interventions. Again, there is a dearth of methods for validating certain interventions as in mental health domain. Although BDA supports health policies, still there exist several obstructions. For instance, privacy and protection of data hinder analysis through a combined approach of healthcare provider and service. Likewise, unstructured health records are also a problem for BDA.

When there will be scalable, methodological and privacy friendly outcomes assisted by advanced statistical methods, it will pave way for development of precise and effective interventions. With advent in technology, one can synergistically amalgamate data from healthcare environment and information

from society. Social networks, forums, blogs etc can befit health environment by providing a wealth of data which is directly implemented for benefit of public health. When one combines information from informal and data emerging from diagnosis and surveillance, it is possible to achieve a early detection of disease outbreaks and transmission. As for instance, one can cite about the ARGO model which was basically a predictive model. This has been accomplished through amalgamation of tracking of disease, spread dynamics and surveillance by adopting social networking means like Twitter. Owing to well-planned analysis of these data with incorporation of travel, trade and climate change, it is possible to attain a predictive model for population based interventions as well as improved treatment of individual patient. Apart from this, with early detection of disease outbreak, the government experts can well coordinate important strategies, like, quarantine and vaccination.

d) Precision Medicine

Value based healthcare is now a buzzword. This can be treated as a guiding principle for sustainable health. This actually comprises of units like patient reported outcome and the amount of cost incurring care path and the eventual decision of payment. In such case, the healthcare is offered incentives when it surpasses some performance index. This is not to be confused with treatment of patient, rather it has link to patient related outcome. For proper implementation of VAHC, there must be a streamlined flow of collection, analysis and aggregation of data by inclusion of total care path, cost etc. While doing so, patient linked health outcomes require monitoring in three stages. They are during, before and after treatment. Certain challenges in this connection can be cited like lacking of an updated admin care, which can cater to associated specific care paths to produce an accurate estimate of expenditure incurred. In case there is a succinct connection of care processes as well as care paths with assistance from a huge database, empirical evidence based decision for specific therapies will be materialized. To do that, there is immediate necessity of standardized and authenticated methods.

e) Optimizing workflows in Healthcare

In industrial sectors, most of the things are predictive. Hence, priority, objectives are well defined. Nevertheless, in healthcare sector, this is completely opposite. It is quite a volatile system. Influenced by patients and their need along with service providers, the productivity becomes a lot more challenging unless the stakeholders are well apprised of the functionalities of the healthcare domain. This calls for requirement of necessary tools, which will pave way for integrated multi stream flow of data spanning electronic health records, patient monitoring data, laboratory data, nursing operation data etc. for smoother functioning along optimal utility of resource utilization.

4. Privacy, ethics and security

These three words are very much essential in defining data. With advent of increasing data services, everyone has access to data from multiple sources where one has the liberty of combining all of them.

This has led to misuse of this. Accordingly, question arises like destination of data, user identification and motto behind use of data. Consequently, there is an utmost need of regulation. Good news is that there has been an updated General Data Protection (GDPR) replacing the old version. As per this, it is no longer required to have a national legalization. It will cater to both public and private sector bringing all organizational sectors under its domain.

5. Technical challenges

In the following, technical challenges and opportunities are discussed regarding the application of Big Data technologies in healthcare.

a) Data quality

As the term goes, quality of data is very much vital. Because of expensive processes involved in medical and pharmaceutical sector, the reliability and reproducibility are two stringent measures. Hence, it is dealt with caution that how data is generated, executed and transformed before readying them for storing. With up gradation of analytical methods and complexity of operations, source of data is extremely important as they can significantly affect the conclusion.

b) Data quantity

After quality comes another vital part data quantity. As stated earlier, BDA is actually dependent on voluminous data. Since it spans clinical, genetic, behavioral, environmental, financial and operational data, hence, there should be effective mechanism to tackle such big wealth of information so as to retrieve valuable insights towards improvement of healthcare in terms of quality and efficiency. This will ensure not only optimization of existing products and services but also propositions of new rules.

6. Recent application of IoT driven BDA

Since the outbreak of Covid-19 in Wuhan, China, the pandemic as declared by WHO has shattered the global scenario. Most part of the world has been undergoing the lock down phase. The SARS-COV-2 virus has caused fatalities amounting to 150 K worldwide and the no. is increasing each day. Apart from that, most of the countries are also undergoing economic turmoil. All big organizations such as FDA, CDC, USA are at a fix to find out the best possible vaccine as well as effective medicine in order to combat this Covid-19 pandemic. In this context, IoT driven BDA has been widely used by health professionals to find out best remedies in the fight against Covid-19. Accordingly, we have complied two very recent application of IoT enabled BDA.

Supercomputer Taking part in data analysis part for Covid-19

We are all aware of the much hyped IBM' Blue Gene supercomputer. Such was the power of its computation skill that it efficiently surpassed the petascale barrier around sixteen years ago. Based on that, this supercomputer played a crucial role in analyzing the sequencing of human genome, thereby paving way for designing novel drugs and treatments. It has also successfully simulated one percent of the

most complex machine of the earth; human brain. Ideally, this supercomputer has been destined for such complex computing process. However, the Covid-19 since its outbreak is spreading very fast and it has infected two million people globally. As a result, this pandemic situation has to be tackled efficiently. Accordingly, the Dept of Energy of United States of America, which has been severely affected by this Covid-19, deployed a powerful ally-the IBM-built Summit supercomputer in combating COVID-19 [19]. In general, the infection of cells by virus is caused by an injection of spike to genetic material belonging to the host cells. The practitioners in wet labs patiently inspect the reaction of the microorganism in response to new compounds. Although, this kind of practices although bears fruit; in terms of time and handling, they may prove cumbersome. In such cases, computer simulations will not only save ample amount of time but also provide unique solutions. In the cautionary aspect, these simulations being endowed with the ability of analyzing reactions of different virus corresponding to several variables handle terabytes of data pertaining to each variable. Consequently, the multiple simulations emerge out to be a very tedious process in the context of time as well as hardware involved. In order to integrate all computing hardware, IoT enabled BDA seems to be the only savior. Accordingly, through an efficient IoT enabled BDA scheme, Summit is helping researchers to simulate approximately 8,000 compounds in a short span of time. The main goal was to identify the optimum model for constraining the infections of SARS-COV-2. As such, seventy-seven small-molecule compounds, such as medications and natural compounds, have been identified; exhibiting ability of weakening COVID-19's ability to dock with and infect host cells. Although, these finding are not a direct mean to cure the viral disease; however, the results will be a direct boon for future studies and provide a basis such that experimentalists will deploy these compounds and find the best suitable one as a potential tool for mitigating the COVID-19. The modeled protein by Summit is illustrated below in Figure 2
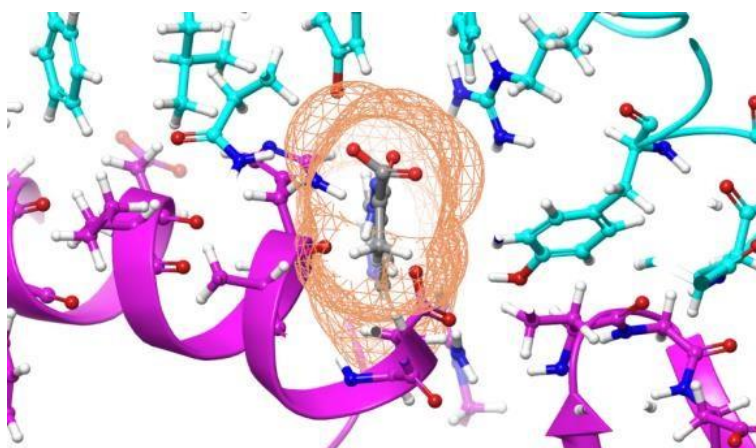
.

Figure 2. Computed compound (grey color) with the SARS-Cov-2 spike protein (represented in cyan color), thus prohibiting docking with the Human Angiotensin-Converting Enzyme 2, shown in purple. (Courtesy: Micholas Smith/Oak Ridge National Laboratory, U.S. Dept. of Energy)[19]

Innovative Search engine in the framework of BDA for COVID-19

BDA in this time of epidemic (Covid-19) has achieved another feat. Throughout the world, researchers are in a process of digging out all articles related to Covid-19 with the objective of having ample information about the virus activities and eventual curative measures. Accordingly, they accrued 50K articles, which is a mammoth data. In order to make optimum use of it, one enterprise named Verizon Media has made a praiseworthy task. Utilizing Vespa, being an open-source, big data processing program, they develop BDA based a coronavirus academic research search engine: CORD-19 Search [20].
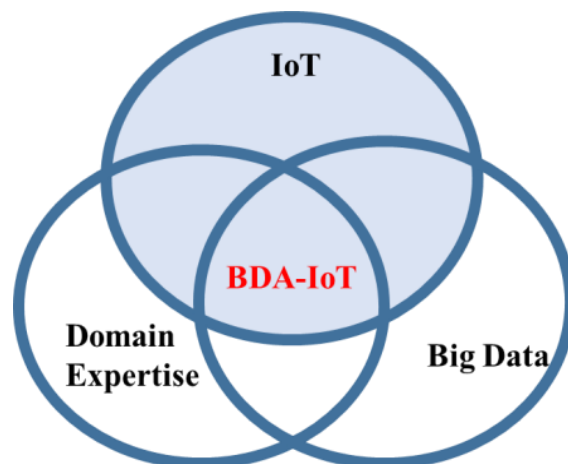

Figure 3. A visionary approach for IoT and BDA

Accordingly, this engine makes best use of the COVID-19 Open Research Dataset (CORD-19) with the sole goal of assisting medical researchers. These medical researchers are expected to able to get better insight in combating SARS-CoV-2. The repository gets piling with the advent of new peer-reviewed research publications. Apart from this, the archival platforms such as arxiv, bioRxiv, printxiv, medRxiv give another boon to the repository. Additionally, the repository entails documents being linked to publication databases such as PubMed, SAGE, Microsoft Academic, and the WHO COVID-19 database of publications.

The core of this innovative search engine is the utilization of semantic similarity through scibert-nli model. This model is referred to as a pre-trained data-mining language model with the capability of searching scientific text with utmost efficiency. Precisely, Vespa exploits text and structured search synergistically. Apart from this, it is equipped with article recommendation, user personalization, ad targeting; which is further accompanied by application programming interface (API) for advanced users. Because of such novel features, it is proving very beneficial to the researcher saving ample time in the search of oceanic volume of Covid-19 articles.

7. Conclusions and Recommendations

Herein, the characteristics of IoT oriented BDA in healthcare are comprehensively highlighted. It can be seen that BDA helps in rendering far-reaching, targeted and cost-effective health care. It can be understood that BDA cannot be fully exploited until and unless there is any targeted research endeavor. Proper access and quality of big data are some impending challenges. There is a need to explore appropriate and effective ways , which is in synergy with privacy and ethical principles, to monitor big data so that one can have a deeper insight in understanding the objectives of implementation and quality which will eventually lead to developed and optimized care processes, early diagnosis etc. Figure 3 shows a visionary approach for IoT, BDA and domain expertise. As can be seen, it is imperative to have a multidisciplinary field, which positions itself at the intersection of internet of things (sensors and networks), big data technology, and domain specific analytics. This will eventually lead to a combination of the tools and methods aiding transformational solutions for industries, government, and society.

Declaration of Competing Interest:

Author declares no competing interest.

REFERENCES

[1]. Akgül, Ceyhun Burak, et al.(2011), "Content-based image retrieval in radiology: current status and future directions." Journal of Digital Imaging, 24(2), pp. 208- 222.

[2]. Bertsimas, D, Bjarnadóttir, M, Kryder, J, Pandey, R, Vempala, S, Wang, G. (2008), "Algorithmic prediction of health-care costs", Operations Research, 56(6), pp. 1382-1557.

[3]. Bhattacherjee, A & Hikmet, N, (2007), "Physicians' resistance toward healthcare information technology: a theoretical model and empirical test", European Journal of Information Systems: Including a Special Section on Healthcare Information, 16, pp. 725–737.

[4]. LeRouge, C., Mantzana, V. & Vance Wilson, E. (2007), 'Healthcare information systems research, revelations and visions', European Journal of Information Systems, 16, pp. 669 – 671.

[5]. Müller, Henning, et al. (2004) "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions." International journal of medical informatics 73(1), pp.1-24.

[6]. Shmueli, G., & Koppius, O. (2011), 'Predictive analytics in information systems research', MIS Quarterly, 35(3), pp. 553-572.

[7]. Sarkar, Indra Neil.(2010), "Biomedical informatics and translational medicine."

Journal of Translational Medicine, 8(1), 22.

[8] Walker, M., Fonda, SJ., Salkind, S. & Vigersky, R. (2012), "Advantages and

disadvantages of real-time continues glucose monitoring in people with type 2 diabetes", Blood glucose monitoring, 8(1), p. 25.

[8]. Hao Wang, Ottar L. Osen, Guoyuan Lit, Wei Lit, Hong-Ning Dai, Wei Zeng, Big   Data   and Industrial Internet of Things for the Maritime Industry in Northwestern   Norway, 10.1109/TENCON.2015.7372918, TENCON 2015 - 2015 IEEE Region Conference

[9]. Zheng Li , Diego Seco and Alexis Eloy Sánchez Rodríguez, (2019), "Microservice-   Oriented Platform for Internet of Big Data Analytics: A Proof of Concept", Sensors 9 19, 1134

[10].   Darabseh, A.; Al-Ayyoub, M.; Jararweh, Y.; Benkhelifa, E.; Vouk, M.; Rindos, A. SDStorage: A Software Defined Storage Experimental Framework. In Proceedings of the 3rd International Conference on Cloud Engineering (IC2E 2015),

[11].   Tempe, AZ, USA, 9–13 March 2015; IEEE Computer Society: Tempe, AZ, USA, 2015; pp. 13 341–346.

[12].   Djedouboum, A.C.; Ari, A.A.A.; Gueroui, A.M.; Mohamadou, A.; Aliouat, Z. Big Data Collection in Large-Scale Wireless Sensor Networks. Sensors 2018, 18. 4474.

[13].   Kang, J.M.; Bannazadeh, H.; Rahimi, H.; Lin, T.; Faraji, M.; Leon-Garcia, A. Software-Defined Infrastructure and the Future Central Office. In Proceedings of the 2nd Workshop on Clouds Networks and Data Centers, Budapest, Hungary, 9– 13 June 2013; IEEE Press: Budapest, Hungary, 2013; pp. 225–229.

[14].   Kang, J.M.; Lin, T.; Bannazadeh, H.; Leon-Garcia, A. Software-Defined Infrastructure and the

SAVI Testbed. In TridentCom 2014: Testbeds and Research Infrastructure: Development of Networks and Communities; Leung, V.C., Chen, M., Wan, J., Zhang, Y., Eds.; Springer: Cham, Switzerland, 2014; 137, pp. 3–13.

[15].   Li, C.S.; Brech, B.L.; Crowder, S.; Dias, D.M.; Franke, H.; Hogstrom, M.; Lindquist, D.; Pacifici, G.; Pappe, S.; Rajaraman, B.; et al., (2014), "Software defined environments: An introduction" IBM J. Res. Dev, 58, pp.1–11.

[16].   Nunes, B.A.A.; Mendonca, M.; Nguyen, X.N.; Obraczka, K.; Turletti, T, (2014), "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks", IEEE Commun. Surv. Tutor, 16, pp. 1617–1634.

[17].   Villari, M, Fazio, M, Dustdar, S, Rana, O, Ranjan, R, (2016), "Osmotic Computing: A New Paradigm for Edge/Cloud Integration", IEEE Cloud Comput, 3, pp. 76–83.

[18].   Vögler, M.; Schleicher, J.M.; Inzinger, C.; Dustdar, S. Ahab, (2017) "A Cloud-based Distributed Big Data Analytics Framework for the Internet of Things", Softw. Pract. Exp. 47, pp. 443–454.

[19].   https://newsroom.ibm.com/US-Dept-of-Energy-Brings-the-Worlds-Most-Powerful-Supercomputer-the-IBM-POWER9-based-Summit-Into-the-Fight-Against-COVID-[Accessed   on 14/04/2020]

[20].   https://www.zdnet.com/article/verizon-introduces-open-source-big-data-coronavirus-search-engine/[Accessed on 14/04/2020]