

# Research Issues and Challenges on Big Data: A Survey

ISSN (e) 2520-7393  
ISSN (p) 2521-5027  
Received on 25th Aug 2020  
Revised on 20th Sept, 2020  
www.estirj.com

Muhammad Jameel Arshad<sup>1</sup>, Adnan Ali<sup>2</sup>, Irshad Ahmed Sumra<sup>3</sup>, Ahmad Naeem Akhtar<sup>4</sup>

<sup>1,2</sup>Department of Computer Science, Bahria University Lahore.

<sup>3,4</sup>Department of Information Technology (IT), Lahore Garrison University, Lahore.

Email: [jameelasmi127@gmail.com](mailto:jameelasmi127@gmail.com)<sup>1</sup>, [adnan.ali001@gmail.com](mailto:adnan.ali001@gmail.com)<sup>2</sup>, [irhadahmed@lgu.edu.pk](mailto:irhadahmed@lgu.edu.pk)<sup>3</sup>, [drnaeem@lgu.edu.pk](mailto:drnaeem@lgu.edu.pk)<sup>4</sup>

**Abstract:** In advanced world, information is produced from different sources, because of this advancement in everywhere the big data growth is the main big issue. It is analyzed from all fields with the large amount of data. The big data are structured, semi structures and unstructured forms in different bytes like terabyte, petabytes and hex bytes. The big data create through 5V's which refers to volume, velocity, variety, value and the last one is veracity. In this paper, authors provide a survey on big data and its challenges and also discuss in detail the security issue related to big data processing. It is also explain the different types of tools which are used to process the data in different types of network.

*Keywords:* Data, 5Vs, security issues, challenges, network.

## 1. Introduction

Generally, it is define as the daily use of data in amount of terabytes and these large data sets that cannot be possible to process through old database management systems tools and application. The big data are structured, semi structures and unstructured forms in different bytes like terabyte, petabytes and hex bytes. The big data create through 5V's which refers to volume, velocity, variety, value and the last one is veracity. Volume is huge data which is generated every day every minute in our daily life and velocity is the rate of growth of the data. Variety it is the data is semi structured, structured and UN structured. The Figure 1 describes the 5V's of big data. The best objectives of big data analysis to process the data in high volume, velocity, variety, volume and veracity these through old and estimating techniques [1]. The various these data extraction methods to get the helpful information was explain by Gandomi and Haider [2]. So the big data is basically is a problem and the definition is not exact about big data but in future it very helpful for decision making. In future possible the amount of big data is increase from 25 billion to more and more [3]. The context of information and communication technology, in today's era the big data has powerfully motivation to upcoming generation of intelligence technology in industries and many other fields [4]. It is most powerful to next generation in technology and information. It refers thought the big data, cloud computing and internet the data is stored and generate from thing it is used, social websites and for storage the is managed by data warehouses the big issue in it data is how we can extract the useful and best knowledge from the big data sets. The epistemological is need of effects and the

data is the key of evolution [5]. The most of the approaches is not enough to handle the large data in data mining .the problem is generate when the tools getting no understanding in database systems and data analysis .but the solution is we decide and make the best pattern for the specific and big problems that really help us to solve the problem in any way. on the other hand also know the problems pattern and design algorithms the representation form of the problem describe all its abstract that easily guide us where we work and which pattern we following data [6] [7] [8]. So, all the data which in big data are not useful for analysis in data and decision making on unrefined big data. The section 2 describes the big data challenges and section 3 discusses in detail the Big Data analytics in cloud computing environment. The section 4 discusses in detail the tools use in big data environment and section 5 concludes the paper.

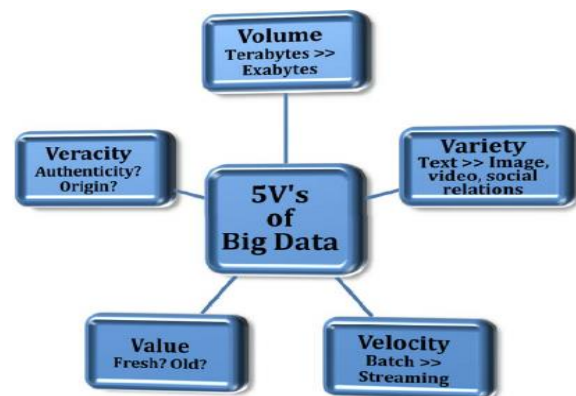


Figure 1. 5V's of Big Data [39]

<sup>2</sup>Electronic Engineering Department Mehran University Jamshoro,

<sup>3</sup>Department of Information and Communication Technology IIU Malaysia,

<sup>4</sup>Department of Electrical and Electronic Engineering Universiti Teknologi PETRONAS, Malaysia,

## 2. Challenges in Big Data Analytics

From the past few years the data is collected from such domains like, health care, public administration, retail, biochemistry and other scientific researches. The web applications face big data frequently such that internet text, social computing, documents and internet search. Social computing have social networks and analysis system, online communities for data analysis, data reputation systems and data prediction markets. The big data provides a great advantage as well with new opportunities for large data to knowledge in processing tasks while the new researches arrives in future. Opportunities gets from some challenges which we follow to get that, while handle the challenges we need to know about the different methods to analyze the big data. Like, some methods that do well with the small data size not for the large scale volume data .as well with computational techniques in small data. For the field of health most of the researchers work for many challenges [9] [10]. The challenges of big data analytics are divided into four categorize:

- knowledge discovery
- computational complexity
- scalability
- visualization

### 2.1. Data Storage and Analysis

In past few years the data is increase as the technology and usages of it .data is increase and now its new name big data analysis which is generate by mobile devices , radio frequency waves , sensors and technologies etc. these data are stored if they have enough space or ignore and deleted if they not have a space to store. So the first challenges for us are storage of big data and also give higher performance in input/output data in speed. In this case the main priority in data is its accessibility for knowledge discovery and data representations. In the past researches the data analysts use hard drives to store the data but it slower the working and performance to overcome this problem the two things solid state drive (SSD) and phase change memory (PCM) was developed to use them for this problem. The second challenge in big data analysis is its attribute diversity of data .with the raped growth of the data related to data sets the data mining tasks also increase .when we deal through the huge data sets we use the data reduction techniques, data selection frame works and feature selection is the most important tasks for it .we use these advanced techniques because the old and existing techniques and algorithms may not always works proper on large data sets on time. In big data exclude clustering they help to analysis best in big data [11]. So, it is trying to develop and make new machines learning algorithms for consisting the last years data is based on consisting so it is the major in big data. Through the advance technologies that are hadoop and reduce map its make us possible to get the large amount of structure and semi structure data in less time. Das and kumar [12] describe to use a framework to analyze in data. The result of analysis the data is also shared with details on its paper through tweets from Das and Ali [13]. For this it use the process to convert the semi structured and unstructured data

into structured data, and after that we applying the data mining algorithms to extract the knowledge from the large data .so in these the most important is to design storage systems and use efficient data tools to well sure provide accurate data when the data comes from the different sources. However, it is sure to design the machine learning algorithms for analyze the data for improve efficiency and scalability in the data.

### 2.2. Knowledge Discovery and Computational Complexities

Knowledge discovery and computational complexities these are the main issues in big data. It has a several sub fields too. There are many tool to use in knowledge discover and representation these are fuzzy set [14], rough set [15], soft set [16], near set [17] and formal sets/concept analysis [18] ,principal component analysis [19]. Many other techniques are helpful to solve real life problems. But the techniques work according to the problem. Moreover some techniques not for the large data set in sequence. On the other hand while the growth of the big data increase very speedily then the available and current working tool is not efficient to get the meaningful information through the data. The best approach to management the data warehouses. Because data ware house is the responsible to store all the data they got through the different resources. Computational complexities are most of the requirement to analysis the large dataset. So, the major issue is here that how we handle the inconsistent and irregular data which is present in the data. Many researches is going on and about to use for solve these regular problems. While analysis the large data it of sure we must use computational complexities. Many survey and research projects has use machine learning techniques with low memory. So the basic objective in the paper is to reduce the computational cost and complexities [20] [21] [22].

### 2.3. Scalability and Visualization of Data

This is one of the most important challenges in big data analysis. The increase of big data day by day as faster as the CPU speed, then according to the data increasing the CPU's core is also embedded faster than the data produce [23]. Online markets like amazon flip kart and e-bay have billions of users who use daily and sell and buy goods through it this generate a lot of data .some companies use many software tools like tableau for big data visualization. This software have capacity to convert a huge and complex data into easy and pictorial representation form. This is very helpful to manage the company's for check and balance, monitor the customers and feedback, sell/purchase. But unfortunately currently the tool for big data analysis is very bad in use and gives the poor performance and functionalized in every response time.

### 2.4. Information Security

In big data most of the data are correlated, analysis and mined in understanding patterns. All companies and organization have different rules to safe their company's information from all over the world either info related to company and customers. Because it is the most key factor element in big data. Big data have a huge security issues. So that why information security is becoming major

problem [23]. Security is managed with great techniques and security algorithms like authorized new advanced techniques is the one who only accessed their account example; encryption/decryption codes. Here we have challenge to build a multi-level security /privacy models for big data. various data applications face the network scale , many different devices , real time security , monitory and lack of instruction [25] [26].

### 3. Cloud Computing for Big Data Analytics

Virtualization technology is the great built to make data computing is accessible and less costly. Enormous Information and distributed computing advances are produced with the significance of building up a versatile and on interest accessibility of assets and information. Distributed computing fit monstrous information by on interest access to configurable figuring assets through virtualization procedures. allotted computing are mentioned in detail by way of numerous scientists which capabilities the problems in data the executives, information collection and speed, records stockpiling, records making ready, and asset the board [29], [30]. So, it allotted computing allows in constructing up a plan of action for all assortments of utilizations with foundation and devices. Big information software using allotted computing need to bolster statistics logical and development moreover, allotted computing need to likewise empower scaling of apparatuses from virtual improvements into new advancements like sparkle, R, and one of a kind types of great facts dealing with systems. Massive data shapes a shape for analyzing distributed computing options. Contingent upon specific need, consumer can visit the industrial middle and buy basis administrations from cloud expert agencies, as an instance, Google, Amazon, IBM, programming as an administration (SaaS) from an entire group of organizations, for instance, NetSuite, Cloud9, Job science and so on. Some other favorable role of allotted computing is distributed storage which offers a viable technique to putting away massive facts. The undeniable one is the time and fee that are predicted to switch and download big facts inside the cloud circumstance. It finishes up difficult to manipulate the dissemination of calculation and the basic system. Be that as it is able to, the serious troubles are safety concerns figuring out with the facilitating of facts on open servers, and the ability of data from human examinations. Each such an issue will take big facts and allotted computing to an ordinary state of improvement.

#### 3.1. Bio-inspired Computing for Big Data Analytics

Bio-roused registering is a device enlivened my nature to deal with complicated certifiable problems. Natural frameworks are self-composed without a focal manage. A bio-roused cost minimization tool hunt and find the ideal information management association on considering cost of facts the board and management help. Those systems are produced through common atoms, for instance, DNA and proteins to direct computational counts which include placing away, recuperating, and managing of data expansions of advances are developing like big facts, IoT, allotted computing, bio inspired figuring and so on while concord of information need to be possible just through selecting right level to interrupt down large and outfit

financially savvy outcomes. Bio-propelled processing techniques fill in as a key process in astute information investigation and its software to huge information. those calculations assist in appearing records digging for massive datasets due to its advancement software. The maximum favorable function is its effortlessness and their rapid meeting to best arrangement [31] even as looking after administration arrangement issues. Some applications to this stop using bio propelled registering were pointed out in element through Cheng et al [32]. From the dialogs, we can see that the bio-propelled processing fashions provide extra astute communications, unavoidable records misfortunes, and help is handling ambiguities. Henceforth, it's far trusted that in future bio-enlivened figuring might also help in looking after big facts to a huge diploma.

#### 3.2. Quantum Computing for Big Data Analysis

A quantum laptop has memory that is exponentially larger than its physical length and may manage an exponential schedule of records assets at the same time [33]. This exponential enhancement in laptop frameworks can be possible. Inside the event that a true quantum computer is available now, it can have tackled problems which are quite difficult on ongoing desktops, surely current large data issues. The major functional problem in create quantum pc may want to earlier than lengthy be doable. Quantum figuring offers an method to mix the quantum mechanics to procedure the facts. In normal computer, facts is displayed by lengthy collection of bits which encode both a zero or a one. on the other hand a quantum pc makes use of quantum bits or qubits. The comparison mid qubit and bit is that, a qubit is a quantum framework that encodes the 0 and the only into discernable quantum states. Thusly, it favor to be won by using the wonders of superposition and entrapment. It implies that various sizeable data issues can be tackled loads quicker by using larger scale quantum pcs contrasted and set up computers. Finally it is a check for this age to manufacture a quantum laptop and encourages quantum figuring to attend to big data troubles.

### 4. Tools for Big data Processing

In this segment, it examine some present procedures for breaking down enormous information with accentuation on three vital developing instruments in particular MapReduce, Apache Sparkle, and Tempest. The common work stream of huge information venture talked about by Huang et al is featured in this segment. For instance Dermal and Apache Drill are the huge information stages that help intelligent examination. An awesome rundown of huge information instruments and methods is additionally examined by many analysts [6],[34]. The greater part of the accessible apparatuses focus on cluster preparing, stream processing, and intelligent investigation. These devices help us in building up the huge information ventures. Extensive quantities of devices are accessible to process enormous information. Stream information applications are for the most part utilized for constant systematic. Most bunch preparing apparatuses depend on the Apache Hadoop framework, for example, Mahout and Dryad.



#### 4.1. Apache Hadoop and MapReduce

The master node divides the input into smaller sub problems and then distributes them to worker nodes in map step. Moreover, Hadoop and MapReduce works as a powerful software framework for solving big data problems. Hadoop works on two kinds of nodes such as master node and worker node. The most established software platform for big data analysis is Apache Hadoop and Mapreduce. It consists of hadoop kernel, mapreduce, hadoop distributed file system ( HDFS ) and apache hive etc. Thereafter the master node combines the outputs for all the sub problems in reduce step.

#### 4.2. Apache Mahout

Center calculations of mahout including bunching, arrangement, design mining, relapse, dimensionality decrease, developmental calculations, and cluster put together shared sifting keep running with respect to best of Hadoop stage through guide diminish system. Apache Mahout Apache mahout plans to give adaptable and business machine learning methods for substantial scale and canny information examination applications. The objective of mahout is to construct an energetic, responsive, assorted network to encourage talks on the undertaking and potential use cases.

#### 4.3. Apache spark

Apache Flash Apache flash is an open source enormous information preparing structure worked for speed handling, and advanced investigation. Flash comprises of segments in particular driver program, bunch administrator and laborer hubs. The driver program fills in as the beginning stage of execution of an application on the flash bunch. Flash keeps running over existing Hadoop circulated document framework (HDFS) foundation to give improved and extra usefulness. The real favorable position is that it offers help for sending flash applications in a current Hadoop groups. The prime focal point of flash incorporates flexible dispersed datasets (RDD), which store information in-memory and give adaptation to non-critical failure without replication. It is written in scale programming language and keeps running on java virtual machine (JVM) condition. Another advantage is that, a client can run the application program in various dialects, for example, Java, R, Python, or Scala. The principal advantage is that notwithstanding MapReduce, it likewise bolsters gushing information, machine learning, and chart calculations. The spark runs an application in Hadoop group, up to multiple times quicker in memory, and multiple times quicker when running on plate.

#### 4.4. Dryad

It is another popular programming model for implementing parallel and distributed programs for handling large context bases on dataflow graph. A dryad application runs a computational directed graph that is composed of computational vertices and communication channels [34].

#### 4.5. Storm

Radiance is accountable for disseminating code over the tempest group, planning and doling out assignments to laborer hubs, and observing the entire framework. On tempest group clients run distinctive topologies for various tempest assignments though hadoop stage actualizes map lessen occupations for comparing applications. Storm Tempest is a disseminated and blame tolerant constant calculation framework for handling expansive gushing information. A tempest bunch comprises of two sorts of hubs, for example, ace hub and specialist hub. The tempest bunch is obviously like hadoop group.

#### 4.6. Apache drill

Likewise it has a goal to scale up on 10,000 servers or more and achieves the capacity to process patabytes of information and trillions of records in a flash. Apache drill is another dispersed framework for intelligent investigation of huge information. It has greater adaptability to help numerous sorts of question dialects, information organizations, and information sources.

#### 4.7. Jasper soft

It additionally have a capacity to assemble ground-breaking hypertext markup language (HTML) reports and dashboards intuitively and specifically from huge information store without ETL necessity. It is an adaptable enormous information investigative stage and has an ability of quick information representation on prevalent capacity stages, including MangoDB, Cassandra, Redis and so on.

#### 4.8. Splunk

Its trait include ordering organized, unstructured machine produced facts, ongoing looking for, revealing logical outcomes, and dashboards. Splunk is a regular and insightful level produced for misusing machine created great records. The most critical intention of Splunk is to present matrices to various software, examine problems for framework and facts innovation foundations, and insightful help for commercial enterprise sports. It encourages purchaser to seek, display screen, and dissect their system created records through internet interface. As of past due a superb deal of data are created through system from business organizations. The transformation of facts into expertise is by no means an clean undertaking for high performance large-scale facts processing, consisting of exploiting parallelism of modern and upcoming pc architectures for statistics mining. The amount of facts accrued from numerous programs all over the world throughout a huge sort of fields today is predicted to double every years. The development of powerful computer systems is a boon to implement those techniques main to automatic systems. This necessitates the improvement of strategies which can be used to facilitate large facts analysis. Many distinct models like fuzzy units, rough sets, gentle units, neural networks, their generalizations and hybrid models obtained through combining or more of those fashions were located to be fruitful in representing information, greater importantly, these new challenges might also include, on occasion even become worse, the performance, efficiency and scalability of the devoted statistics in depth computing systems. Expressing statistics

get entry to requirements of applications and designing programming language abstractions to exploit parallelism are a direct want [38]. More regularly than no longer, huge statistics are decreased to include simplest the important traits vital from a selected examine point of view or depending upon the application location.

## 5. Conclusion

The different strategies used for the analysis consist of statistical evaluation, system mastering, information mining, intelligent evaluation, cloud computing, quantum computing, and information movement processing. To this lead to this paper, we survey the diverse research troubles, challenges, and tools used to analyze these big statistics. it believe that during destiny researchers will pay more interest to these techniques to clear up issues of large statistics efficiently and efficiently. From this survey, it is understood that each big statistics platform has its man or woman attention.

## References

- [1] M. K.Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, *International Journal of Application or Innovation in Engineering & Management*, 2(8) (2015), pp.228-232.
- [2] A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35(2) (2015), pp.137-144.
- [3] C. Lynch, Big data: How do your data grow?, *Nature*, 455 (2008) , pp.28-29.
- [4] X. Jin, B. W.Wah, X. Cheng and Y. Wang, Significance and challenges of big data research, *Big Data Research*, 2(2) (2015), pp.59-64.
- [5] R. Kitchin, Big Data, new epistemologies and paradigm shifts, *Big Data Society*, 1(1) (2014), pp.1-12.
- [6] C. L. Philip, Q. Chen and C. Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on big data, *Information Sciences*, 275 (2014), pp.314-347.
- [7] K. Kambatla, G. Kollias, V. Kumar and A. Gram, Trends in big data analytics, *Journal of Parallel and Distributed Computing*, 74(7) (2014) , pp.2561-2573.
- [8] S. Del. Rio, V. Lopez, J. M. Bentez and F. Herrera, On the use of mapreduce for imbalanced big data using random forest, *Information Sciences*, 285 (2014), pp.112-137.
- [9] MH. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki and D. K. Grunwell, Health big data analytics: current perspectives, challenges and potential solutions, *International Journal of Big Data Intelligence*, 1 (2014), pp.114-126.
- [10] R. Nambiar, A. Sethi, R. Bhardwaj and R. Vargheese, A look at challenges and opportunities of big data analytics in healthcare, *IEEE International Conference on Big Data*, 2013, pp.17-22.
- [11] Z. Huang, A fast clustering algorithm to cluster very large categorical data sets in data mining, *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, 1997.
- [12] T. K. Das and P. M. Kumar, Big data analytics: A framework for unstructured data analysis, *International Journal of Engineering and Technology*, 5(1) (2013), pp.153-156.
- [13] T. K. Das, D. P. Acharjya and M. R. Patra, Opinion mining about a product by analyzing public tweets in twitter, *International Conference on Computer Communication and Informatics*, 2014.
- [14] L. A. Zadeh, Fuzzy sets, *Information and Control*, 8 (1965), pp.338-353.
- [15] Z. Pawlak, Rough sets, *International Journal of Computer Information Science*, 11 (1982), pp.341-356.
- [16] D. Molodtsov, Soft set theory first results, *Computers and Mathematics with Applications*, 37(4/5) (1999), pp.19-31.
- [17] J. F.Peters, Near sets. General theory about nearness of objects, *Applied Mathematical Sciences*, 1(53) (2007), pp.2609-2629.
- [18] R. Wille, Formal concept analysis as mathematical theory of concept and concept hierarchies, *Lecture Notes in Artificial Intelligence*, 3626 (2005), pp.1-33.
- [19] I. T.Jolliffe, *Principal Component Analysis*, Springer, New York, 2002.
- [20] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis and K. Taha, Efficient machine learning for big data: A review, *Big Data Research*, 2(3) (2015), pp.87-93.
- [21] Changwon. Y, Luis. Ramirez and Juan. Liuzzi, Big data analysis using modern statistical and machine learning methods in medicine, *International Neurourology Journal*, 18 (2014), pp.50-57.
- [22] P. Singh and B. Suri, Quality assessment of data using statistical and machine learning methods. L. C.Jain, H. S.Behera, J. K.Mandal and D. P.Mohapatra (eds.), *Computational Intelligence in Data Mining*, 2 (2014), pp. 89-97.
- [23] A. Jacobs, The pathologies of big data, *Communications of the ACM*, 52(8) (2009), pp.36-44.
- [24] H. Zhu, Z. Xu and Y. Huang, Research on the security technology of big data information, *International Conference on Information Technology and Management Innovation*, 2015, pp.1041-1044.
- [25] Z. Hongjun, H. Wenning, H. Dengchao and M. Yuxing, Survey of research on information security in big data, *Congresso da sociedade Brasileira de Computacao*, 2014, pp.1-6.
- [26] I. Merelli, H. Perez-sanchez, S. Gesing and D. D.Agostino, Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives, *BioMed Research International*, 2014 , (2014), pp.1-13.
- [27] N. Mishra, C. Lin and H. Chang, A cognitive adopted framework for iot big data management and knowledge discovery prospective, *International Journal of Distributed Sensor Networks*, 2015, (2015), pp. 1-13
- [28] X. Y.Chen and Z. G.Jin, Research on key technology and applications for internet of things, *Physics Procedia*, 33, (2012), pp. 561-566.
- [29] M. D. Assuno, R. N. Calheiros, S. Bianchi, M. a. S. Netto and R. Buyya, Big data computing and clouds: Trends and future directions, *Journal of Parallel and Distributed Computing*, 79 (2015), pp.3-15.

- .....
- [30] I. A. T. Hashem, I. Yaqoob, N. Badrul Anuar, S. Mokhtar, A. Gani and S. Ullah Khan, The rise of big data on cloud computing: Review and open research issues, *Information Systems*, 47 (2014), pp. 98-115.
  - [31] L. Wang and J. Shen, Bioinspired cost-effective access to big data, *International Symposium for Next Generation Infrastructure*, 2013, pp.17.
  - [32] C. Shi, Y. Shi, Q. Qin and R. Bai Swarm intelligence in big data analytics, H. Yin, K. Tang, Y. Gao, F. Klawonn, M. Lee, T. Weise, B. Li and X. Yao (eds.), *Intelligent Data Engineering and Automated Learning*, 2013, pp.417-426.
  - [29] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, New York, USA 2000.
  - [34] M. Herland, T. M. Khoshgoftaar and R. Wald, A review of data mining using big data in health informatics, *Journal of Big Data*, 1(2) (2014) , pp. 1-35.
  - [35] T. Huang, L. Lan, X. Fang, P. An, J. Min and F. Wang Promises and challenges of big data computing in health sciences, *Big Data Research*, 2(1) (2015), pp. 2-11.
  - [36] G. Ingersoll, Introducing apache mahout: Scalable, commercial friendly machine learning for building intelligent applications, White Paper, IBM Developer Works, (2009), pp. 1-18.
  - [37] H. Li, G. Fox and J. Qiu, Performance model for parallel matrix multiplication with dryad: Dataflow graph runtime, *Second International Conference on Cloud and Green Computing*, 2012, pp.675-683.
  - [38] D. P. Acharjya, S. Dehuri and S. Sanyal *Computational Intelligence for Big Data Analysis*, Springer International Publishing AG, Switzerland, USA, ISBN 978-3-319-16597-4, 2015.
  - [39] Kaynak, Okyay & Yin, Shen *Big Data for Modern Industry: Challenges and Trends*. *Proceedings of the IEEE*. 103. 143-146. 10.1109/JPROC.2015.2388958, 2015.