



## International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

# Big Data Predictions, Challenges and Applications through Machine Learning and Deep Learning

Shraddha Mishra<sup>1</sup>, Mukesh Kumar Chandrakar<sup>2</sup>, Anup Mishra<sup>3</sup>

Undergraduate Student, Bhilai Institute of Technology, Durg, Chhattisgarh, India

Assistant Professor, Bhilai Institute of Technology, Durg, Chhattisgarh, India

Professor, Bhilai Institute of Technology, Durg, Chhattisgarh, India

**ABSTRACT:** With all the internet of things, internet and cloud computing, data produced in 2017 was roughly 2.5 quintillion bytes a day. That is staggering amount of data given that 7.55 billion population on an average had 2 smart devices each, as in 2017. This data consisted of all the information about the physical world and took between machine to machine, man to machine and man to man. Our paper presents what this data means to mankind, its impact and significance, optimization techniques that work with most of the problems, challenges and solutions to the challenges faced by the big data and also its applications to elevate the development of mankind, society and civilization all together.

**KEYWORDS:** Big Data, Machine Learning, Deep Learning, data analysis.

### I. INTRODUCTION

The amount of data generated each day by an individual creates an imprint of the individual in the cyberspace. This data, to be exact, has represented and also changed the identity, social, geological, and even political establishments around the world. This data emerging from all the sources can be termed as the big data. There, however, is no universally accepted fixed definition of the big data but Wikipedia defines big data as "Big data is data sets that are so voluminous and complex that traditional data-processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source." [2]. The human society through ever expanding internet of thing devices, all connected through the internet and cloud is inevitably reflecting the variables of physical world in cyberspace, embodied as big data.

It's just not the data, which used to be the collective form of the information, which used to be used for computational purposes earlier, for which computers were invented. Now computers work on knowledge as input and wisdom as processing asset. Information is whereas, now a currency. Reliance Industries Chairman Mukesh Ambani nearly doubled his wealth in two years in 2017 by giving world's second largest country India, free data services and said, "Data is the new oil and the foundation of the fourth industrial revolution is connectivity" [3].

Hundreds of science and technology journals like Nature, Science, IEEE, etc. are registering research and development papers on the topic of big data every day. This technology is now converging to something that will integrate cloud computing, artificial intelligence, Internet of Things (IoT), Big Data analytics, robotics and nanotechnology with humanity for eternity, in inseparable and progressive way.

Big data is backed by deep learning in such a way that, every sphere that needs optimization whether that be social, economical, ecological or even political, data is being transformed into the wish of the controlling entities. On the one hand where it leads to the generation of more jobs and revenue, on the other had it becomes fragile asset that needs to be protected and treasured from falling into wrong hands. The latest example being 2016 US Presidential Elections, in which Russia with the help of Cambridge Analytica, which is a British political consulting firm, combined data brokerage, data mining, and data analysis and used it on the big data sourced from the biggest social network Facebook

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 3, March 2018

to tamper the election process and even got successful in doing so. Indian political circumstances had also been affected by the same firm in past, as confirmed by CA whistleblower Christopher Wylie.

So, if the big data is that powerful and combined with machine learning techniques, especially Deep Learning, its significance, challenges and legal applications must be identified. Our paper in its first part explains the big data in technical way and its optimisation methods through deep learning. In the second part, we deal with its significance and the challenges faced by various fields in holistic way. In the last part, we aim to provide solutions for alleviating the limitations and challenges and also possible legal applications.

## II. THE BIG DATA BASICS

The word big data was coined in 2005 by Roger Magoulas from O'Reilly media. He described it as the data that is huge in magnitude and cannot be extracted, analyzed, processed and stored through the current available methods. He had then asserted that "The future belongs to them who can transform big data into products".[1][7]

The big data can be seen as collection of complex individual sets of data of various patterns and structures that are on their own capable of interacting some meaning and collectively they represent the every aspect of information administered by the data analysis. It is characterised by the 5 attributes:

1. Volume
2. Velocity
3. Variety
4. Variability
5. Veracity

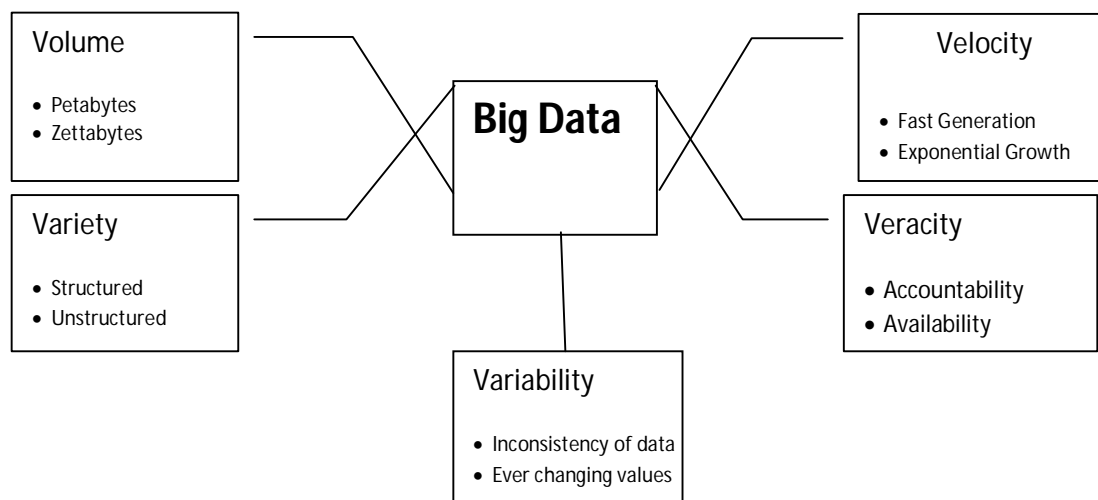


Figure 1: Attributes of Big Data: The 5 V's

These attributes not only define the big data and differentiate it from the traditional form of total data, but also suggest the problems and challenges associated with the big data.



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

Let us understand what it means by each of these attributes:

Attribute	Impact on big data
Volume	It is estimated that by 2020, there will be 35 Zettabyte of the data in cyber space. The data is understandably humongous given 1 Zettabyte= $10^{12}$ GB. This data cannot be estimated, analysed or processed, contained or stored, by any presently available data processing methods like machine learning, deep learning, etc. Also, the data contains so many dimensions of information that with the help of various theories, fields, data extraction techniques, this creates more data on processing.
Velocity	The data nowadays is being communicated on 4G/5G speeds and research on many more generations of data speed technology is still underway. To calculate and process this amount of huge data would take time and computational processes lagging the urgency of the processed data would obviously be detrimental of the data needs.
Variety	The data generated from various sources is also diverse in the forms and patterns. The data might be audio, visual, cryptic, structured or unstructured or even mixture of all. To process such complex data is challenge when it is already accompanying the huge volume.
Variability	The data makes sense only when it is consistent about the information being conveyed by the data. Big data since consists of various layers of information in huge volume being generated at exponential rates, the inconsistency of the value of information is inevitable.
Veracity	With the ever building data structures, and ever changing values how one does processes the accountability of the data? This data is coming from all sorts of sources and authenticity of any value of the data analyzing entity's interest might be compromised.

These challenges are in fact the strength of the big data processing. The multi platform sources of the data give all sorts of dimension to the data which when aided with proper data cleaning techniques and data processing methods give highly optimised solutions of data analysis. Deep learning is one of the most sought after data processing methods. Optimisation through deep learning of the big data is preferred almost every time through stochastic gradient descent methods (SGDs). But some cases also present other optimization methods superior than that of SGDs. For low dimensional problems, where the number of parameters are relatively small, for e.g., Convolutional Neural Networks, Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm is better performer. Similarly, for high dimensional problems, Conjugate gradient (CG) with line search can significantly simplify and speed up the process of pre-training deep algorithms. Also, CG and L-BFGS mostly outperform SGDs on training sparse auto-encoders. Algorithms that exploit the problem structures may perform much better when both CG and BFGS don't perform up to the mark.[5]

### III. SIGNIFICANCE AND CHALLENGES OF THE BIG DATA

Started as offspring of the initially vague technology, cloud computing, big data is a precursor of the next generation of information technology industrial revolution which is broadly built on the third platform, mainly referring to big data, internet, cloud computing, IoTs, and social media business. As it is a still developing technology, current data processing method like machine learning, deep learning, etc., and analysis tools such as data mining and statistical analysis are not adequate on their own in dealing with the processing of such data. However, it is to be noted that all data available in the form of big data is not useful for analysis or decision making process since academia and industry are interested in disseminating the findings of big data.[6] Big data is a cost effective and efficient ingredient of optimization where academia and industry scientists operate upon big data and simulations to bring out the new findings, core technologies and incorporate that in their decisions.[8]



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

An example of this market strategy can be taken from a leading online Indian fashion website Myntra. It uses machine learning techniques on the big data generated by the users on the current trends and incorporates their shopping preferences, strategies and tastes to develop the understanding of what prints, designs, cuts, styles and colors they are actually buying. They even generate clothing designs through machine learning softwares and tools to optimize their sales and profits.

The challenges to big data arise from its huge volume, high speed, large variety and tremendous variability. All these factors affect the veracity of the big data. These challenges are collectively explained as follows:

## 1. Too large volume creates analysis, processing and storage woes

Not only the general citizens are generating a lot of the data through the use of internet through their smart phones, computers, IoT devices, etc but the governments of the nations worldwide are also generating very high speed data in the form of the population surveys, defense and security data, administration data, research and development data, etc.. Industries and R&D departments have their own share of data generation too.

Now, all this data is generated at huge costs and is mostly deleted just because the available servers and storage technologies are just not sufficient to store this amount data. Also, analysis and processing is a daunting task here because so much white noise is there that what is relevant and legit and what is vague and unreliable data is a blurred line. Additionally, data selection, feature selection, data reduction, and setting up relevant parameters is an essential task especially when dealing with large datasets.[9]

Data mining becomes very time consuming and complex process because available algorithms and concepts may or may not deliver optimum results on every problem. Also, methods used for generalization and optimization vary vastly from problem to problem. And for any complex and urgent data processing, expensive hardware and strong data servers are also another necessity. The servers not only consume a lot of energy but also need investment in physical space and generate heat that is harmful to the environment.[10]

Cloud computing may provide a small solution for the storage problem but again, it is susceptible to the data breach and hacking which at no cost can be compromised.

## 2. High speed of data generated causes strain on computing processes

The storage technology, the processing speeds and the present developed methods of data computation for big data are easily outrun by the data that is being generated everyday at high velocity with different types of pattern, structure and types. Not only has this added to the complexity of the data but also by the inherent function of big data processing, it can be processed by any field of science and technology incorporating any theory, concept or algorithm.

So, with such processing, it is inevitable to cause strain on computing process and this may result in suspended working of the processing assets, limitations in optimization or compromise in data authentication, management, representation, preservation, archiving, and information retrieval.[7]

## 3. Large variety of data challenges the meaningfulness of the information

The amount of data keeps on increasing exponentially, which adds on new values to the information being carried out by the big data. Hence, the results of data processing of one batch of data may not be similar to the another batch because of the difference in the value of information carried out by them. Since, the computational process of such data is already very complex, this variability in the data aided with the large variety of the data which is hard to analyze and process makes the meaningfulness of the data very fragile entity due to the inconsistencies and uncertainty present in the datasets. This is why current big data analysis tools have somewhat poor performance in handling computational complexities due to limitation in present hardware and processing methods. More development in computational mathematics needs to be done to counter this impediment.

## IV. SOLUTIONS FOR THE BIG DATA CHALLENGES

### 1. Faster processors and storage solutions

The processor's speed of the computing device should always match the data it needs to work upon that is being generated. This makes the computing and storage of the data efficient and cost effective. Proper digital infrastructure is necessary for any competing organization or country. This can be achieved by virtualization of computing technologies.



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

Virtual computing, also known as cloud computing, has been one of the most robust big data techniques. Big Data and cloud computing technologies provide a scalable and on demand availability of data and resources. Cloud computing harmonizes all sorts of massive data by on demand access to configurable computing resources on premium or free basis.[6]

## 2. Data security and Breach Intelligence

The datasets that make up the big data are universally considered as the assets of the individual or the country generating it. That is because of the privacy and ownership of the data based on utopian concept. All the nations consider their data of national importance and protect it from being used illegally or for bad intentions. Nowadays, a nation's competency with the other countries is its technology to protect, enhance and use its big data resources and its advancement in the artificial intelligence technologies.

Security of big data is so important to the nations that they are constantly developing new data security technologies and more advanced counter-breach intelligence. Civilians are too aware of the fragility and importance of the security of their data. This big data needs to be protected from any theft, breach or snooping, for the information is then black-marketed for illegal intentions that may also pose risks to the physical insecurity.

## 3. Developments in mathematical modeling and deep learning algorithms

Deep learning techniques and big data technology go hand in hand due to relatively less complex algorithms that deliver great optimizations. However, these solutions are entirely dependent on the problems and may work on problems selectively. This limits the human understanding of dealing and manipulating the big data to its fullest potential for the good of mankind.

Hence, constant research in computational mathematics for developing the algorithms, computing, data cleaning and processing methods should be done. Everyday new simulation softwares are being developed for better understanding of the working and implementation of the data. Having expertise in multiple scientific fields also helps.

## 4. Smart IoTs and structured network architecture

Advanced digital infrastructure is now the objective of every leading and developing nation around the world. That includes fast internet, cities interconnected with fast optic wires for better data transmission, smart Iots everywhere and overall structured network architecture. Since, most of the work that governments do is uploaded and done through online platforms, this data structure must be encrypted and secured.

Knowledge acquisition from IoT data is the biggest challenge that big data professional are facing.[6] This issue can be managed through systematic and structured network architecture where all the data that is sourced from different devices is administered by the processing organizations concerned with specific analysis.

## V. APPLICATIONS OF THE BIG DATA THROUGH DEEP LEARNING

### 1. Smart internet and Cloud Computing

The new internet is restructuring the way people live, the art of doing business, global interrelations and cultural revolutions. It has even generated new methods of earning and generating revenues which are more transparent and easy to access than previous traditional ways of the same. The IT giants like Google, IBM, Apple, etc have invested heavily into the artificial intelligence and machine learning advancements. They are creating smarter internet through the use of big data processing methods like deep learning in image classification, speech recognition, language translation, language processing, sentiment analysis, recommendation through the Natural Language Processing methods and algorithms.

### 2. Medicine and Biology

The biggest achievement of big data application through deep learning is the advancement in medicine, pharmacy and biological studies. With the help of artificial intelligence, humans have achieved the impeccable feats of medical advancements in both surgical and targeted medication. Stem cell treatment's advanced applications, cancer cell detection, diabetic grading, drug discovery, individual's customized treatment suiting to his body are some of the few areas where deep learning has immensely benefitted the mankind. [12]



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 3, March 2018

### 3. Healthcare

Incorporation of this technology with medical equipments delivers treatment with clinical and surgical efficiency. Much more research is going on in this area, and this application requires the expertise of the medical and technical science fields. Also the mortality rates due to late detection, operation risks and inevitable possible human errors can be controlled if big data from around the world about all the possible medical cases is processed through machine learning techniques and incorporated in healthcare.[12]

### 4. Media and entertainment

Entertainment is one of the professions that is audience targeted and generates huge job opportunities and revenues. The big data generated by the individuals indicates about their preferences, interests and also moods. Companies utilize this data to develop and suggest entertainment that suits their tacit demands which is more likely to succeed and generate greater profits. Also the technology is used in video captioning, real time translation and smart video search. This way viewers of different languages can watch the videos of the languages they don't understand, hence, increasing the audience base of the entertainment providers increasing their revenues.[7]

### 5. Security and Defense

Nowadays, personal security and surveillance is taken care of by IoT devices. Also, national defense is majorly monitored by big data of aerial or satellite images and through the data analytics. All this mechanism is closely guarded by law enforcement departments and is responsible for the law and order of the concerned organization or country. Face detection, video surveillance and satellite imagery is mostly administered through artificial intelligence methods and are managed through cloud computing.

### 6. Automation

Deep learning applications provide huge deal of control over the electrical devices, intelligent systems and anything that can be connected to internet or network. This might be employed to save the energy, work more efficiently, to work in cost effective or precise way. Many industries employ automation to reduce their work force, do work that needs programming, hard for a human to do or is repetitive. Automation in initial stages of the production controls the quality of products and enhances the security of work personnel, refines the product being produced and increases the efficiency of the process.

Nowadays, the concept of smart homes and auto driving cars indicates the future of our living standards. All this technology is being integrated with our lives in seamless ways. This way, humans have more time to do other important things that need not necessarily be done by them.

### 7. Advancement of research fields

So, with the extra time that automation saves, intellectual millenials are doing the things that make their future more advanced. The time that technology saves and the help that it provides is being utilized in development of further new technologies and researches. The new way of research and development now doesn't depend upon the known fields to the individual. Now, through internet, big data, machine learning and simulation softwares, any field can be researched upon and can be merged with the knowledge of the other fields. The discoveries aren't dependent upon knowledge of theories and success of experiments, but on the simulations guided by the big data along with the machine learning and deep learning techniques.

Newer fields are being developed every half decade and with the further advancement in IoT devices, cloud computing, big data analytics and machine learning and deep learning technology, research fields correlate and merge like never before.

## VI. CONCLUSION

We analyzed the big data trends, developments and the present most advanced data analysis techniques. We reviewed some of the most cited and technologically forward research papers to build our own analysis of the future trends associated with the machine learning and deep learning applications of big data. The future will be driven by the big data technologies. Since, the semiconductor technologies keep upgrading at regular intervals at fast pace, the speed of



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 3, March 2018

computer processors and the storage devices are sure to see technological revolutions. Moreover, since the inception of internet, this decade has seen phenomenal digital revolutions and trends, which is only the beginning. The big data and cloud computing are the permanent technologies of the future which will bloom at exponential rates. For their analysis, by the looks of current trends, Cloud Computing, Data Mining, Statistical Analysis, Intelligent Analysis, Machine Learning, Quantum Computing and Data Stream Processing Techniques will be advanced and many more technologies are sure to come in the way. This review of current and future trends of the big data technology marks the importance of two most important techniques- machine learning and deep learning because of their better compatibility with big data analysis. We conclude this paper with the hope of better future for us all.

## REFERENCES

- [1] Xiaolong Jin , Benjamin W. Wah a,b, Xueqi Cheng a, Yuanzhuo Wang, “ Significance and Challenges of Big Data Research”, Big Data Research Elsevier, Pages: 2-6, 2015.
- [2] [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data), Big data, 2018.
- [3] <https://economictimes.indiatimes.com/news/company/corporate-trends/mukesh-ambani-says-data-is-new-oil-for-fourth-industrial-revolution/articleshow/57173843.cms> , Mukesh Ambani says data is new oil for fourth industrial revolution, 2017.
- [4] Quoc V. Le, Jiquan Ngiam, Adam Coates, Abhik Lahiri, Bobby Prochnow, Andrew Y. Ng, On Optimization Methods for Deep Learning, 28<sup>th</sup> International Conference on Machine Learning, Bellevue, WA, USA, Pages: 1,7, 2011.
- [5] D. P. Acharjya, Kauser Ahmed P, A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, Pages: 511-515, 517, 2016 .
- [6] Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey, Business intelligence and analytics: from big data to big impact, MIS Quarterly Vol. 36 No. 4, Pages: 1165-1188, December 2012.
- [7] Elena Geanina Ularu, Florina Camelia Puican, Anca Apostu, Manole Velicanu, Perspectives on Big Data and Big Data Analytics, Database Systems Journal vol. III, no. 4, Pages: 3,4,8,12, 2012.
- [8] Sumithra. J, Mining Big Data: Current Status, and Forecast to the Future, International Journal of Computer Science and Information Technology Research Vol. 2, Issue 3, Pages: 39-43, July - September 2014.
- [9] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, Edin Muharemagic, Deep learning applications and challenges in big data analytics, Journal of Big Data, Pages: 1-19, 2015.
- [10] Furqan Alam, Rashid Mehmood, Iyad Katib, Nasser N. Albogami, Aiiad Albeshri, Data Fusion and IoT for Smart Ubiquitous Environments: A Survey, IEEE Access VOLUME 5, Pages: 9533- 9549, 2017.
- [11] Alexandra L'heureux, Katarina Grolinger, Hany F. Elyamany, Miriam A. M. Capretz, Machine Learning With Big Data: Challenges and Approaches, , IEEE Access VOLUME 5, Pages: 7776- 7794, 2017.
- [12] Chandrakar Mukesh Kumar and Mishra Anup “Review of Medical Image Analysis, Segmentation and Application using Deep Learning” Scopis SJR, JARDCS ,ISSN 1443-023X , 01-Special Issue, Pages: 549-553, 2018.