Visualization of COVID-19 Pandemic: An Analysis Through Machine Intelligent Technique Toward Big Data Paradigm

Manash Sarkar, Saptarshi Gupta, Bhavya Gaur, and Valentina E. Balas

Abstract The word pandemic is scary to whole world. It is a pestilence of an irresistible ailment that has spread over an enormous district, for example, different mainland’s or around the world, influencing a generous number of individuals. An across the board endemic illness with a steady number of contaminated individuals is definitely not a pandemic. Far reaching endemic maladies with a steady number of contaminated individuals. The people are infected from various part of the world. The number of confirmed cases of COVID-19 in increasing abruptly day by day. The world economy is hampered due to this pandemic. To improve the world economy, it is required to adapt an optimistic decision for during this pandemic as well as for post-pandemic. In this research paper, a real-time data set from World Health Organization (WHO) is collected and analyzed. The nature of the data set is as the concept of big data. A statistical analysis is performed on the data set and produce the results. The data for India is only considered for analysis. The number of death cases and number of COVID-19 confirmed cases in India during the period of 15 weeks from 28th February 2020 to 7th June 2020. Finally, the analysis reported that in India rate of death cases is less than rate of cure cases.

Keywords Pandemic · COVID-19 · Big data · Sampling distribution · Data visualization
1 Introduction

Massive quantity of data can be used for various profits in daily life, business, health care, banking, retail, etc. Proper analysis of big data can be generate revenue as well as human health can also be monitored and proper treatment can be provide to the patients. The patients data can be collected from hospital patient personal file, doctor prescription, Generic databases, wearable devices, IoT healthcare device database, pathology test report, medical imaging, etc. [1, 2]. If massive outbreak of any disease takes place in any county or a particular place that data can be monitored from the government official website. Proper analysis of those data can be led to find some prominent findings and decision. Big data can be utilized to anticipate if there is a pestilence epidemic, remedial diseases, understanding a patient’s medical record, keep away from preventable fatalities, etc. Furthermore, patient treatment will be simpler if his health information is collected and the recognition of any severe illness is done in the early hours [3, 4]. A pandemic is an epidemic that is spread various nations. Not all infectious disease terms are made equivalent; however, regularly they are erroneously utilized conversely. The qualification between the words “pandemic,” “epidemic,” and “endemic” is normally obscured, even by clinical specialists. This is on the grounds that the meaning of each term is liquid and changes as infections become pretty much pervasive after some time. While conversational utilization of these words probably would not require exact definitions, realizing the thing that matters is essential to assist you with bettering comprehend general wellbeing news and fitting general wellbeing reactions. The coronavirus pandemic is probably going to keep going up to two years and would not be controlled until around 66% of the total population is invulnerable, a gathering of specialists said in a report.

In India, the COVID-19 pandemic is a part of the overall pandemic of coronavirus infection 2019 (COVID-19) brought about by serious intense respiratory condition coronavirus 2 (SARS-CoV-2). The first instance of COVID-19 in quite a while, which started from China, was accounted for on 30 January 2020. Starting at 10 June 2020, the MoH and FW has affirmed an aggregate of 276,583 cases, 135,206 recuperations (counting 1 movement) and 7745 passing in the nation. India right now has the biggest number of affirmed cases in Asia, with the quantity of absolute affirmed cases breaking the 100,000 blemish on 19 May and 200,000 on 3 June. India’s case casualty rate is generally lower at 2.80%, against the worldwide 6.13%, starting at 3 June. Six urban communities represent around half of every single revealed case in the nation—Mumbai, Delhi, Ahmedabad, Chennai, Pune, and Kolkata. Starting at 24 May 2020, Lakshadweep is the main district which has not revealed a case.

The United Nations (UN) and the World Health Organization (WHO) have applauded India’s reaction to the pandemic as “far reaching and hearty,” naming the Lockdown limitations as forceful yet essential for containing the spread and building important medicinal services foundation. The current year’s World Health Statistics report clarifies that the worldwide endeavors in ongoing decades have been paying off. Taking a gander at the most forward-thinking information organization have on a portion of these imperative Sustainable Development Goals (SDG) pointers, it
uncovers wellbeing patterns across Member States, areas, and the whole world. The information shows that people are proceeding to gain enormously promising ground from numerous points of view—yet in addition that they need progress in different manners. Disparity endures, with certain districts despite everything falling behind. They should keep on cooperating to stay concentrated on our objectives. Absence of accessible information called information blind spots, and stay a dire test yet in addition an incredible chance known as rarity of information assortment. Since any place people can quantify, they can gain ground. In light of its capacity to spread from individuals who do not seem, by all accounts, to be sick, the infection might be more enthusiastically to control than flu, the reason for most pandemics in late history, as indicated by the report from the Center for Infectious Disease Research and Policy at the University of Minnesota. Individuals may really be at their generally irresistible before side effects show up, as per the report.

The future of the world of post-pandemic depends on the effect of the presents. To increase the dross domestic product (GDP) rate, an intelligent analysis is required. Due to the COVID-19, people are infected and some worst cases death also. Therefore, the data set which is handled by WHO is increasing abruptly. The nature of the data set is increasing volume, velocity, variety, value, and veracity which satisfied the concept of big data. In this research paper, the data from WHO is analyzed and prepared a statistical report on Indian situation during this pandemic. The data is considered from 28th February 2020 to 7th June 2020. Uniform distribution and sample distribution are applied to determine the degree of closeness for a given sample with COVID-19 confirm cases. The rate of spreading of COVID-19 in India with respect to rest of the world is also compared.

The rest of the paper is defined as follows: details of literature survey is shown in Sect. 2 as a related works followed by fundamental of big data in Sect. 3. Pandemic in perspectives of big data is described in Sect. 4. Mathematical analytics is explained in Sect. 5 followed by data preparation in Sect. 6. Results and simulations are presented in Sect. 7 and discussion in Sect. 8. Finally, conclusion and further scope of research are discussed in Sect. 9.

2 Related Works

As per the history, last few century world is suffered from various pandemic. In 1920, due to influenza pandemic huge number of deaths and immunoprotection cases occurred. In 1817 to 1824, first Asiatic cholera pandemic or Asiatic cholera spread south Asia and South-East Asia. The world is experiencing different new outburst in past years, e.g., Zika virus (Identified in 1947) [5], SARS-CoV—severe acute respiratory syndrome coronavirus (Identified in 2003) [6], Ebola virus (Discovered in 1976) [7], and now novel coronavirus disease 2019 [COVID-19]. The consequences of the 2014 Ebola upheaval in West Africa, general wellbeing experts around the world, were approached to plan for the following pestilence. The biggest flare-up of its sort, the Ebola infection caused 28,600 cases and 11,325 passing’s, as per the
US Habitats for Disease Control and Prevention [8, 9]. Serving not just as an exercise for the future, the flare-up additionally featured the significance of enormous information in these sorts of crisis circumstances. Chen et al. [10] depicted for worldwide and national scourges; large information has progressed to a spot where it can give crisis reaction groups ongoing devices and innovation that can possibly screen, contain, and even stop the spread of infection. Conventional wearable devices have various drawbacks, such as un-comfortable for long-term wearing, and insufficient accuracy, etc. Thus, health monitoring through traditional wearable devices is hard to be sustainable. In order to obtain and manage healthcare big data by sustainable health nursing, the system design “smart clothing,” enabling unobtrusive collection of various physiological indicators of human body. To offer persistent cleverness for smart clothing erection, mobile healthcare cloud stand is constructed by the usage of mobile internet, cloud computing, and big data analytics. This paper announces design facts, key tools, and applied implementation methods of smart dress system. Typical claims powered by smart clothing and big data clouds are presented, such as medical backup response, emotion care, disease diagnosis, and real-time tangible interaction. Wang et al. [11] in their research, they estimated the lung cancer using double dispensation system. The image dispensation system was familiarized into double for early prediction. The challenge in this progression was recognition of tiny nodes which comprehends early cancer detection. The unstipulated knobs in lungs can be spotted using ridge recognition algorithm. Kelvin et al. [12] discussed the smoking behavior of the user. The e-cigarette has a small electrical resistance coiled wire in 1.5 Ω which is connected to the positive and negative poles of the device. When the button of e-cigarette is pressed, the resistance coil can be connected with electrical supply under the immersion of some “E-liquid,” the coil heats up and transform the E liquid to vapor, which can be inhaled by the smokers. It monitors the smoking behavior of the user in order to prevent the patient from cancer.

3 Fundamental of Big Data

Big data is a domain in which it deals with the extraction of information from huge amount of data or large data set and analyze the large data components. Nowadays due to huge and large amount of data, the conventional data processing software is not very much suitable for the data analysis and information extractions.

As per Gartner [13] the big data can be clarified as follows:

Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.

Data complexity increases day by day and this is related to the combinations of five “V” (Characteristics of big data [14, 15]. They are volume, velocity, variety, value, and veracity.

- **Volume**: Huge amounts of data, the data sets size may be near about 1024^7 bytes.
- **Velocity**: High speed data from live or real-time source.
- **Variety**: Data come from dissimilar sources. Data may be in different formats, e.g., structured (database table), semi-structured (Extensible Markup Language—XML data), unstructured (text, video, image and many more) etc.
- **Value**: Information in itself is of no utilization or significance; however, it should be changed over into something important to separate Information.
- **Veracity**: Veracity essentially implies the level of dependability that the information brings to the table. Since a significant piece of the information is unstructured and superfluous, big data needs to locate a substitute method to sift them or to decipher them through as the information is urgent in business improvements.

The big data term is not only indicated the large amount of data also it indicates the data arrives in fast speed having complex format from variety of sources. Basic properties of big data are graphically represented in Fig. 1.

### 3.1 Types of Data

Web Data, Text Data, Weather Data, Class Room Data, Time and Location Data, Email Data, Social Network Data, Sensor Data, Data from different Industries and Organizations, etc. (Fig. 2).

### 3.2 Big Data Ecosystem Components

The excursion to big data is to comprehend the levels and layers of abstraction, and the parts around the equivalent. Some basic segments of big data investigative stacks and their combination with one another. The proviso here is that, in a large
portion of the cases, HDFS/Hadoop structures the center of the greater part of the big-data-driven applications, however that is not a generalized rule or guideline.

3.2.1 Technologies for Capturing, Storing, and Accessing Big Data

Big data’s processing should be possible in four layers as appeared in Fig. 3. The primary testing task is to gather colossal measure of data from various sources in various arrangement. As the data is unstructured, it gets hard for conventional database the executives framework to separate information out of it yet huge data take care of this issue since it might help in extricating the information from organized, semi-organized, and unstructured data.

Initial step is to gather the data that is assembled from various sources and afterward store this gathered data in some common spot. To give distributed file system (HDFS) for conveyed capacity and adaptation to non-critical failure, Apache Hadoop is normally being utilized nowadays. MapReduce is a programming model utilized in Hadoop for handling enormous measure of data rapidly. Machine learning (ML) algorithm can be applied for accomplishing fast examination on input data and make the data which can be utilized for delivering data in processing layer.
3.2.2 Analytical Techniques

Data analysis (DA) is the way toward analyzing data set (inside the type of text, sound and video), and reaching determinations about the data they contain, all the more usually through explicit frameworks, programming, and strategies. Data analytics advancements are utilized on a modern scale, across business ventures, as they empower associations to make determined, intelligent decision for business.

Big data analysis mixes conventional measurable data analysis approaches with computational ones. Statistical sampling from a universal set is perfect when the whole data set is accessible, and this condition is commonplace of conventional bunch preparing situations. Be that as it may, big data can move clump handling to real-time preparing because of the need to understand spilling information. With
gushing information, the data set aggregates after some time, and the information is time-requested. A large portion of the broadly utilized analyzing techniques are listed as follows:

- Statistical methods
- Machine learning
- Data mining
- Data warehouse
- Database querying.

Globally, enterprises are bridling the intensity of different various information investigation methods and utilizing it to reshape their business models. As innovation grows, new examination programming rises, and as the Internet of Things (IoT) develops, the measure of information increments. Enormous information has advanced as a result of our expanding development and association, and with it, new types of removing, or rather “mining,” data.

### 3.2.3 Visualization

Data visualization is the practical depiction of data. It incorporates conveying pictures that pass on associations among the addressed data to watchers of the photos. This correspondence is practiced utilizing an efficient mapping between reasonable engravings and data regards really taking shape of the portrayal. This mapping sets up how data regards will be addressed ostensibly, choosing how and how much a property of a reasonable engraving, for instance, size or concealing, will change to reflect changes in the estimation of a datum. To grant information indisputably and capably, data discernment uses genuine structures, plots, information delineations, and distinctive devices. Numerical data may be encoded using spots, lines, or bars, to apparently bestow a quantitative message. Convincing portrayal helps customers with separating and reason about data and evidence. It makes complex data progressively accessible, legitimate, and usable. After analysis of the data, graph needs to be plotted to see the results and data visualization is necessary. “R” language can be used for data visualization.

### 3.3 Applications of Big Data

In today’s world, the data is wealth. Based on the previous data analysis, the future plan would be accepted by any organizations as well as individuals also. Today, people are very much flexible to access their information through Internet. In the concept of Internet of Things (IoT)-based smart cities, the life is totally controlled by the internet. Due to precious time in human life and their precious data to keep in a secure place, people are going to store their data in various cloud platform. Therefore, the volume of the data and their velocity, variety are also changes very
rapidly. In every fields like social networking, health care system, online shopping, online traveling system, academic system, finance system, agriculture, concept of big data is explored. There are some application of big data [18] are discusses as follows:

- **Segmentation and prediction**: According to the customers behavior, segmentation can be done and different predictions can be formed for growth of the business.
- **Recommender systems and marketing**: Various recommendation schemes are already available in existing world but that may be placed in a proper place. Nowadays, online recommendation system become very famous, and by using this recommendation system, business industries can target the appropriate customers or audience that can lead to target marketing.
- **Sentiment analysis**: This is one of the new trend in data analysis. This kind of examination offer data on what the market is stating, thinking & feeling about an association. Mostly the sentiment of a customer or individuals is analyzed by the data taken from social media.
- **Operational analytics**: Without the human involvement, automatic decision can be taken in consumer service in different industries so that quick and fair decision can be made for all the consumer.
- **Prediction for optimal decision**: This is the one of most important domains where the application of big data is fully explored. Analysis of previous data set and trace out an optimal decision for the near future. Most of the organizations even also individuals are considered for this services. In the field of marketing, production, banking sector, health care, economic are the most useful domain for the application of big data.

In this paper, a data set of pandemic of current scenario is considered and analyzed to predict the infection of COVID-19 for the people of India. This is one kind of big data application. The data value of the data set is changed abruptly with variety of values.

### 4 Pandemic: Perspectives of Big Data

The episode of COVID-19 has seriously influenced all parts of human life, forcefully affected the worldwide economy, and has put an enormous strain on the worldwide wellbeing framework. In an offer to contain the pandemic, organizations and pioneers at all levels are utilizing huge information and investigation devices, among different arrangements, to bring down the effect of the infection. They are utilizing AI-enabled framework, analyzing, and visualization tools for big data to conceive the eventual fate of the pandemic, track the continuous spread of the infection, discover cures against COVID-19.

Various associations likewise have begun to utilize huge information examination advances to speed up sedate improvement against COVID-19 and better see
how the invulnerable framework vanquishes the infection. Pharmaceutical organizations GlaxoSmithKline (GSK) and Vir Biotechnology, for example, joint powers to progress coronavirus treatment improvement utilizing artificial intelligence.

Abruptly spreading of COVID-19’s and effect have caused individuals in the whole world to feel vulnerable, terrified, and baffled. Be that as it may, the once in a while news and disclosures about the fix and utilization of current innovation sponsored by data makes them to feel safe and expectation in the terrible “pandemic” time.

While expressions of social distancing, lockdowns, and straightening the bend are turning out to be basic aphorisms in a typical man’s life, big data is demonstrating clever in considering the measurements about populace developments across districts, check open consistence in following the lockdown and wellbeing conventions, size and recurrence of individuals with higher paces of temperature from the information by temperature scanners, all of which help in foreseeing how the curve with develop or decrease in the coming weeks.

In order to establish the decision on data-driven and convey powerful administrations, organizations, policymakers, government substances, and others over the globe are understanding the operational preferences of utilizing enormous information examination. Moreover, intensified speculation from government and intelligence (G&I) and areas of social health care to deal with the pandemic will drive the market in the post-pandemic world. The capacity of big data to gather, store volumes of information, speed to permit continuous information stream for quicker preparing, bolster an assortment of data form: unstructured or structured, numeric-emplomatic, from various sources and inconstancy in taking care of them according to request loads with higher veracity makes it an extraordinary favorable position when utilized alongside artificial intelligence, machine learning, IoT applications, etc. Methods of big data analytics are appropriate for tracing and managing the spread of COVID-19 around the world.

5 Mathematical Analytics

In this research, data analysis is done where the volume of the data is enormous and the pattern of the data is unstructured and there is a veracity of the information.

The data set of COVID-19 is considered as a test bed. The properties of the data set are nothing but big data. Various state of India has been affected by COVID-19, but number of infected persons are vary from state to state.

Assuming total population of India is considered as a formal with coordinate with system \{x_1, x_2, \ldots, x_n\} Let consider the density function \(\delta(x_1, x_2, \ldots, x_n)\) which is the limit for the ratio of the number of infected in different state. The population of a small region is \(\Delta v\) round the point \(\{x_1, x_2, \ldots, x_n\}\) to \(\Delta v\). The total population set will be equilibrium if the infected persons are distributed as closet as possible to equidistribution over the total population with kinematic restriction.

The sample infected of position \(\{x_1, x_2, \ldots, x_n\}\) has \(E(x_1, x_2, \ldots, x_m)\) the mean of infected per each sample is given constant as
\[ \int E S \, dv = d \]  \hspace{1cm} (1)

where \( dv \) indicates the unit volume.

At uniform distribution, determine \( \delta \). The degree of closeness for a given sample distribution to uniform distribution. As the pattern of data as a big data, high degree of closeness is required. The entropy of the sample is defined by Eq. (2).

\[ -\int S \log s \, dv \]  \hspace{1cm} (2)

The solution of Eq. (2) as maximizing is subjected to Eq. (1)

The optimum choice of samples \( S \) for which \( -\int S \log s \, dv \) will be maximize depends on \( \int E S \, dv = d \) and is defined as

\[ S = e^{\lambda E + \mu} = \alpha e^{\lambda E} \]  \hspace{1cm} (3)

where \( \alpha \) is selected such that \( \int S \, dv = 1 \)

As per the inequality

\[ \int S \log \frac{s}{q} \geq 0 \]  \hspace{1cm} (4)

For every application of any two alternative density is defined as

\[ -\int S \log s \, dv \leq -p \log q \, dv = -\int Z(\lambda E + \mu) \, dv = -(\lambda d + \mu) \]  \hspace{1cm} (5)

If Eq. (5) is applied for Eq. (1), then the fact will be integral of \( S \) is unity.

The data samples for infected persons are considered for different state of India.

The rate of infection for different state would be different. Sampling distribution is deployed to transform the rate of infection at various level. For linear transformation \( Y = BS, \forall S \rightarrow Y \).

If \( B \) is considered a non-singular matrix then

\[ \frac{DY}{DS} = |B| \text{ as a positive sign} \]  \hspace{1cm} (6)

The interrelated differential elements are defined by the relational equation in Eq. (7).

\[ dy_1 \, dy_2 \ldots dy_n = |B| \, ds_1 \, ds_2 \ldots ds_n \]  \hspace{1cm} (7)

From Eq. (6)

\[ dY = |B|dS \quad \forall B \quad \text{is an orthogonal matrix and } |B| = 1 \]  \hspace{1cm} (8)
\[ Y = BS \] which transforms into a quadratic form

\[
S' S \rightarrow Y' Y \\
(S - \mu)' (S - \mu) \rightarrow (Y - \eta)' (Y - \eta) \quad \text{where, } \eta = B\mu
\] (9)

After determine the dissimilarity of the infected and suspected patients, distance from threshold value is evaluated by applying partitioning matrix.

\[
B = \begin{pmatrix}
B_1 \\
\vdots \\
B_k
\end{pmatrix}
\text{where, } B_i = n_i \times n \quad \text{and} \quad \sum n_i = n
\] (10)

The partitioning matrix is partitioned into \( k \)th sub-matrices. All sub-matrices are orthogonal to each other, but not be orthogonal themselves.

6 Data Preparation

A real data set is collected from World Health Organization (WHO) [19]. Data is collected on 10 June 2020 from official website of WHO. Only nineteen individual data records are collected. Each record contains two attributes like number of confirmed cases and number of deaths for seven days intervals. Each record contains the same details of three different regions in the world; India, South-East Asia Region, and world. In this research, the data is considered from 28th February to 7th June 2020; total 15 weeks in intervals. In this data set, total number of confirmed cases from each states in India are not mentioned. As per the requirement of the proposed methodology, data set is prepared. The modified data set is shown in Table 1.

For further classification and analyzing, different derived data sets are generated. Table 2 is derived from Table 1 and it shows number of increases for confirm cases in India during every seven days intervals. Table 2 contains only the records of India for the same time intervals.

Table 2 is a derived from Table 1. The data set, collected from WHO, is considered the death cases only caused by COVID-19 (Table 3).

7 Results and Simulations

The results are evaluated and simulated with MATLAB 2018a and R statistic tool. Values from the tables are used to evaluate the results. In this research, main focus is the record of COVID 19 confirmed for India. The data is taken only for 15 weeks
| S. No. | Date       | No. of State in India | Confirmed (India) | Death (India) | Confirmed (South-East Asia Region) | Death (South-East Asia Region) | Confirmed (World) | Death (World) |
|--------|------------|-----------------------|-------------------|--------------|-----------------------------------|-------------------------------|--------------------|--------------|
| 1      | 28-Feb-2020| 1 (Kerala)            | 3                 | 0            | –                                 | –                             | –                  | –            |
| 2      | 09-Mar-2020| 11                    | 44                | 0            | –                                 | –                             | –                  | –            |
| 3      | 14-Mar-2020| 13                    | 84                | 2            | –                                 | –                             | 132,758            | 4955         |
| 4      | 22-Mar-2020| 23                    | 360               | 7            | 979                               | 38                            | 266,073            | 11,184       |
| 5      | 28-Mar-2020| 27                    | 909               | 19           | 2536                              | 79                            | 462,684            | 20,834       |
| 6      | 05-Apr-2020| 29                    | 3577              | 83           | 6528                              | 267                           | 1,051,635          | 56,985       |
| 7      | 12-Apr-2020| 31                    | 8447              | 273          | 14,161                            | 617                           | 1,610,909          | 99,690       |
| 8      | 19-Apr-2020| 32                    | 16,116            | 519          | 27,319                            | 1185                          | 2,241,359          | 152,551      |
| 9      | 26-Apr-2020| 32                    | 26,917            | 826          | 41,073                            | 1658                          | 2,719,897          | 187,705      |
| 10     | 03-May-2020| 32                    | 40,263            | 1306         | 60,490                            | 2256                          | 3,267,184          | 229,971      |
| 11     | 10-May-2020| 32                    | 62,939            | 2109         | 90,808                            | 3204                          | 3,855,788          | 265,862      |
| 12     | 17-May-2020| 32                    | 90,927            | 2872         | 127,995                           | 4201                          | 4,425,485          | 302,059      |
| 13     | 24-May-2020| 32                    | 131,868           | 3867         | 182,278                           | 5556                          | 5,103,006          | 333,401      |
| 14     | 31-May-2020| 32                    | 182,143           | 5164         | 260,579                           | 7431                          | 5,934,936          | 367,166      |
| 15     | 07-Jun-2020| 32                    | 235,657           | 6642         | 336,577                           | 9316                          | 6,663,304          | 392,802      |
Table 2  Number of increases of confirmed cases

| Days | Number of increases in India |
|------|-----------------------------|
| 0    | 0                           |
| 7    | 41                          |
| 14   | 40                          |
| 21   | 276                         |
| 28   | 549                         |
| 35   | 2668                        |
| 42   | 4870                        |
| 49   | 7669                        |
| 56   | 10,801                      |
| 63   | 13,346                      |
| 70   | 22,676                      |
| 77   | 27,988                      |
| 84   | 40,941                      |
| 91   | 50,275                      |
| 98   | 53,514                      |

Table 3 Increases of death with respect to time

| Days | Death cases in India |
|------|----------------------|
| 0    | 0                    |
| 7    | 0                    |
| 14   | 2                    |
| 21   | 7                    |
| 28   | 19                   |
| 35   | 83                   |
| 42   | 273                  |
| 49   | 519                  |
| 56   | 826                  |
| 63   | 1306                 |
| 70   | 2109                 |
| 77   | 2872                 |
| 84   | 3867                 |
| 91   | 5164                 |
| 98   | 6642                 |

started from 28th February 2020 to 7th June 2020. During this time period, the number of COVID 19 confirmed cases is gradually increased up to sixth week then suddenly it is changed very abruptly.

In Fig. 4, the nature of increasing through 15 weeks is plotted. At seventh week,
the graph rapidly increases. After seventh weeks, how the confirmed cases increase and corresponding their rate of increasing also changes based on the number of confirmed cases. In Fig. 5, the rate of increasing is shown with time domain.

After analyzing Fig. 4, it is clear that the rate of increasing of confirmed COVID-19 is abruptly changed at 42 days. As the rate of confirmed cases is increased, contemporary number of deaths due to COVID-19 are also increased. The number of death cases due to COVID-19 is directly depended on the rate of confirm infected cases. Therefore, the number of death cases is also increased as the number of confirmed cases is increased.

Figure 6 represents the visualization of the analytical report of death cases due to COVID-19. The death rate is also abruptly increased from seventh week onwards. How number of death cases and number of confirmed cases are varied with each
other’s in a particular time domain, is shown in Fig. 7. In India, after 15 weeks, the total number of death cases are 6642, whereas the number of confirmed cases are 235,657.

After 15th weeks, the rate of death in respect to total confirmed cases is very less, it is 2.81%. In Fig. 7, it is shown that the graph of death cases is very closest and parallel to Y-axis whereas, the graph of confirmed cases is increased extremely. From this research analysis, it is clear that the rate of death is gradually decreases in respect to number of confirmed cases.

In support of this statement, another death report of other organization is shown Fig. 8 which convergences to this proposed research analysis. Report at Fig. 8 is
Fig. 8 Fifteen days records in India [20]

taken from 27th May 2020 to 10th June 2020. The report shows that the percentage of death on 27th May 2020 is 4.1% while it is 3.7% on 10th June 2020.

8 Discussion

As per the report of WHO, the data is collected and analyzed through equidistribution over the total population. The main focus of this paper is the analysis for spreading of COVID-19 in India and also compares the rate of increasing with rest of the world. If a clear look into the data of India, then it is found that on 28th February 2020, India has only three confirmed cases of COVID-19, whereas, on 7th June 2020, it has been increased on 235,657. The report of spreading coronavirus in India is very shocking. But if think deeply, then it is found that the number of deaths during this 15 weeks is changed from 0 to 6642. The number 6642 is really mean as a death cases. But if a statistical analysis is explored, then find the mean ($\mu$) value and standard deviation ($\sigma$) in both cases.

$$\mu_{\text{Confirmed}} = \frac{1}{n} \sum_{i=1}^{n} \text{Confirm}_i \quad \text{and} \quad \sigma_{\text{Confirmed}} = \sqrt{\frac{\sum (p_i - \mu_{\text{Confirmed}})^2}{n}}$$
where $i = 1, 2, \ldots, n$.

$$
\mu_{\text{Death}} = \frac{1}{n} \sum_{i=1}^{n} \text{Death}_i \quad \text{and} \quad \sigma_{\text{Death}} = \sqrt{\frac{\sum (p_i - \mu_{\text{Death}})^2}{n}}
$$

where $i = 1, 2, \ldots, n$.

where

$n \rightarrow$ Size of the population

$p_i \rightarrow$ Each value from the population.

In this paper, the value of $n = 15$

$$
\mu_{\text{Confirmed}} = 15,710.27 \quad \text{and} \quad \sigma_{\text{Confirmed}} = 18,988.04 \\
\mu_{\text{Death}} = 1579.26 \quad \text{and} \quad \sigma_{\text{Death}} = 2134.092
$$

Now, look into the death rate during this considerable 15 weeks. From Table 1, it is found that on 28th February 2020, in , death cases are zero whereas, on 14th 2020, March, death cases are 2. On 7th June, death cases in India are 6642. Now, evaluate the percentage of death cases in respect to the number of confirmed cases.

In India, the rate of death cases based on the number of confirmed cases during different time period are shown in Table 4. The death rate of India is compared with the death rate of world.

Figure 9 shows a Pie chart to describe the death rate of India with death rate of world due to coronavirus COVID-19 on 7th June 2020. From Table 4, it is cleared that in India, the death cases are not increased so abruptly as the number of confirmed cases. It gives some hopes of better tomorrow.

| Table 4 | Rate of death cases at various time period |
|---------|------------------------------------------|
| Time duration | Percentage of death cases (%) |
| On 14th March 2020, percentage of death | 1.17 |
| On 7th June 2020, percentage of death | 2.81 |
| 31st May to 7th June, death rate increases | 2.76 |
| Average death rate (14th March to 7th June) | 1.64 |
9 Conclusion

The COVID-19 pandemic, otherwise called the coronavirus pandemic, is a continuous pandemic of coronavirus sickness 2019 (COVID-19), brought about by serious intense respiratory disorder coronavirus 2. The infection is principally spread between individuals during close contact, regularly by means of little beads delivered by hacking, wheezing, and talking. The beads generally tumble to the ground or onto surfaces as opposed to going through air over significant distances. The whole world is suffering from COVID-19 corona virus. Number of people is infected abruptly and contemporary death rate is also increased. In this paper, a statistical analysis is performed over the data set collected from WHO. A sample distribution is performed on the data of India from 28th April 2020 to 7th June 2020. During this 15 weeks how the rate of change of COVID-19 confirmed cases and also death cases in India are determined and visualized by simulations outputs. On 14th march 2020, the death rate of India is 1.17% in respect to whole infected people, whereas it is increased up to 2.81% on 7th June 2020. From the analysis report, it is cleared that the rate of increasing death is very less than rate of increasing infected cases. Thus, it implies that the rate of cure is greater than rate of death. The analysis of the pandemic data set as a big data is useful to make an optimistic decision for post-pandemic scenario. Based on this analysis, the economic condition, GDP rate, market analysis will be considered as future research directions.

References

1. Mathawan, R. (2020). Big Data in Healthcare, March 25, 2020. https://techstory.in/big-data-in-healthcare/. Accessed on 12/06/2020.
2. Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. Journal of Big Data, 6(54), 1–25.
3. Saranya, P., & Satheeskumar, B. (2016). A survey on feature selection of cancer disease using data mining techniques. International Journal of Computer Science and Mobile Computing.
5(5), 713–719.
4. Hamilton, B. (2012). Big Data is the future of healthcare, cognizant white paper, 2012. https://www.cognizant.com/industries-resources/healthcare/Big-Data-is-the-Future-of-Healthcare.pdf. Accessed on 10/06/2020.
5. https://www.who.int/news-room/fact-sheets/detail/zika-virus#:~:text=Zika%20virus%20is,a%20mosquito,Americas%20and%20Asia%20and%20the%20Pacific. Accessed on 10/06/2020.
6. https://www.who.int/ith/diseases/sars/en/. Accessed on 10/06/2020.
7. https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease. Accessed on 10/06/2020.
8. https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html. Accessed on 10/06/2020.
9. https://onlinedegrees.unr.edu/blog/the-role-of-big-data-in-global-epidemics/. Accessed on 10/06/2020.
10. Chen, M., Ma, Y., Song, J., Lai, C.-F., & Bin, Hu. (2016). Smart clothing: Connecting human with clouds and Big Data for sustainable health monitoring. Mobile Networks and Applications, 21, 825–845.
11. Wang, W., & Wu, S. (2006, June 25–28). A study on lung cancer detection by image processing. International Conference on Communications, Circuits and Systems.
12. Tsoi, K. K. F., Kuo, Y.-H., & Meng, H. M. (2015). A data capturing platform in the cloud for behavioral analysis among smokers an application platform for public health research. In 2015 IEEE International Congress on Big Data, New York, NY, USA, 27 June–2 July 2015.
13. Beyer, M. A., & Laney, D. (2012). The importance of ‘Big Data’: A definition. Gartner.
14. O’Leary, D. E. (2013). Artificial intelligence and big data. IEEE Intelligent Systems, 28, 96–99.
15. Berman, J. J. (2013). Introduction. In Principles of Big Data (pp. xix–xxvi). Morgan Kaufmann, Boston.
16. https://www.ubuntupit.com/best-big-data-applications-in-todays-world/. Accessed on: 23/05/2020.
17. Zikopoulos, P., et al. (2011). Understanding Big Data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media.
18. Su, X., Introduction to Big Data, NTNU Learning material is developed for course IINI3012 Big Data.
19. World Health Organization (WHO). https://www.who.int/india/emergencies/coronavirus-disease-(covid-19)/india-situation-report. Accessed on 10/06/2020.
20. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India. Accessed on 10/06/2020.