# An Innovative Pandemic Knowledgebase Using Machine Learning


**Ajanta Das, Sameya Ashraf, and Shubam Prasad**

## 1 Introduction

In the current situation, the pandemic COVID-19 is the worst thing that can happen to humans. The long-lasting corona virus and its exponential spreading nature have brought more devastation worldwide. The long-lasting effects may vary from one country to another. However, the affected cases are increasing day by day with a series of deaths that have been observed for every country, and the pattern is the same everywhere. Obviously, people are also recovering and maybe the recovered numbers are higher than death counts, but it is very hard to predict when the spike of death and affected cases will decline sharply. Lots of precautions and prevention activities such as washing hands with soap for 20s rigorously, sanitizing all the surroundings frequently, wearing face mask, and keeping social distancing are notable. In the twenty-first century, living in a nuclear family is increasing gradually. Due to the very small size (15–100 nm) of virus, it is doubtful whether the face mask can prevent corona virus inhaling. So, people can only protect themselves by restricting their outside activities, and confine and seal themselves at home safely. Once again, it is not possible for human beings to be safe at home day after day and month after month. As a result, technology-driven and social media-based living is prioritized. Meeting people, friends, and near and dear ones on online platform gives some kind of relief from living alone or living with own family members for long within a residential flat. With the restriction on going out from home, helping hands for household jobs has also stopped as a prevention measure. Simultaneously, with equal opportunity of working and more responsibilities at home, working women are under much pressure, including those who are working for home and those working from home digitally.

A. Das (✉) · S. Ashraf · S. Prasad
Amity Institute of Information Technology, Amity University, West Bengal, Kolkata, India



S. Bhatia et al. (eds.), *Intelligent Healthcare*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-030-67051-1_7

In order to sustain the job, leading life in connectivity with various electronic devices is mandatory. As a result, exponential growth of data has arisen from various smart devices. Hence, data analysis and representation of data are necessary to prevent the spread, and are essential to get back to normal life scenarios. Data science and machine learning techniques and algorithms may need to be used to represent or interpret the data properly for further protection and awareness of general people [1]. Discovery of knowledge and decision-making information extracted from social media play a vital role. Analysis of previous epidemic and pandemic cases is presented in Sect. 2. Section 3 presents a brief study about the present pandemic situation and sustainability issues for existing healthcare policies. This chapter presents the importance of data analytics and knowledgebase in Sect. 4. Furthermore, this chapter presents two separate case studies for international and national scenarios, both with respect to the pattern of spreading of the corona virus in Sect. 5. Next, Sect. 6 proposes a knowledgebase based on the available data and then presents a rule-based learning in healthcare to handle corona virus categorically in the current pandemic situation. Finally, the chapter concludes with Sect. 7.

## 2 Historical Analysis of Epidemic and Pandemic

This section presents the analysis of previous epidemic or pandemic cases preceding COVID-19 cases since 1918. This historical analysis is presented based on the knowledge gathered from previous cases, such as, when it occurs, how it spreads, what is the rate of spreading, whether any symptoms are present or not, what is the lifetime of the virus, and how it is controlled or how people have protected themselves from the infection of the virus. With reference to the responses to all these abovementioned questions, analyses are presented for five virus types, namely Spanish Flu, Asian Flu, Swine Flu, West African Ebola, and Zika virus, which are as follows:

(a) S*panish Flu*: This is the oldest devastating pandemic, found from the archives or history. It is found from historical data that 500 million people were infected worldwide during 1918–1920 [2]. The severity of pandemic was highest globally, but the recovery time was 1–2 weeks. Infections spread with influenza symptoms, while very less cases were asymptotic. It was airborne and the virus life time was 2–4 weeks. People used to wear masks and sheltered themselves in their homes as a protective measure. As a consequence, offices, schools, and businesses were closed for a long time to fight the pandemic, as the fatality rate was 3%.
(b) *Asian Flu*: From east Asia, a new virus was first reported in Singapore in 1957. The pandemic was named Asian flu, and it had a high death rate. From the record, around 1.1 million deaths occurred worldwide. The virus spread both ways with and without symptoms. Compared to Spanish flu, the fatality rate

was very less, <0.2%. The lifetime of the virus depended on the severity of the infection, on average, 14–21 days and the recovery time was normally 2–4 days.

(c) *Swine Flu*: After more than 50 years, in 2009, pandemic by influenza A (H1N1) virus was reported. It started in the United States and quickly spread across the world. The virus spread asymptotically 90% with very less symptoms. However, the fatality rate was very less, <0.03%. The lifetime of the virus was a maximum of 10 days and the recovery time was normally 3–7 days.

(d) *Ebola Epidemic*: The epidemic spread during 2013–2016 in western Africa with the outbreak of Ebola virus disease (EVD). The most affected and disrupted places were Guinea, Liberia, and Sierra Leone. It was neither seasonal nor airborne. It spread through blood or body fluid of the infected person. Hence, it did not spread widely, but the fatality rate was very high, 50%. According to the report, 28,616 people were infected, while the number of deaths was 11,310 [3].

(e) *Zika Virus Epidemic*: Simultaneously with Ebola epidemic in Brazil, during 2015–2016, Zika fever spread in the United States [4]. A few cases of the same virus spread were reported in Asian and Pacific island regions. After a year of spreading cases and based on analysis, it was named Zika epidemic in November 2016 by the World Health Organization. Zika epidemic spread mostly with symptoms like normal influenza, joint pain, and conjunctivitis [5]. Although it had not taken the shape of pandemic, the rate of fatality was maximum compared to other cases, 8.3%, due to the virus life span of 2–42 days.

## 3 COVID – 19: Pandemic

The virus is invisible in nature but can be seen through microscope. The virus looks like a "crown" and hence the name corona, as the meaning of corona is crown. The crown is surrounded by protein which causes the speedy transmission into human beings. Till date, it is suspected that the origin of this outbreak of coronavirus is Wuhan, China, in December 2019. The infection of the virus initiated with common cold, leading to severe respiratory syndromes and mortality, and spreading throughout the globe. This most devastating disease is famous with the name COVID-19.

Often, people get confused between outbreak, epidemic, and pandemic. In case of outbreak, the spreading of any infectious disease is within a community. While this outbreak lasts for a few weeks and goes beyond control within that community or location, it is well known as epidemic. Lastly, when epidemic travels throughout the world, it is identified as pandemic. Revisiting the historical data, an example of epidemic is *Ebola Virus,* while *Spanish flu* is identified as pandemic. Since none of us are relieved from corona virus globally, COVID-19 is a pandemic affecting all countries.

The virus is usually active from 2 to 14 days. However, the removal of effect from the virus varies between 30 and 49 days, depending on the severity of the cases. These symptoms are usually mild and increase gradually. Lack of attention to these symbols may lead to severe complications followed by fatality. The virus is spreading asymptomatically also. At the outset, people need to remember that "prevention is always better than cure" and hence most effective protection and prevention ways for COVID-19 are as follows:

(a) Clean the hands frequently and thoroughly.
(b) Wear face mask whenever necessary, if possible, always.
(c) Avoid touching eyes, mouth, and nose.
(d) Cover while coughing with the bend of elbow or with a tissue.
(e) Maintain a distance of at least 1 m from others.

The next section presents the machine learning techniques for preparing the knowledgebase.

## 4 Data Analytics and Knowledgebase

Data analytics reviews raw data and reexamines the datasets to present the pattern of the raw dataset. Data analysis is the basis of data analytics. Analysis of the same raw datasets needs to be reiterated before presenting analytics. Now machine learning is a method of data analysis which identifies data pattern and performs decision-making process without human intervention. Well-known machine learning algorithms include supervised and unsupervised clustering or regression to automate data analytics through mechanical processes. The main objective of data analytics is to optimize the performance; more specifically, optimization of business performance is done based on data analytics.

Knowledge discovery plays an important role in extracting knowledge from fuzzy datasets and presenting the meaningful pattern from all these fuzzy datasets [1]. Various techniques like *classification*, *clustering*, *dimensionality reduction*, and *collaborative filtering* for processing these datasets are used to perform predictive analytics. Classification technique or algorithm works for well-known groups. In case of supervised technique, learning is mandatory based on training data and then predicting the model for categorization of new data. Unlike classification, clustering techniques are applicable for unknown datasets and the grouping or categorization of data is prepared based on similarity index or degree of similarity. So clustering technique is iterative. Next, dimensionality reduction is applicable for reducing various dimensions or eliminating non-desired dimensions through feature selection and feature extraction, as each feature is not important for decision-making or prediction. Last, collaborative filtering gathers preferences from users' similar interests and then generates personalized recommendations.

Hence, machine learning explores useful knowledge by applying all these above-mentioned techniques, then matches the extracted information or knowledge with

historical information or existing patterns in the datasets. Usually, the existing patterns help identify the new patterns and new patterns need to be validated before integrating with existing knowledge database. Machine learning technique, specifically decision tree, can be easily applied to COVID-19 dataset for the preparation of knowledgebase and historical information. The datasets for asymptomatic or based on some symptoms are not binary, rather fuzzy. As we know, all influenza types with cough and sneezing patients are not identified as COVID-19 patients, so validation of this data is very important and specific knowledgebase can be ready for better awareness. Hence, extraction of the relevant information can be integrated with existing pandemic knowledgebase for future usage.

## 5 Novel Case Studies

In the twenty-first century, people are sealed inside their residential premises only as preventive strategy against COVID-19. Most of the countries around the globe somehow have been infected by this virus. We are trying to keep a record or analysis of the spread of COVID-19 around the world. Because of the ongoing pattern and multiple characteristics of the corona virus, the dataset is changing every day. The data for analysis are collected from various media sources, such as daily newspapers, bulletins, or press conferences on television, and mostly from social media [6, 7].

This section presents two different case studies for international and national scenarios to present the ratio between number of deaths and number of active cases between number of deaths and number recovered, wherever possible.

### *International*

The virus spreading started with China, Spain, France, and Italy gradually. As the virus spread in India since late March 2020, to prepare the knowledgebase for COVID-19, it is important to study the data pattern for international countries first. This section collects data for 10 different countries like Australia, Canada, China, Italy, France, Germany, Spain, the United Kingdom (U.K.), the United States, and India for the last 6 months from February 2020 to July 2020. Table 1 presents the data for the last 6 months of the total infected cases of the above-mentioned countries.

Next Table 2 presents active cases and deaths with respect to the total infected cases during February 2020–July 2020.

Figure 1 presents the death percentages with respect to the total cases from February 2020 to July 2020. The chart also presents comparison between all the above-mentioned countries. Regarding death cases, France, Italy, and United Kingdom are showing devastation. Compared to that, results for Germany and the United States show little control over death cases. Although in case of India and

**Table 1** Infected total cases during February 2020–July 2020

| Country | February | March | April | May | June | July |
|---|---|---|---|---|---|---|
| Australia | 25 | 4738 | 1991 | 441 | 641 | 9069 |
| Canada | 20 | 8592 | 44624 | 37711 | 13257 | 12108 |
| China | 68033 | 1730 | 1308 | 139 | 530 | 761 |
| Italy | 1128 | 104664 | 97799 | 29388 | 7599 | 6959 |
| France | 100 | 52028 | 77453 | 22172 | 13048 | 23118 |
| Germany | 79 | 71729 | 91201 | 20485 | 12338 | 14833 |
| Spain | 58 | 95865 | 143417 | 47169 | 9842 | 39251 |
| U.K. | 23 | 22769 | 132359 | 93774 | 34328 | 19928 |
| U.S.A. | 63 | 194051 | 905991 | 754843 | 873908 | 1977033 |
| India | 3 | 1394 | 33466 | 155746 | 395183 | 1111262 |

**Table 2** Infected active cases and deaths during February 2020–July 2020

| Country | Total cases | Active cases | Deaths |
|---|---|---|---|
| Australia | 16905 | 6726 | 197 |
| Canada | 116312 | 6510 | 8935 |
| China | 72501 | 684 | 4634 |
| Italy | 247537 | 12422 | 35141 |
| France | 187919 | 76154 | 30265 |
| Germany | 210665 | 9141 | 9224 |
| Spain | 335602 | Not available | 28445 |
| U.K. | 303181 | Not available | 46119 |
| U.S.A. | 4705889 | 2222756 | 156747 |
| India | 1697054 | 564856 | 36551 |

Australia death cases are very less, the corona virus started spreading late compared to other countries.

It reflects from both Tables 1 and 2 that the spreading of the virus will decline only after reaching its peak and it is very hard to predict when the peak time will arrive.

Next, the chapter presents the mobility view of Corona virus throughout the world. The daily or monthly cases for various country reflect different patterns. So, it is very difficult to arrive at a conclusion as to how and when the virus will spread around the countries or stay within the same country. A study or analysis is carried out for six different countries like China, Italy, France, Germany, Australia, and United States of America. The graph represents the number of total COVID cases for the month, specifically for the week; then out of these total cases, number of deaths or fatality, number recovered, and finally, the active cases are presented for all these above-mentioned six countries in Figs. 2, 3, 4, 5, 6, and 7 respectively. The pattern reflects that the virus started spreading in China first and its peak season occurred in February; then it spread mostly to Italy, France, Germany, and in Australia during the first week of March, it started spreading, but within a span of 2–3 weeks, it rose to a peak and within one month, the fatality rate reached the highest. At the outset, in the United States, the virus started spreading during

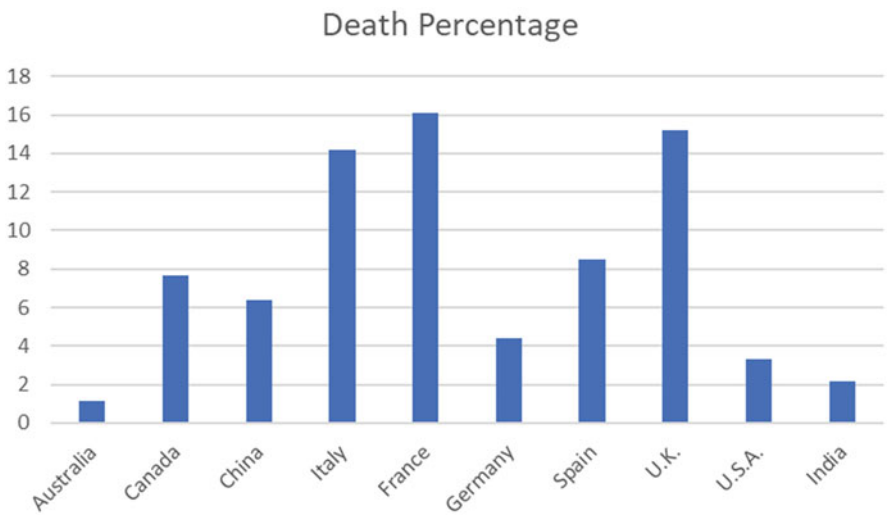


**Fig. 1** International death percentage during February 2020–July 2020

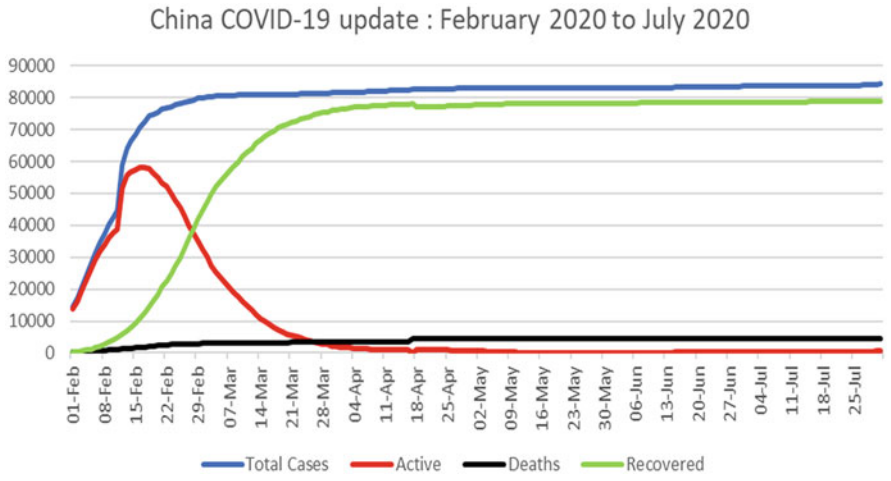


**Fig. 2** COVID data analysis in China

March but it showed sharp decline during July. However, fatality was controlled. The fatality rate is highest in Italy and France among all countries. The analysis presents that while the graph for active numbers declines, the number of total cases is highest but recovered numbers are also increasing steadily and deaths are steadily declining. It provides the knowledgebase that may be almost everyone or more than 30% of population will be infected by corona virus and people will recover soon by following guidelines properly.

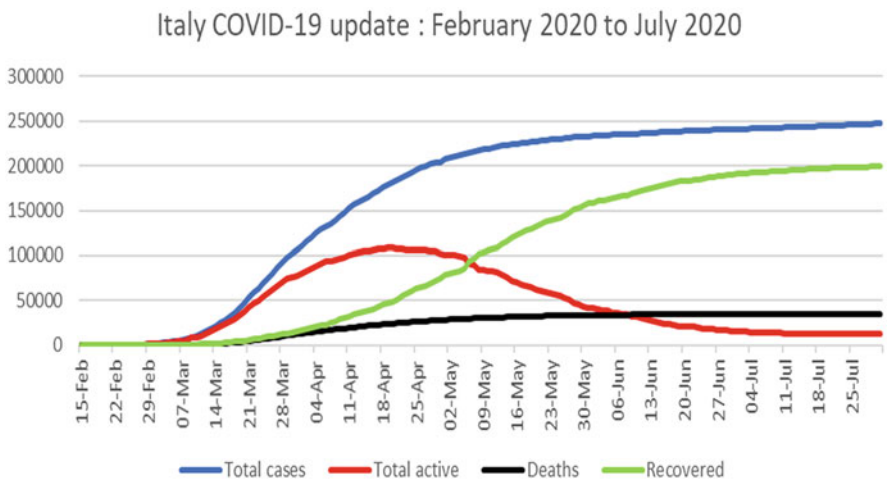


**Fig. 3** COVID data analysis in Italy

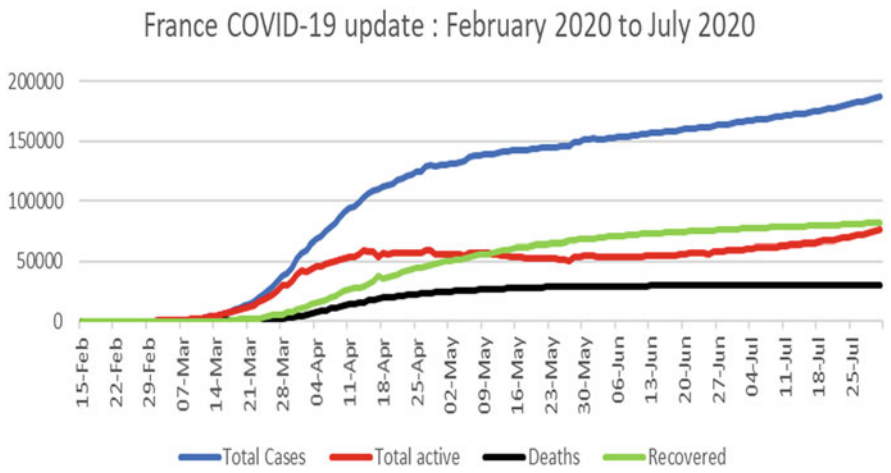


**Fig. 4** COVID data analysis in France

## *National*

The case studies for 10 different states, Bihar, Delhi, Haryana, Kerala, Maharashtra, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal, in India are presented based on observation and data collected from various social media [8, 9]. It presents the spreading pattern in various towns and villages for six consecutive months. Further, the classification of these dataset is categorized based on "A" as Active Cases, "R" as Recovered, and "D" as Deaths. It is observed that no case is reported in February 2020. So, Table 3 reflects datasets during March 2020 to July

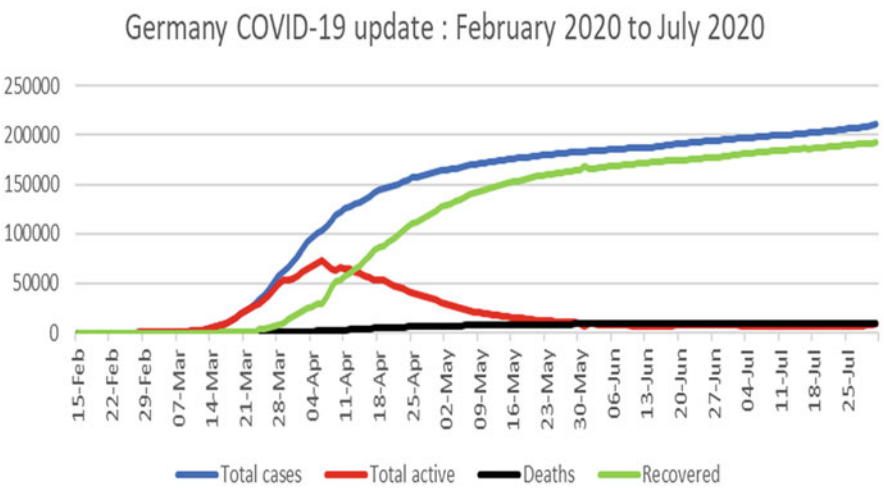


**Fig. 5** COVID data analysis in Germany

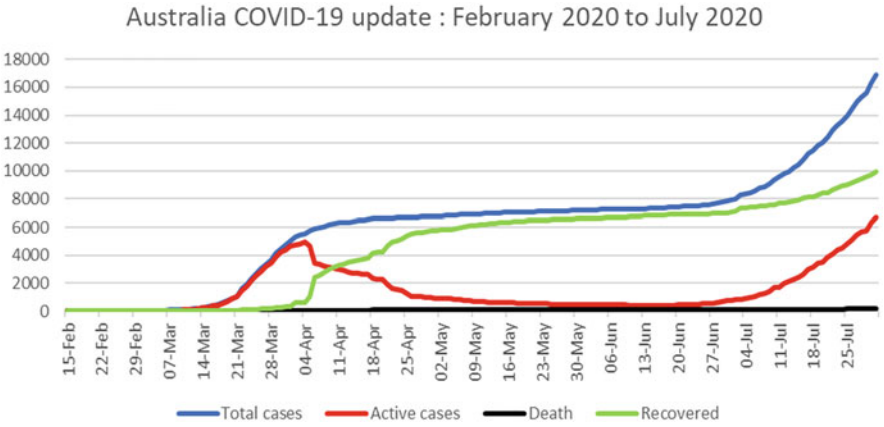


**Fig. 6** COVID data analysis in Australia

2020 and Table 4 presents total active cases, recovered, and deaths for the above-mentioned states within the same timeline.

The case study presents that infected cases are increasing because the disease spreads primarily from one person to another person through the small droplets from the nose or mouth of those who have been infected with COVID-19 virus [10]. These droplets can land on objects and surfaces around the person such as tables, door-knobs, and handrails, etc., and people can become infected by touching these objects or surfaces, and then when they touch their faces.

Figure 8 presents the death percentages with respect to the total cases for all the above-mentioned states of India. The observation shows that the most infected state

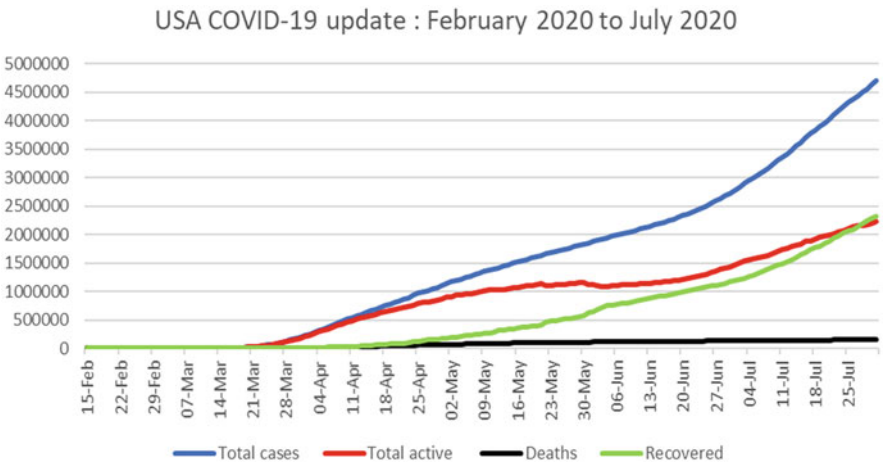


**Fig. 7** COVID data analysis in United States of America

**Table 3** COVID-19 datasets of various states during March 2020–July 2020

| | March | | | April | | | May | | | June | | | July | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| State | A | R | D | A | R | D | A | R | D | A | R | D | A | R | D |
| **Bihar** | 20 | 0 | 1 | 339 | 84 | 1 | 2264 | 1436 | 21 | 2376 | 6024 | 45 | 17039 | 26106 | 230 |
| **Delhi** | 112 | 6 | 2 | 2362 | 1088 | 57 | 10893 | 7384 | 414 | 26270 | 49870 | 2269 | 10705 | 62582 | 1221 |
| **Haryana** | 19 | 24 | 0 | 100 | 211 | 4 | 1023 | 813 | 16 | 4340 | 8924 | 216 | 6317 | 18255 | 185 |
| **Kerala** | 215 | 24 | 2 | 111 | 359 | 2 | 670 | 207 | 6 | 2114 | 1714 | 15 | 10517 | 10719 | 49 |
| **Maharashtra** | 252 | 39 | 11 | 8266 | 1734 | 448 | 36040 | 27556 | 1827 | 75995 | 61582 | 5569 | 74683 | 150966 | 7139 |
| **Punjab** | 37 | 1 | 4 | 356 | 103 | 16 | 231 | 1883 | 25 | 1557 | 1880 | 99 | 4999 | 6867 | 242 |
| **Rajasthan** | 90 | 3 | 0 | 1633 | 890 | 58 | 2605 | 5139 | 136 | 3375 | 8188 | 219 | 11558 | 15625 | 267 |
| **Tamil Nadu** | 117 | 6 | 1 | 1038 | 1252 | 26 | 9400 | 11499 | 149 | 38892 | 37317 | 1025 | 57968 | 133882 | 2734 |
| **Uttar Pradesh** | 87 | 17 | 0 | 1620 | 534 | 40 | 3015 | 4292 | 177 | 6711 | 11241 | 480 | 34968 | 32778 | 933 |
| **West Bengal** | 31 | 3 | 3 | 601 | 121 | 30 | 3027 | 2033 | 284 | 5761 | 9973 | 351 | 20233 | 36244 | 913 |

**Table 4** State-wise total active cases, recovered, and deaths

| | Total | | |
|---|---|---|---|
| State | A | R | D |
| **Bihar** | 17039 | 33650 | 298 |
| **Delhi** | 10705 | 120930 | 3963 |
| **Haryana** | 6317 | 28227 | 421 |
| **Kerala** | 10517 | 13023 | 74 |
| **Maharashtra** | 150966 | 256158 | 14994 |
| **Punjab** | 4999 | 10734 | 386 |
| **Rajasthan** | 11558 | 29845 | 680 |
| **Tamil Nadu** | 57968 | 183956 | 3935 |
| **Uttar Pradesh** | 34968 | 48863 | 1630 |
| **West Bengal** | 20233 | 48374 | 1581 |

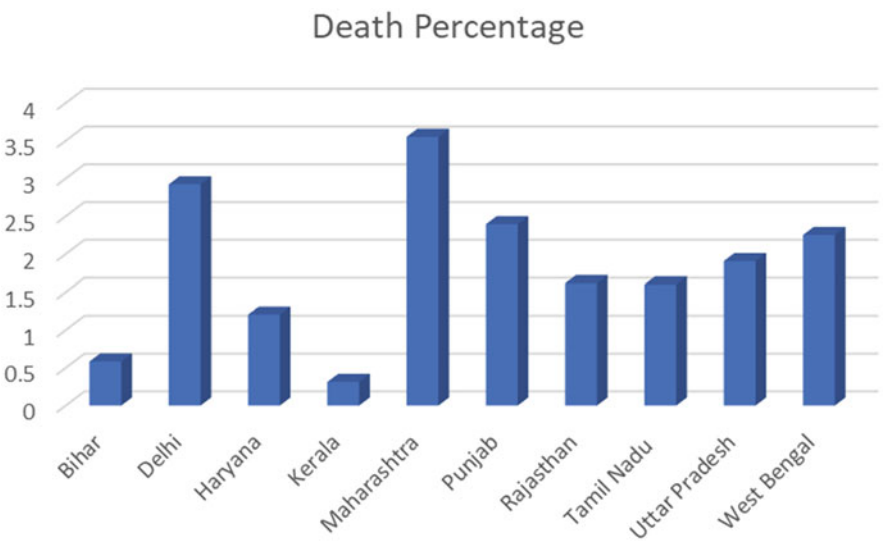


**Fig. 8** National death percentage during February 2020–July 2020

is Maharashtra and the least infected state is Punjab. Similarly, in recovered cases, Punjab is also the lowest one in recovered cases among all selected 10 states and Maharashtra is the highest one in recovered cases also. And in death cases, Kerala is the lowest one among all selected 10 states and Maharashtra is the highest which proves that more devastation has occurred in all three cases.

The chapter also presents detailed analysis for six different states of India. The analysis is carried out in the same manner; data collection was from social media or from media and gathered total COVID cases were for the last 6 months from February to July 2020 [11]. However, no such cases for COVID has been reported, so the study period is considered from March and carried out till end of July. Further, total COVID cases are subdivided among the number of recovered, fatality, and finally active cases for the month. Data are presented for six states, namely, Delhi, Kerala, Maharashtra, Tamil Nadu, Uttar Pradesh, and West Bengal, in Figs. 9, 10, 11, 12, 13, and 14, respectively. Considering the virus' life time (2–3 weeks), number of active cases or recovered are counted in next month also. For example, some infected cases have been identified in last week of the month, then recovery of these cases will be floated in next month.

Hence, the analysis shows that in Delhi, recovered numbers are exceeding the total infected cases in July. It inferred that new infected cases are declining and existing infected cases have recovered. This result is very good, positive, and motivating. So, the people from other states can also believe that the fight against corona virus can overcome the situation. Analysis for rest of the states also represents that recovered numbers are higher in high magnitude compared to active and deaths.

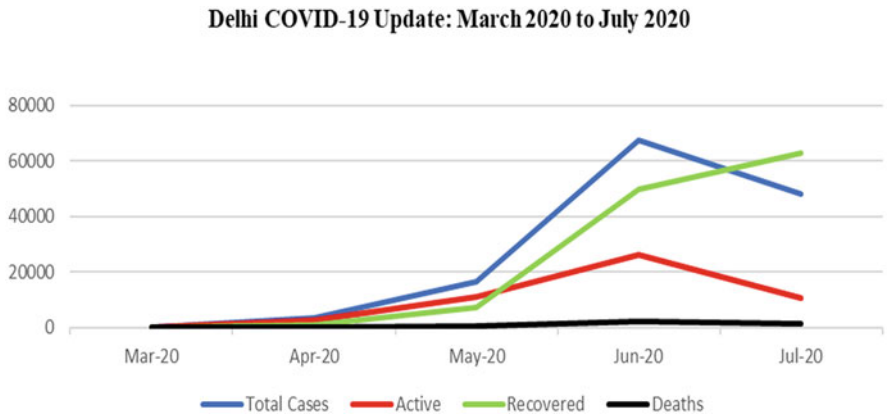


**Fig. 9** COVID data analysis in Delhi

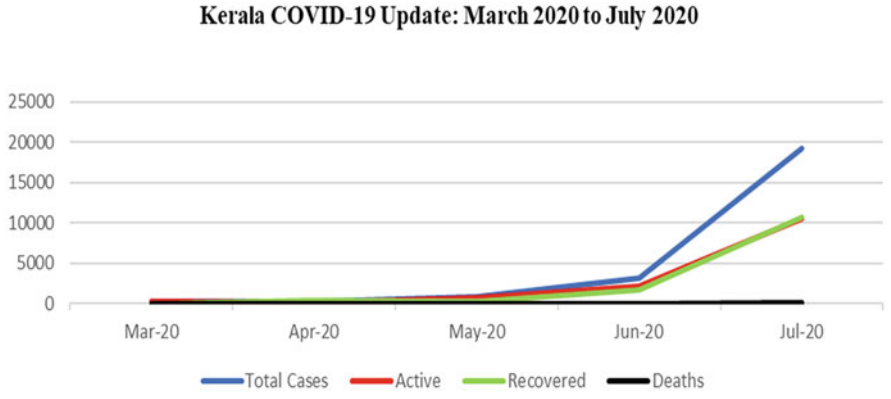


**Fig. 10** COVID data analysis in Kerala

After the prolonged lockdown for consecutive 60 days, the state of West Bengal moves toward "new normal" lifestyle. In case of this "new normal" living, people are aware of corona virus. So, it is expected that even common people will also take necessary prevention while going out for work or for some other purposes. However, through "*test, track, and trace"* methodology, a few locations are identified for complete lockdown. *Complete lockdown* is enforced where the testing shows more number of "positive" results, *partial lockdown* means the locality is released from restrictions for a few hours for sustainable living, and *no lockdown* releases the location for free living. Location or region with "*complete lockdown*" is referred as "*red zone*" or high-risk zone; location or region with "*partial lockdown*" is referred as "*orange zone*" or low-risk zone, and rest of the locations are identified as "*green zone*" [12].

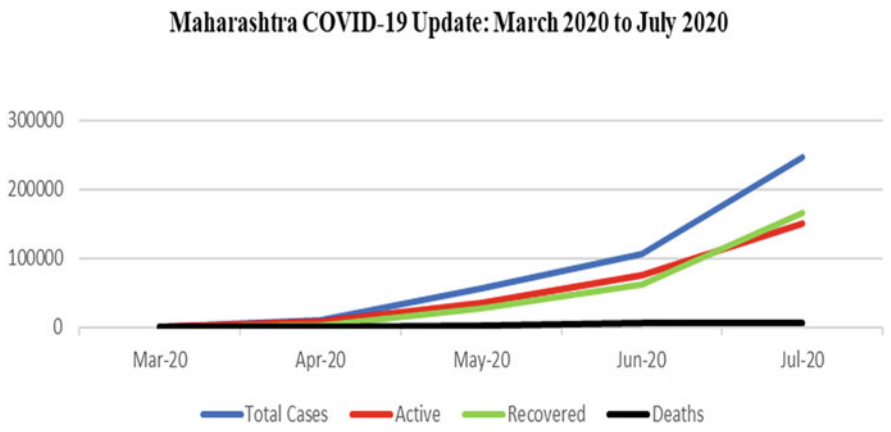


**Fig. 11** COVID data analysis in Maharashtra

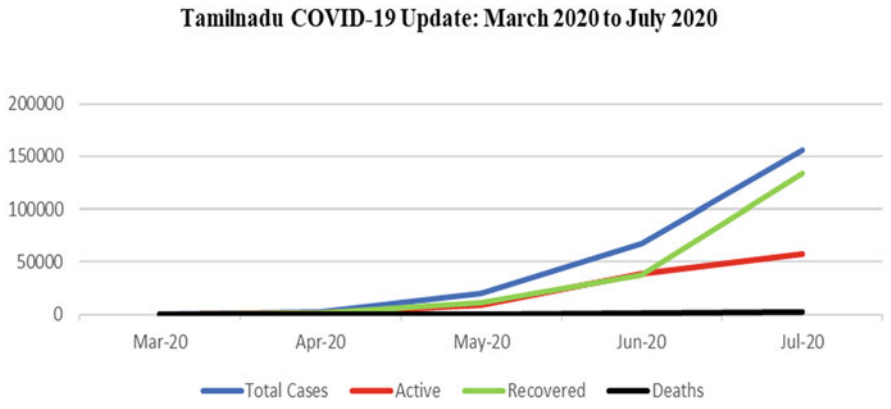


**Fig. 12** COVID data analysis in Tamil Nadu

Table 5 presents the list of zones such as red, orange, and green, which specifically show that various districts of West Bengal fall in these zones.

As of July 31, according to the list, 10 districts of West Bengal are in red zones where the maximum number of coronavirus cases has emerged; five districts have been identified as orange zones in the state, while eight districts have been classified in the green zone.

Next section presents a rule-based learning for the coronavirus pandemic, applicable in healthcare.

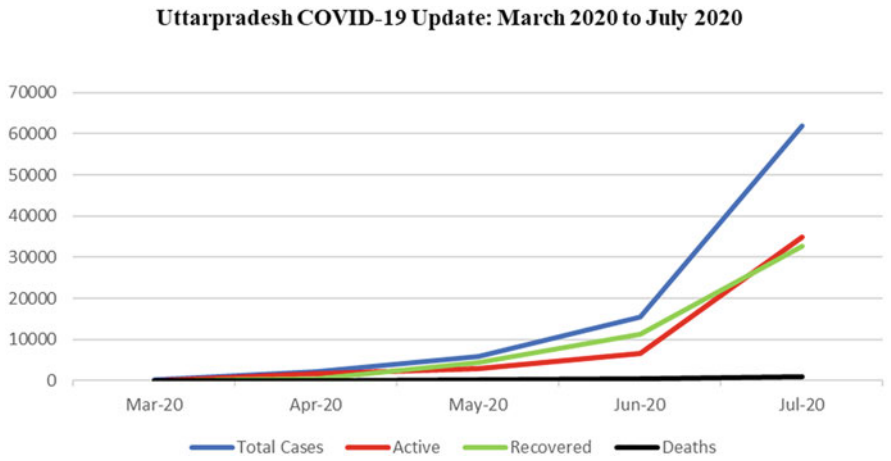


**Fig. 13** COVID data analysis in Uttar Pradesh

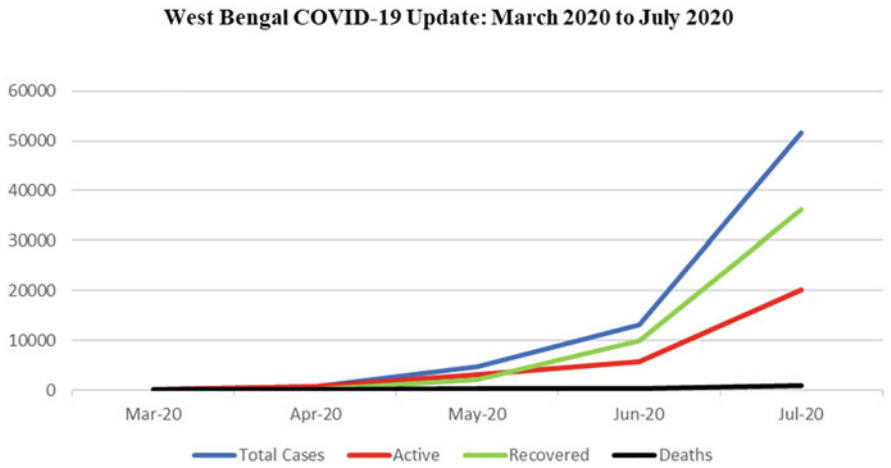


**Fig. 14** COVID data analysis in West Bengal

## 6 Rule-Based Learning in Healthcare Using Pandemic Knowledgebase

The above case studies show that irrespective of the country, caste, gender, and age, people are devastated by the spreading of COVID-19. It is also inferred that the speed of transmission and number of deaths vary from one location to other location. Moreover, it is observed that the spreading of virus may be controlled in a more conservative manner, since the case study illustrated that death cases vary from urban to rural living style. Maharashtra being the most affected state, death

**Table 5** Zone-wise district lists in West Bengal

| District | Zone |
|---|---|
| Kolkata | Red Zone |
| Howrah | Red Zone |
| North 24 Parganas | Red Zone |
| South 24 Parganas | Red Zone |
| Paschim Medinipur (West Midnapore) | Red Zone |
| Purba Medinipur (East Midnapore) | Red Zone |
| Darjeeling | Red Zone |
| Jalpaiguri | Red Zone |
| Kalimpong | Red Zone |
| Malda | Red Zone |
| Hooghly | Orange Zone |
| Paschim Bardhhaman | Orange Zone |
| Nadia | Orange Zone |
| Purba Bardhhaman | Orange Zone |
| Murshidabad | Orange Zone |
| Uttar Dinajpur | Green Zone |
| Bankura | Green Zone |
| Birbhum | Green Zone |
| Cooch Behar | Green Zone |
| Dakshin Dinajpur | Green Zone |
| Purulia | Green Zone |
| Alipurduar | Green Zone |
| Jhargram | Green Zone |

rates are still maximum 3.5% of its total number of cases till date. However, from the international case study, it is proved that top three death rate percentages lie between 14% and 16% of its total number of infected cases.

To control or reduce the spreading, the role of **3-T**'s is very important. These **3-T**'s are *testing*, *tracking,* and *tracing*. The first T, testing, can confirm identification of positive cases. However, India, being a country with a high population density of almost 1380 million people, cannot control the spreading only through testing. So, tracking and tracing are also very important steps for identifying the location. With the speedy transmission of virus, it is also obvious that all the positive cases are not having prominent symptoms, rather the asymptomatic cases also increase gradually [13].

The very common symptoms for COVID-19 are similar to normal influenza: cold and cough [14]. Hence, similar disease is possible through both these viruses, and due to lack of care, respiratory problems may arise. Moreover, according to the analysis of various symptoms from the COVID-19 patients, a few identified symptoms with various measures are listed below:

(a) **Fever** – below 100 °F, between 100 °F and 103 °F, and above 103 °F
(b) **Cough** – moderate to severe, more severe, leading to chest pain

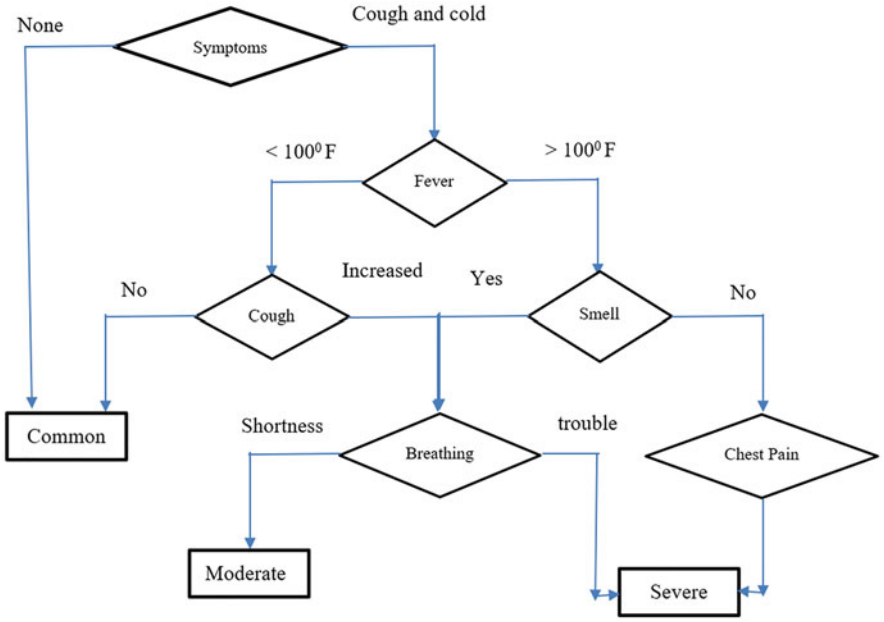


**Fig. 15** Decision tree-based symptoms analysis of COVID-19

(c) **Breathing** – shortness, trouble breathing, and severe
(d) **Smell** – loss of smell, loss of taste, and both persist longer

Decision tree is an important supervised learning technique in machine learning, which is applied to both categorical and continuous variable datasets [15]. Using decision tree-based learning, classification of symptoms of COVID-19 is presented in Fig. 15 [16]. This broad classification can easily identify *common* people who have very less infection rate of the virus; Next category, *moderate* suggests chances of having the positive cases through testing, and *severe* suggests advanced stage of the infected case.

The pattern of spreading clustered in identified area or community transmission may be represented using decision tree-based analysis datasets. The rule-based learning also recommends some preventive measure to control mortality [17]. Table 6 presents a suggested rule-based preventive measure. This preventive measure will be used as knowledgebase for controlling the devastation of further spreading and lead to speedy recovery. At the outset, according to the World Health Organization, the virus is now floating in air also and most probably there is high chance of infection of everyone in every room. Hence, the campaign needs to be carried out to make people aware enough to do self-control by taking proper precaution for not to be infected. Even if they are infected, the rule base will lead them to speedy and proper recovery.

**Table 6** Suggested rule-based preventive measure

| Symptoms category | Suggested prevention measure |
|---|---|
| Common | Home isolation and continue immunity-boosting habits |
| Moderate | Medical care and quarantine in safe home with continuation of immunity-boosting habits |
| Severe | Mandatory hospitalization and follow quality healthcare |

It needs to be mentioned that people need to increase and continue immunity-boosting habits like a few wake-up Ayurvedic starters, such as lemon water or turmeric in empty stomach and mild breathing exercises. These also prevent the spreading of virus and control the transmission speed of the corona viruses. Having major similarities to influenza virus, the major deviation is that children are main carriers of influenza virus, while in the case of COVID-19, children are less affected. The analysis presents that due to lot of vaccinations, children (age group above 10) can strongly fight against COVID-19. Hence, elderly people are at the highest risk of the coronavirus.

## 7 Conclusion

This chapter presents meticulous data analysis of pandemic COVID-19 to suggest rule-based knowledgebase for speedy recovery of the affected patients. The rule-based knowledgebase is prepared using machine learning technique, specifically decision tree-based algorithms. The common and severe symptoms are presented with various boundaries to make aware people of any age, color, caste, and gender. During recent years, a wide range of coronaviruses that cause a wide assortment of human and veterinary sicknesses have happened. Almost certainly, these infections will proceed to rise and cause both human and veterinary episodes inferable from their capacity to recombine, change, and contaminate various species and cell types.

Future research on coronaviruses will keep examining numerous parts of viral replication and autogenesis. In the first place, understanding the inclination of these infections to bounce between species, to build up disease in another host, and to distinguish huge repositories of coronaviruses will drastically help our capacity to foresee when and where potential waves may happen.

## References

1. M. Awad, R. Khanna, *Machine learning and knowledge discovery* (Apress, Berkeley, 2015). https://doi.org/10.1007/978-1-4302-5990-9_2

2. Spanish flu. Accessible at https://en.wikipedia.org/wiki/Spanish_flu. Accessed on 1 Aug 2020 (n.d.)
3. Western African Ebola virus epidemic. Accessible at https://en.wikipedia.org/wiki/Western_African_Ebola_virus_epidemic. Accessed on 1 Aug 2020 (n.d.)
4. M.M. Haby, M. Pinart, V. Elias, L. Reveiz, Prevalence of asymptomatic Zika virus infection: A systematic review. Bull. World Health Organ. **96**(6), 402 (2018)
5. A.J.L.A. da Cunha, M.C. de Magalhães-Barbosa, F. Lima-Setta, R. de Andrade Medronho, A. Prata-Barbosa, Microcephaly case fatality rate associated with Zika virus infection in Brazil: Current estimates. Pediatr. Infect. Dis. J. **36**(5), 528–530 (2017)
6. Coronavirus COVID-19 Live tracker Johns Hopkins. https://www.grainmart.in/news/coronavirus-covd-19-live-cases-tracker-john-hopkins/ (n.d.)
7. COVID-19 World dataset, Worldometer. https://www.worldometers.info/coronavirus/. Accessed on 1 Aug 2020 (n.d.)
8. COVID-19 India dataset. https://www.covid19india.org/. Accessed on 1 Aug 2020 (n.d.)
9. WHO Coronavirus disease. https://covid19.who.int/table. Accessed on 1 Aug 2020 (n.d.)
10. India fights COVID-19. https://www.mygov.in/covid-19. Accessed on 1 Aug 2020 (n.d.)
11. COVID-19: Pandemic in India. https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_India#cite_note-4. Accessed on 1 Aug 2020 (n.d.)
12. English Jagaran: Coronavirus Hotspots in West Bengal: Red Zones in state, 516 areas identified as containment zones. https://english.jagran.com/india/coronavirus-hotspots-in-west-bengal-red-zones-green-zone-orange-zone-bengal-containment-areas-check-full-list-here-10011474. Accessed on 1 Aug 2020 (n.d.)
13. ICMR to study prevalence of asymptomatic COVID-19 infected people to check for community spread. Accessible at https://economictimes.indiatimes.com/news/politics-and-nation/icmr-to-study-prevalence-of-asymptomatic-covid-19-infected-people-to-check-for-community-spread/articleshow/75631365.cms?from=mdr. Accessed on 1 Aug 2020 (n.d.)
14. Everything you should know about coronavirus. https://www.healthline.com/health/coronavirus-covid-19#coronavirus-types. Accessed on 1 Aug 2020 (n.d.)
15. Decision Tree Algorithm. https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb. Accessed on 1 Aug 2020 (n.d.)
16. Introduction to Decision Tree. https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html. Accessed on 1 Aug 2020 (n.d.)
17. A. Das, A. Desarkar, Decision tree-based analytics for reducing air pollution. J. Inf. Knowl. Manag. **17**(02), 1850015 (2018)