Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Study and impact analysis of COVID-19 pandemic clinical data on infection spreading

Sasmita Parida\(^2\), Aisworya Mohanty\(^1\), Suvendu Chandan Nayak\(^2\), Bibudhendu Pati\(^1\), Chhabi Rani Panigrahi\(^1\)

\(^1\)DEPARTMENT OF COMPUTER SCIENCE, RAMA DEVI WOMEN’S UNIVERSITY, BHUBANESWAR, ODISHA, INDIA; \(^2\)DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY, GITA AUTONOMOUS COLLEGE, BHUBANESWAR, BPUT, ODISHA, INDIA

1. Introduction

The novel coronavirus disease 2019 (COVID-19) is a continuous general well-being crisis of universal noteworthiness [1]. There are huge information holes in the study of disease transmission, transmission elements, examination apparatuses, and the board. The advancement of another coronavirus plague started since the most recent three decades. The differing transmission designs, in particular, social transmission, and spread through somewhat symptomatic cases are a zone of concern. In December 2019, the novel coronavirus was detected in the city of Wuhan, China [1]. The significance of this disease was completely hidden and the exact pharmaceutical invention related to this disease is still unidentified. For this reason, more and more people are affected and this disease has spread all over the world in a very short time [2]. The number of infected people is growing day by day and thousands of people have died in this dangerous tragedy, but there is no appropriate way to stop the disease. Everyone in the world is hopeless, and within 4 months the virus enters 212 countries in the world. The United States, Italy, China, Spain, Germany, France, Iran, the United Kingdom, Switzerland, and Netherlands are the top 10 harshly affected countries in the world. In January 2020, the coronavirus infection was declared as a global health emergency by the WHO because of its harmful and super spreading nature, as it affected not only China but also the whole globe. Finally, the novel coronavirus was renamed as COVID-19 in February 2020 [1]. Then the COVID-19 outbreak was declared as a world pandemic on March 2020 [3]. Due to life risk and because it specially affects the respiratory system, the virus was called as “severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)” by the International Committee on Taxonomy of Viruses (ICTV), and WHO announced the new disease as
COVID-19 [4,5]. The number of infected people is growing rapidly day by day, as there is no medication to stop the spread.

COVID-19 is an infectious disease caused by the novel coronavirus mutated to the human body and survived for the first time [1]. It is highly contagious and spreads from person to person easily through contamination. People may be sick and detected with symptoms after 14 days from getting infected. The most common symptoms are high fever, tiredness, dry cough, and difficulty in breathing [6], and in some cases, symptoms such as muscle pain, confusion, headache, sore throat, rhinorrhea, chest pain, diarrhea, nausea, and vomiting may be found [7]. The COVID-19 virus enters the human body through the nose, mouth, and eyes and directly affects the cells of the body. Then human body cells produce the protein called angiotensin-converting enzyme 2 (ACE-2), which binds the COVID-19 virus and produces new coronaviruses. The COVID-19 virus first infects the cells that line the throat. Inside the human body, different cells start forming new coronaviruses and every infected cell is able to produce millions of viruses that make a human being unable to survive. In infected patients the virus is generally located in the back side of nose termed as nasopharynx and oropharynx. Till now, there is no proper medication and preventive vaccine, but scientists are effortlessly searching and working on it. In the present scenario, only precaution is the best way to save human beings from this deadly pandemic. Staying at home and sanitizing hands again and again are advised as precautionary steps for preventing COVID-19. But unfortunately, some people are infected by this deadly disease and are advised to contact the health department immediately for proper medical observation and contact tracing. Similarly, the infected people are warned to undergo home isolation or hospital isolation for 14 days to avoid the spread of COVID-19 [8].

Day by day, more number of patients suffer from this hazardous disease in the world and increasing rapidly with respect to recovered patients from this syndrome with a huge number of dead patients. Thus different countries all over the world are locked down for some days to save their citizens. Similarly, the Indian government announced a 21-day lockdown phase wise to fight against COVID-19 and a complete restriction was being enacted on people from moving out of their homes, as it is very important to break the chain of infection. To save the health crisis, the world economy has crashed severely by this outbreak. In other words, the spread of COVID-19 should be controlled analytically to save the whole mankind. As there are no proper pharmaceutical products or specific medications available to avoid this outbreak, mostly researchers follow data analytics to analyze the clinical data of different patients [9].

Machine learning (ML) is a subset of artificial intelligence (AI) that deals with different types of learning algorithms [10]. Fig. 12.1 shows interrelationship among ML, AI, deep learning, and data science. Data scientists mainly used ML approaches to process the business as well as medical data to draw outcomes [11]. Those outcomes help develop intelligent models for future use. ML can be applied in many fields such as agriculture, anatomy, banking, computer vision, healthcare, economics, etc. It is broadly categorized into three types [12]: supervised learning, unsupervised learning, and reinforcement learning.
In addition to the current practices and therapies, AI can assure a new healthcare model. Several specific AI methods are used to evaluate data and decision-making processes based on ML algorithms. Basically, for early detection and diagnosis of diseases, AI-based tools are more useful [13]. We know prevention is better than cure, so by using several ML techniques, researchers can try to predict the disease in its early stage for customary medication treatments, solution decoction, and medication treatments. It is essential to handle such situations using different computational technologies. ML-driven software can be used to classify COVID-19 outbreaks with prediction of extension of their spread around the world. Generally the basic principle of ML-driven tools is that they need enough training data and should process them [14]. A conventional ML technique requires a spotless arrangement of commented on information, with the goal that classifiers can be all around prepared, which falls under the scope of supervised learning.

Over the past decades, huge advancements have been made in settling numerous issues of a few medical pandemic outbreaks. Fig. 12.2 shows a brief idea about the applications of ML in the medical industry [15]. ML requires a lot of training information to be prepared from the COVID-19 clinical data [16]. The essential thought behind the utilization of ML is not only to maintain a strategic control over COVID-19 for spreading but also to focus on building a model for precautionary measures. Nowadays, prediction of the coronavirus infection is becoming a demanding topic, as it has affected a large number of the population in a very short period and a large number of people have died due to the lack of proper medication [17]. ML plays a vital role in the continuous development of complex medicines for diseases. So many researchers are putting their efforts to predict the disease in the early stage.
Decision tree is a type of ML algorithm that is based on the classifier model in the custom of tree structure to solve classification, regression and feature selection problem [18], and the process of breaking the larger data into smaller form. According to the rules studied from the given dataset, it intends to predict the target class [19]. It is one of the best algorithms used by researchers because of its advantages. It works well in data preprocessing system and solves classification problems with limited resources of dataset. It has the capability to handle continuous and categoric data [20]. It is also useful for diagnosis of existing chronic diseases and their impact on COVID-19 [1,5].

The chapter is organized as follows: Section 2 discusses the state of the art in the current study and tools yet to be designed for COVID-19. The clinical data processing and how ML mechanisms can be useful is discussed through different steps in Section 3. A case study of clinical data analysis of India is presented in Section 4. Section 5 highlights the ML-based model for prediction as the primary model for further research. The chapter concludes with some future scope of work in Section 6.

2. Related work

In the past few months, various techniques such as data analysis, ML, deep learning, etc. are used to design models to control the spread of COVID-19 as a precautionary measure. Huwen et al. [21] in their work estimated the actual growth of the infection, an average number of infections caused by each infected person, a secondary infection caused by each infected person, the number of new infections decreased in a particular
time, and the number of death rate in a certain period in Wuhan, China. Jasper et al. [22] studied by taking a medical history of a few patients in a house; they saw that the patients who are infected transmitted the infection to other healthy members in that house and they ensured that the virus is transmitted from one person to another by their close contact with each other.

Santosh [23] proposed AI-driven tools that had many advantages to identify COVID-19 outbreaks and also guessed the cause of spread all over the world. These tools are predictable to have an active learning-based cross-population train or test models that hire multitudinal and multimodal data. Without any social distancing as well as close contact with a family member or anyone else the infection has a strong capability to spread from one person to another easily. Ghinai et al. [24] focused on consideration of the rise and fall rate of the number of patients affected in COVID-19 using parameters such as age, stages of the disease, fatality rate, and the number of patients died. Jennifer [25] compared COVID-19 with severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). In this paper, by combining the microbiologic research as well as the well-known information about SARS-CoV-2, the authors provided actionable and attainable guidance to decision makers, building operators, and all indoor occupants, trying to diminish infectious virus transmission through environmentally arbitrated ways.

The 2019-nCoV infection is more expected to affect old-age patients with comorbidities and can result in severe respiratory diseases such as acute respiratory distress syndrome (ARDS). In general, features of patients who died with reference to MULBSTA score, which is an early warning technique for predicting mortality in viral pneumonia [6]. MULBSTA is a calculated score using the factors as multilobular infiltration, hypolymphocytosis, bacterial coinfection, smoking history, hyper-tension and age. Khot and Nadkar [26] presented in their work that COVID-19 is a continuing public health crisis. In this work, the authors collected different confirmation data about the disease and then compared different terminologies related with SARS, MERS, and COVID-19. Heymann [2] explained about the early stages of the infection and features of patients who are infected with 2019-nCov, MERS, and SARS. Ramadan and Shaib [27] presented a brief description of MERS-CoV such as the symptoms related to COVID-19, along with a health approach and how to control and prevent the spread of the disease.

Res et al. [28] reported that 104 (1.8%) out of 5911 SARS patients were tested positive for COVID-19 and the data was collected from 52 districts of 20 states/union territories. The most of the COVID-19-positive patients were men and the patients’ age was more than 50 years. Among all patients, 40 (39.2%) COVID-19 cases were having no travel history from infected countries and no contact history with infected persons. States et al. [9] provided a brief description of the terminologies related to COVID-19, such as incubation period, quarantine period, period of virus shedding, transmission process of virus, the character of droplets, aerosols and fumets about the infection control and prevention, importance of personal protection, severity of this virus, and fatality rate with regard to the United States. Wang et al. [21] stated the severity related to the
COVID-19 outbreak and the classification of different stages, with the stages represented as $R_0$, $R_1$, and $R_T$. Estimation of the epidemic movement in Wuhan, China, assuming the current avoidance and control procedures was not sufficient.

Mubarak et al. [29] presented information on both innate and adaptive immune responses to MERS-CoV, the recent state of the MERS-CoV vaccine development, and also the development of the MERS-CoV immune response that successively added to the control of MERS-CoV infections. Mandal et al. [30] discussed the main goal about how to find out the best way to avert or postpone the local outbreaks of COVID-19 through restraints on travel from abroad and in case if the virus has already entered the country. The authors also reported the importance of the entry screening of all travelers who journeyed from affected countries and then the role of quarantine of symptomatic patients to avoid the infection.

3. Clinical data analysis

Mostly the effects of outliers in medical datasets should be considered either for selection or to filter out noisy data from a given training medical dataset so that the output has a positive impact on final imputation tests. Fig. 12.3 shows the clinical data observation rate toward chronic disease solutions [31].

By comparing the outputs of classification obtained through combination of instance selection and imputation are represented on the base line. Thus relevant attributes that are related to the COVID-19 outbreak were presented here. The various important attributes related to COVID-19 is given in Table 12.1. It is the list of attributes of suspected patients, either having COVID-19 symptoms or not and not tested positive or not recovered.

It is found by clinical data analysis that in most cases men are more affected by COVID-19 than women. This is because generally men go outside for many reasons. In some countries, previously, governments recorded that old people have maximum chances to be affected by COVID-19 because of their low immunity and have very less chance to recover from this outbreak. It is also observed that patients with medical history related to heart disease, high blood pressure, and diabetes are more prone to be
infected by the COVID-19 pandemic. It is also said that low-level immunity increases the risk factor to recover from COVID-19. The final outcome is to predict the patients having COVID-19 positive (+ve) report or negative (−ve) report. Table 12.2 presents a variety of symptoms related to the COVID-19 pandemic in different regions.

### Table 12.1 List of COVID-19 attributes.

| Attribute name               | Explanation                                                                 |
|-----------------------------|-----------------------------------------------------------------------------|
| Patient’s identification number | Unique identification of the patient                                         |
| Gender                      | Patient is male or female                                                    |
| Age                         | Age of the patient (numeric)                                                |
| Asymptomatic                | Does not carry any symptoms or unusual symptoms                             |
| Symptomatic                 | Symptoms such as high fever, dry cough, difficulty in breathing, etc.        |
| Chronic disease             | Mostly previous medications for (heart disease/diabetes/blood pressure/kidney disease) |
| Travel history              | It shows the international/domestic travel history of the patient or related relatives |
| Reporting date              | It shows starting date of patient admitted to hospital for medication        |
| Immunity power              | Immunity level to fight against COVID-19                                     |
| COVID-19                    | Positive 1 or negative 0 (classification)                                   |

Table 12.1 COVID-19, coronavirus disease 2019.

### Table 12.2 The medical geographies of the COVID-19 outbreak according to different authors.

| Author name     | Fever | Cough | Myalgia and fatigue | Sore throat | Dyspnoea | Diarrhea |
|-----------------|-------|-------|---------------------|-------------|----------|----------|
| Chen et al. [6] | 83    | 82    | 11                  | 5           | 31       | 2        |
| Wang et al. [32]| 98    | 59    | 70                  | 17          | 31       | 10       |
| Huang et al. [33]| 98    | 76    | 44                  | 0           | 55       | 3        |

Table 12.2 COVID-19, coronavirus disease 2019.

3.1 Feature extraction

Nowadays many datasets are available that contain more number of attributes or features and work with all the attributes for a single problem, but it is a very lengthy process for researchers. In this case, before working on a dataset the researcher can moderate the dataset by taking some important features by considering the relevant information related to the problem. In the context of COVID-19, we can extract some important features from the dataset based on our requirement, and in our case study, we consider only nine essential features to perform analysis on the clinical data of COVID-19.

3.2 Feature selection

Feature selection is also known as variable selection. It uses data preprocessing techniques of data mining and is mainly used for declining of data by eradicating some
attribute values. This process progresses the performance of prediction, decreases the training time of the algorithm, and also offers better imagining of data. Feature selections have many application areas such as healthcare, ensemble technique, embedded method, etc. In the ML approach, feature selection method is categorized basically into three types: (1) filter method, (2) wrapper method, and (3) embedded method.

### 3.3 Data preprocessing

Before using the dataset, it is very important to preprocess the data. When we expect better output, it is mandatory to preprocess the data. Normally preprocessing of the data includes two steps. First, it is understood that the dataset contains the value zero for missing data. We detached all the cases that have the value zero for an exact field. Second, data discretization is applied as a method of altering continuous attribute values into a finite set of intervals and combining some precise data with each interval. Discretization expressively develops the quality of discovered data and also decreases the running time of several data mining tasks such as association rule finding, classification, and prediction. Table 12.3 portrays clinical records of patients that were broke down by the Department of Critical Care Medicine [6,32,33]. Epidemiologic, clinical, research facility, and radiologic qualities and treatment and results information were acquired with information assortment structures from electronic clinical records. The information was looked into by a prepared group of doctors. The data recorded by that group includes segment information, clinical history, introduction history, hidden comorbidities, manifestations, signs, research center discoveries; chest computed tomography (CT) outputs, and treatment measures.

### 3.4 Stages of coronavirus disease 2019

According to the WHO, there are four stages of COVID-19 outbreak in India.

*Stage 1:* When cases are only introduced from other affected countries and thus only those who have a travel history from abroad test positive and there is no proof to the spread of the disease locally.

| Author name       | ARDS | Oxygen requirement | ECMO | Shock | Noninvasive ventilation | Invasive mechanical ventilation | Acute cardiac injury | Acute kidney injury | Renal replacement therapy | Death |
|-------------------|------|--------------------|------|-------|-------------------------|---------------------------------|---------------------|---------------------|--------------------------|-------|
| Huang et al. (n-41) % | 55   | 66                 | 5    | 7     | 24                      | 5                              | 12                  | 7                   | --                       | 15    |
| Wang et al. (n-138) % | 31.2 | 76                 | 2.9  | 8.7   | 10.9                    | 12.3                            | 7.2                 | 3.6                 | 1.45                     | 4.3   |
| Chen et al. (n-99) %   | 31   | 76                 | 3    | 4     | 13                      | 4                              | 13                  | 3                   | 9                        | 11    |

*ARDS,* acute respiratory distress syndrome; *COVID-19,* coronavirus disease 2019; *ECMO,* extracorporeal membrane oxygenation.
Stage 2: When cases are familiarized, local transmission from infected persons occurs. This will generally be families or colleagues of those patients who traveled from abroad and test positive and had close contact with the infected persons. A very smaller number of people were affected in stage 2. As the origin of infection is well known, it is easy to do contact tracing and stop the infection by applying 14 days quarantine to all the people who are closely connected with the patients.

Stage 3: It is the stage of community transmission, where people who have not been in close contact with the infected persons or have no travel history to affected nations still test positive, and in this stage, people are incapable of detecting from where they might have contained the virus.

Stage 4: It is the most severe stage of the disease. In other words, in this stage the infection goes to an epidemic condition and rapidly spreads. More number of people are infected and it is much harder to regulate the spread of the disease.

3.5 Incubation period for the COVID-19 virus

Generally the development period of this virus is expected to be in between 2 and 21 days according to some health associations [3].

- According to the WHO, the incubation period for coronavirus is normally 2–10 days.
- China’s National Health Commission (NHC) had expected an incubation period for this virus from 10 to 14 days.
- The United States’ CDC (Centers for Disease Control and prevention) reported that the incubation period for this virus is between 2 and 14 days.
- A leading Chinese online community for physicians and healthcare professionals (DXY.cn) estimates the incubation period for COVID-19 is 3–7 days and it may be up to 14 days.

In some cases the exception occurs while estimating the incubation period for the patient. A case with 27 days of incubation period has been recounted by the local government of Hubei province on February 22, 2020 [34]. Similarly, another case with incubation period of 19 days was reported in a JAMA study on February 21, 2020 [21].

3.5.1 Symptoms developed in incubation period

Day 1: In day 1, symptoms start, and patients may suffer from fever and dry cough. In some other cases, symptoms such as fatigue and muscle pain occur. In case of children may have affected in diarrhea or nausea before 1 or 2 days of the symptoms developed.

Day 5: In severe cases, symptoms start to go downhill. Old-age persons with preexisting medical conditions might experience difficulty in breathing.

Day 7: Generally this is the average period in which the patient infected by the new virus is admitted to the hospital, which is declared by a study conducted by the Wuhan University.
Day 8: This day is very dangerous for a patient. In most cases a patient with severe condition will develop breathing difficulty, pneumonia, and ARDS. Normally ARDS occurs when fluid collection is found in the lungs of patients. It is a very severe case of this virus infection; around 15% patients enter this period.

Day 10: If the disease aggravates in the patient, then the patient might require immediate admission to the ICU. The patients possibly have severe pain in the abdomen and loss of appetite than the starting phase.

Day 17: In average case, patients who recovered from the precarious disease may get discharged from the hospital.

4. Case study

This section presents a detailed case study about the outbreak of COVID-19 in India.

- First finding of the COVID-19 outbreak in India
  The first COVID-19-positive case in India was reported on January 30, 2020, in Thrissur district, Kerala. The patient was studying in Wuhan, China, and also had a travel history from Wuhan, China. After that, within 4 days, another two positive cases were reported in Kerala’s Alappuzha and Kasaragod districts and also these two patients were studying in China and having recent travel history from China. For that reason on February 3, 2020, the state government of Kerala decided to announce a state-level emergency. Then the government started the process of contact tracing of those patients. Around 3400 suspected persons, who had come in contact with them, were quarantined for 14 days and they were kept under 24-hour medical observation. After that no COVID-19-confirmed cases appeared in India until March 1, 2020. On March 2, 2020, again two confirmed cases were found in Delhi and Hyderabad, and those two patients also had a travel history from abroad.

- Different steps initiated by the Indian government
  In India, when the virus was in the first stage, some other countries were in the second or third stage and their situation was worse than that of India. At that time the Indian government had taken the decision to rescue more number of immigrants who are facing lots of troubles in other affected countries. So to avoid transmission of the virus coming through the immigrant people, thermal screening was done at the airport and railway stations to detect the symptoms present in the people who travel from foreign countries (immigrant people). But the screening process alone was not sufficient because some people are asymptomatic and in some cases symptoms appear late. Among them, some had hidden their travel history and lived as usual at home and came in close contact with others instead of living in home isolation and because of this the virus slowly spread all over India.

  After that when India was in the second stage of this pandemic, the Indian government decided to immediately detect the immigrant people and kept them
strictly in quarantine, and also the government opened a registration desk and introduced a toll-free number, 107, for all people. The government requested all the immigrants to call in this number, register their name and phone number, and also give brief information about their travel history. By the time, the number of positive cases increased day by day. For that reason the Indian government declared a Janata Curfew on March 22, 2020, and then declared 21 days nationwide lockdown on March 24, 2020. In that period, people were completely restricted from going outside their homes, all trains and planes were canceled, and all boarders were sealed by the Indian government. After 21 days, the Indian government observed that the current status of COVID-19 is very sensitive in the whole world. But in India the COVID-19 status is far better than that in other well-developed countries. This is only possible for early lockdown in India. After observing the situation, again the Indian government decided to extend the lockdown for 19 days, as only lockdown is the best way to break the corona infection chain.

- **Current status of COVID-19 in India**
  On April 22, 2020, the number of confirmed cases in India was increased by 20,080, the number of patients recovered from this pandemic was 3975, and the number of deceased patients reached 645. Till April 20, a total of 201 laboratories were working in India for testing the coronavirus infection and a total of 4,49,210 samples had been tested in India.

  A total of 28 states and 9 union territories were affected in India till April 22, and Maharashtra was known as the center of COVID-19 in India. Besides that, Gujarat, Delhi, Rajasthan, Madhya Pradesh, and Tamil Nadu are the most affected states in India except states such as Odisha and Kerala, which were successful to fight against COVID-19 according to a report from the Ministry of Health and Family Welfare.

- **Characteristics of this virus**
  It is noticed that the virus can affect men, women, kids, and also senior citizens in India, but more numbers of men are affected in comparison with others. In the European countries, mostly the symptoms of COVID-19 are found within 14 days. But, in India, it is found that in some cases the people are affected by COVID-19 without any symptoms. The reason behind this is reported as the Indian people have good resistance power and hence the symptoms are found after 21 days. This could also be because most of the Indians belong to the Hindu religion and many people prefer vegetarian foods over nonvegetarian. So due to the eating habits of healthy vegetables, Indians have a strong immune system and hence the symptoms are shown lately.

- **Technical support system**
  Now the Indian government developed a freely accessed software application named as “Aarogya Setu”. By this application, one can get complete information about COVID-19 in 11 languages as well as get some useful guidance to prevent it.
The most useful feature of this application is that it can notify the user if a COVID-19-positive case will be identified near his/her location. Another mission working toward this virus is known as ministry of Ayush Mantralaya. Through this mission, public can get the appropriate information about how to improve the immune power of a person by drinking hot water and taking vitamin C, which is available in foods such as orange, amla, etc. All the latest information about COVID-19 is updated in every minute in all social media, and we are getting this information through TV, radio, newspaper, etc.

- **Financial support system**
  India is now in the second stage of this infection, but day by day the situation is getting worse. The people coming from outside states are also kept in quarantine for 14 days. State-to-state and inter-state communications are banned to avoid the transmission of this virus. For this reason, more number of poor laborers lost their work. So to avoid poverty the government offers money, free food products, and other essential products to all the affected laborers.

- **Categorization of the country into color-coded areas**
  On April 15, 2020, depending on the level of infection, the Indian government divided the country into color-coded areas. Red zone specifies an infection hotspot, orange zone indicates some infection, and green zone denotes no infections found in the area. A red “hotspot” district will be entitled as orange zone if no new cases of COVID-19 are reported within 14 days. Similarly if no fresh cases are reported within 28 days then the district will become a green zone. Till now the government has identified 170 hotspot districts in 25 states and 207 nonhotspot districts in 27 states.

- **Containment zone**
  According to the Ministry of Health and Family Welfare, if any COVID-19 new cases are found in a particular geographic area then the area should be declared as a containment zone. In this area, movement of people is completely restricted to prevent further spread of the virus. The government has formed an interministerial team to visit the containment areas and inspect proper lockdown in those areas.

- **Rapid test kit**
  The Indian Council of Medical Research (ICMR) had given permission to use rapid test kits to speed up the screening and detection of COVID-19. But, later it was found that there is a huge variation and inaccuracy with the results of the rapid test kit. Basically it was used for early diagnosis, surveillance, and viral thread detection.

- **Pool testing**
  As the count of the coronavirus-infected patients keeps on going up in India, researchers and clinical specialists are scrambling to discover approaches to control the spread of the virus. Clinical specialists had devised another new methodology to speed up COVID-19 testing in India, which is known as sample pooling (or bunch testing), and aim to gear up testing and screening at more populated and contaminated areas and is a mathematical concept for mass screening.
In pool testing, samples collected from the pool or a group of people are mixed and considered as a single swab for testing to reduce the number of tests required to detect positive samples. If the test report of pool testing is found then each person’s sample is tested one by one. The advantages of pool testing include

- **(a)** prevention of infection against contact spreading,
- **(b)** rapid screening rate,
- **(c)** effective with less resources,
- **(d)** gears up the speed of the test by 3–4 times more.

**Clinical analysis**

According to a study conducted by a global consulting firm Protiviti with times network, the COVID-19 outbreak could reach its peak in India till mid-May [35]. In India, around 75,000 active cases were detected on May 22, 2020. According to this research, if lockdown is extended after May 3 then one infected person can affect 0.3 persons. If the lockdown will extend till May 30, then the virus will disappear in less-affected states in early June but the virus will be active in more affected states till August 30. On September 15 the virus will completely disappear from India. The cumulative data of infected, recovered, and death data day-wise from March 1, 2020 to April 15, 2020 is shown in Fig. 12.4. It is observed that after 30 days, i.e., by the end of March, the infection rate gets doubled and again it gets doubled within countable days.

The day-wise infection is recorded in two digits up to March 26 and after that, as the graph in Fig. 12.4 indicates, the infection increased rapidly. After 17 days of time interval the infection reached to four digits on April 13 and 14, as shown in Fig. 12.5. The number of deaths per day is quite less; a single digit number is recorded for a maximum number of days. On April 14 a maximum number of 42...
death records had been noted. The day-wise recovery was quite better and after April 12 the number jumped to three digits and is continuing as highlighted in Fig. 12.5.

The day-wise death and recovery percentage is shown in Fig. 12.6. The percentage of death rate per day reached a maximum of 9.09%, which was in the initial stage, and 8.42% was recorded as the second highest. The maximum recovery percentage per day was recorded as 21.77%, which was around 22%. As per reports the average recovery rate was increased slowly to more than 22%.

5. Proposed model for the prediction of COVID-19

Gathering enormous measure of information is not unimportant but one needs to wait long. The vast majority of the announced ML-driven models are restricted to verification
of models for coronavirus cases. ML specialists express the way that stipulated data may slant outcomes from the seriousness of the coronavirus pandemic. Hence we used an ML-based mechanism to learn the model over time with limited knowledge about the data. In a nutshell, the ML-driven model promotes self-learning with the presence of medical experts. In this ML-driven model proposed for COVID-19 prediction, we use decision tree, which is a supervised learning-based classifier [36].

Decision tree is a procedure that iteratively breaks the given dataset into at least two sample data. The objective of the strategy is to foresee the class prediction of the outcome variable. It assists with isolating the informational data and generates models to predict the class of unknown sample. It gives optimal value of leaf nodes with high entropy [37–39]. The accuracy of the decision tree depends on a few parameters such as attribute splitting, stop criteria, size of training sample, and outliers.

**Input:** Training dataset.

**Outcome:** Tree-structured decision model.

Fig. 12.7 shows the structure of a decision tree model, which includes decision nodes and leaf nodes. Here the leaf node positive indicates COVID-19(+) and negative indicates COVID-19(−). Each decision node splits into either another level decision node or leaf node based on the condition. Here the choice of attribute for root node is very challenging. It depends on the gain given by the Gini index based on the impurity degree of child nodes. There can be two or more splits of a decision node. The accuracy

![Decision tree model of coronavirus disease 2019 (COVID-19) prediction.](image)
of the decision tree can be measured by entropy, Gini index, misclassification rate, and precision [40]. The decision tree model predicts a tree structure with best impactful attributes toward the correct classification and for better decision.

6. Conclusion and future scope

The COVID-19 is still an ongoing outbreak as well as a serious intimidation to the health of all people across the world. The virus is spreading from person to person and transmitted by close contact via airborne droplets generated by coughing and sneezing. So people should avoid visiting the market, shopping mall, or any public place as far as possible and also stay away from sick people. There is no medical therapy to prevent the coronavirus infection. So in the present scenario, only precaution is the preeminent way to save human beings from this deadly pandemic. Social distancing and sanitizing hands repeatedly are advised as precautionary steps for COVID-19. Although the mortality rate of this virus is less than the virus causing SARS and MERS, because of its high transmission rate it spread all over the world and became pandemic. National and international healthcare organizations publicized apposite harmonization in conduction of this outbreak till now. It is a matter of fact that we have a little information about this virus and also related to the source of the virus, but it is still difficult and unclear that we will develop a vaccine and win over this pandemic COVID-19 outbreak.

In this chapter, we enlightened the possibility of the future of the COVID-19 pandemic. The significance of the ML-driven prediction models and their importance have been presented to control the rapid outbreak. The basic objective is that ML-based models do not always require a complete dataset for training and testing. In addition, considering the spread rate of the COVID-19 outbreak in India, we presented a case study on the current scenario of COVID-19 in India and its analysis. Finally, decision-tree-based solution approach is discussed to predict the COVID-19 pandemic, with consideration of different influencing parameters.

The future scope in this area is identified as follows: various deep learning strategies can be applied on clinical image dataset to identify COVID-19 patients. Hybrid deep learning frameworks and deep transfer learning approaches can also be useful to identify COVID-19 cases at an initial stage. Smart masks with sensors can be developed to detect corona-positive patients. Another direction of research in this area may be an accurate prediction of gathering of people at certain places such as shopping mall, workplace, social events, etc. so that people can be aware of maintaining social distancing, which is one of the important factors to keep ourselves away from the virus.

References

[1] X. Zhang, Y. Tan, Y. Ling, et al., Viral and host factors related to the clinical outcome of COVID-19, Nature (2020), https://doi.org/10.1038/s41586-020-2355-0.

[2] D.L. Heymann. A novel coronavirus outbreak of global health concern, Lancet 395 (2020) 15–18, https://doi.org/10.1016/S0140-6736(20)30185-9.

[3] S. Report, Coronavirus Disease 2019 (COVID-19), vol. 2019, 2020.
Chapter 12 • Study and impact analysis of COVID-19 pandemic clinical data

[4] J.F.W. Chan, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet 395 (10223) (2020) 514–523, https://doi.org/10.1016/S0140-6736(20)30154-9.

[5] D. Pinto, Y. Park, M. Beltramello, et al., Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody, Nature (2020), https://doi.org/10.1038/s41586-020-2349-y.

[6] N. Chen, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet 6736 (20) (2020) 1–7, https://doi.org/10.1016/S0140-6736(20)30211-7.

[7] N. Chen, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet 395 (10223) (2020) 507–513, https://doi.org/10.1016/S0140-6736(20)30211-7.

[8] I. Ghinai, et al., First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA, Lancet 395 (10230) (2020) 1137–1144, https://doi.org/10.1016/S0140-6736(20)30607-3.

[9] U. States, D.B. Jernigan, CDC COVID-19 Response Team, Update: Public health response to the coronavirus disease 2019 outbreak, Morb. Mortal. Wkly. Rep. 69 (8) (2020) 216–219.

[10] P. Shah, et al., Artificial intelligence and machine learning in clinical development: a translational perspective, NPJ Digit. Med. 2 (1) (2019). https://doi.org/10.1038/s41746-019-0148-3.

[11] F. Wu, et al., Deep learning applications and challenges in big data analytics, IEEE Commun. Surv. Tutor. 2 (1) (2019) 262–272. https://doi.org/10.4103/ijmr.IJMR.

[12] S. Das, et al., Applications of artificial intelligence in machine learning: review and prospect, Int. J. Comput. Appl. 115 (2015) 31–41.

[13] F. Jiang, et al., Artificial intelligence in healthcare: past, present and future, Stroke Vasc. Neurol. 2 (4) (2017) 230–243. https://doi.org/10.1136/svn-2017-000101.

[14] Y.K. Chan, Y.F. Chen, T. Pham, W. Chang, M.Y. Hsieh, Artificial intelligence in medical applications, J. Healthc. Eng. 2018 (2018). https://doi.org/10.1155/2018/4827875.

[15] T. Davenport, R. Kalakota, The potential for artificial intelligence in healthcare, Future Healthc. J. 6 (2) (2019) 94–98. https://doi.org/10.7861/futurehosp.6-2-94.

[16] Q. Guo, et al., Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm, bioRxiv (2020). https://doi.org/10.1101/2020.01.21.914044.

[17] J.A.M. Sidey-Gibbons, C.J. Sidey-Gibbons, Machine learning in medicine: a practical introduction, BMC Med. Res. Methodol. 19 (1) (2019) 1–18. https://doi.org/10.1186/s12874-019-0681-4.

[18] N. Nai-Arun, R. Moungmai, Comparison of classifiers for the risk of diabetes prediction, Proc. Comput. Sci. 69 (2015) 132–142. https://doi.org/10.1016/j.procs.2015.10.014.

[19] D. Sisodia, D.S. Sisodia, Prediction of diabetes using classification algorithms, Proc. Comput. Sci. 132 (Iccids) (2018) 1578–1585. https://doi.org/10.1016/j.procs.2018.05.122.

[20] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvard, Machine learning and data mining methods in diabetes research, Comput. Struct. Biotechnol. J. 15 (2017) 104–116. https://doi.org/10.1016/j.csbj.2016.12.005.

[21] H. Wang, et al., Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China, Cell Discov. (2020) 4–11. https://doi.org/10.1038/s41421-020-0148-0.

[22] J.F. Chan, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet 6736 (20) (2020) 1–10. https://doi.org/10.1016/S0140-6736(20)30154-9.

[23] K.C. Santosh, AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data, J. Med. Syst. 44 (5) (2020) 1–5. https://doi.org/10.1007/s10916-020-01562-1.
[24] I. Ghinai, et al., First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA, Lancet 2 (Cdc) (2020) 1–8. https://doi.org/10.1016/S0140-6736(20)30607-3.

[25] M. Jennifer, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China summary of a report of 72,314 cases from the Chinese center for disease control and prevention, JAMA 2019 (2020).

[26] W.Y. Khot, M.Y. Nadkar, The 2019 novel coronavirus outbreak – a global threat, J. Assoc. Physicians India 68 (3) (2020) 67–71.

[27] N. Ramadan, H. Shaib, Review Middle east respiratory syndrome coronavirus (MERS-CoV): A review, Germs 9 (2019) 35–42.

[28] I.J.M. Res, et al., Severe acute respiratory illness surveillance for coronavirus disease, Indian J. Med. Res. 2 (2020). https://doi.org/10.4103/ijmr.IJMR.

[29] A. Mubarak, W. Alturaiki, M.G. Hemida, Middle East respiratory syndrome coronavirus (MERS-CoV): infection, Immunol. Resp. Vaccine Dev. 2019 (Cdc) (2019).

[30] S. Mandal, T. Bhatnagar, N. Arinaminpathy, A. Agarwal, A. Chowdhury, Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: a mathematical model-based approach, Indian J. Med. Res. (2020). https://doi.org/10.4103/ijmr.IJMR.

[31] U. Ali Zia, N. Khan, Predicting diabetes in medical datasets using machine learning techniques, Int. J. Sci. Eng. Res. 8 (5) (2017) 1538–1551.

[32] M. Outcomes D. Wang, B. Hu, C. Hu, Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus—infected pneumonia in Wuhan, China, JAMA 323 (11) (2020) 1061–1069. https://doi.org/10.1001/jama.2020.1585.

[33] C. Huang, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 6736 (20) (2020) 1–10. https://doi.org/10.1016/S0140-6736(20)30183-5.

[34] https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html, web source.

[35] https://www.covid19india.org/, web source.

[36] W. Yue, Z. Wang, H. Chen, A. Payne, X. Liu, Machine learning with applications in breast cancer diagnosis and prognosis, Designs 2 (2) (2018) 13. https://doi.org/10.3390/designs2020013.

[37] J.R. Quinlan, R.L. Rivest, Inferring decision description trees using the minimum length principle, Inf. Comput. 80 (1989) (1989) 227–248. https://doi.org/10.1016/0890-5401(89)90010-2.

[38] C. Panigrahi, J.L. Sarkar, B. Pati, S. Bakshi, E3M: an energy efficient emergency management system using mobile cloud computing, in: IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 2016, 2016, pp. 1–6. https://doi.org/10.1109/ANTS.2016.7947862.

[39] B. Pati, J.L. Sarkar, C.R. Panigrahi, M. Tiwary, ECHSA: an energy-efficient cluster-head selection algorithm in wireless sensor networks, in: International Conference on Mining Intelligence and Knowledge Exploration, 2015, pp. 183–193. https://doi.org/10.1007/978-3-319-26832-3_18.

[40] N. Sneha, T. Gangil, Analysis of diabetes mellitus for early prediction using optimal features selection, J. Big Data 6 (1) (2019). https://doi.org/10.1186/s40537-019-0175-6.