Prediction and Comparative Analysis of Few Most Impacted Countries by Coronavirus

Simran Khanna, Meenal Garg, Seema Aggarwal, and Hardeo Kumar Thakur

Abstract The globe is presently encountering a significant term that has trembled the human's mind, i.e., coronavirus. Carefully examining the unidentified prevalence and danger rate of inhumane life, WHO instantly declared COVID-19 a pandemic on March 11, 2020. The government of all the countries has adopted various policies and measures to control the spread of this disease. Despite many precautions, no. of cases and mortality rate are increasing simultaneously. This analysis targeted to determine the prediction of the COVID-19 cases and deaths among five countries. Especially, it investigates in which country expected maximum and the minimum total number of deaths and number of new cases per million take place in the impending times including cause behind it. Our analysis tends to enforce linear regression to elucidate the founded data to foretell the values. In this context, the "WEKA Tool" has been taken into account which is a renowned machine learning tool. The prediction might be helpful for councilmen and healthcare service providers in suppressing the incremental transmission of this nationwide disease.

Keywords Weka · Owid · COVID-19 · Arff · USA

1 Introduction

COVID-19, this term has shaken human minds and misbalances everyone's life. The outbreak of the novel coronavirus of the twenty-first century affected millions of people. COVID-19 is not only a devastating disease but also a communicative illness, which transfers from one person to another. Because of this infectious disease, the epidemic has spread, and deaths are taking place all over the world. Everything

S. Khanna (✉) · M. Garg · S. Aggarwal
Satyug Darshan Trust, Faridabad 121001, India
e-mail: simrankhanna21jan@gmail.com

S. Aggarwal
e-mail: seema@mru.edu.in

H. K. Thakur
SCSET, Bennett University, Greater Noida, India

in this developed world keenly enjoys science, from the movement of the earth around the sun to the imminent arrival of this deadly coronavirus, all its science. The coronavirus is a zoonotic disease, transmitted between animals and humans. Coronavirus has dispersed its roots throughout the world; countries are severely suffering from immense destruction due to the coronavirus, a direct result of a pseudo-science experiment that typically costs many human lives.

On December 31, 2019, scientists observed the dismissal of a new coronavirus ("SARS-CoV-2") in Wuhan, China, which later give rise to COVID-19. Arrival as well as the gradual spread of the coronavirus have typically led to a rapid worldwide health crisis. Coronavirus caused a severe epidemic that resulted in terrible havoc all over the world. Infectious COVID-19 has irretrievably ruined ordinary life in the world due to its unusual ability to spread rapidly and the inadequacy of effective vaccines. It in common is an infectious disorder that invariably causes rare illnesses, typically ranging from a mild cold to severe respiratory disease. The infectious virus spreads through droplets or nostrils when an infected person coughs or sneezes. Specific symptoms like mild fever, cough, and shortness of breath, profound fatigue, and acute pneumonia may ordinarily occur in the chief sufferers. In key addition, it is typically an asymptomatic disease, which means that some local people without familiar symptoms can also transmit the infectious virus and gently spread it. These rare cases are silent propagators of the infectious virus and are the most difficult to trace. Responsible governments have unanimously adopted several appropriate measures to combat its gradual spread like quarantine, social isolation, and lockdown. Domestic residents were gently urged to wear precise masks and gloves if they ventured outdoors and use sanitizers and cleanse their hands regularly. Local transport has been suspended and non-essential companies, domestic industries, and social services have been typically advised to stay promptly closed until further notice.

Countries around the world have increased the capacity building of laboratory structures, but they until now experience a potential shortage of suitable systems to limit its gradual spread. To prevent this pandemic, we expect a sustainable solution. This pandemic is neither the first nor the last that the world has resolutely faced to be better prepared for the future we need to focus more on the precaution and the solution. To enthusiastically support humane life and stop this devastating contagion, effective vaccines have cautiously emerged as the dominant weapon to encounter this formidable enemy hailed as COVID-19. The whole world has reunited to overcome this problem and fight valiantly against it. Scientists of various countries came out with different vaccines to get rid of the Coronavirus. Covishield and Covaxine have been unanimously approved and adequately prepared by Indian Scientists. Sputnik V is an adenovirus viral vector vaccine for COVID-19 developed by the Russian Ministry of Health. Jansen or Johnson and Johnson COVID-19 vaccine are invented by an "American corporation." Novavax COVID-19 vaccine is invented by an "American biotechnology bureau." Moderna COVID-19 vaccine condensed mRNA-1273 is developed by Moderna (NIAID); and (BARDA).

Even though necessary precautions have been adopted worldwide, the sheer numbers of cases and deaths have fluctuated since the fierce outbreak of the epidemic promptly began. Fortunately, many individuals have recovered without the necessity

of any particular health care, but many need treatment substantially more senior citizens, together with fundamental medication issues say, Cardiac infarction, Brain Tumor, Cholesterol, Diabetes, etc. Coronavirus produces a vast impact on the wealth of the nation as it decelerates the economy and gives rise to extensive damage directed toward the development of the nation. The frugality is deteriorating because the worldwide logistics network is contracting together with many individuals also addressing many additional hindrances. Along with adverse effects, there were also some overwhelming results around the world. As the world was losing out because of the pandemic, our mother nature gained something from it. The environmentally damaging particulate matter was eliminated and most importantly, it was discovered that the largest ozone hole was shrinking, and pollution was decreasing during the pandemic.

1.1 State of the Art

Diverse research papers and studies have been conducted since the inception of the COVID-19. Innumerable studies have been done on analysis and prediction, control measures, prevention and treatment measures, regression analysis for comparison, and many more distinctive features. Like a study named on the prediction of isolation, release, and decease states for COVID-19 patients: A case study in South Korea was performed by Alafif et al. in South Korea. Their prediction will support the nation to isolate areas of infected patients, which will also support the country to carefully limit the spread of COVID-19 [2]. One more study named Analysis of the Worldwide Coronavirus (COVID-19) Pandemic Trend; A Modeling Study to Predict Its Spread was done by Qasim et al. in which they analyzed the most affected countries and explained how countries can improve their health resources [3]. Another successful model was carried out named Prediction of COVID-19 pandemic for the topmost 15 affected countries using "ARIMA Model" by Singh et al. in which they predict what will be the confirmed cases and mortality rate among these countries in the coming times [4]. Another modeling study was done precisely named data-based analysis, modeling, and forecasting of the COVID-19 outbreak by Anastassopoulou et al. in which they predict the count of confirmed cases and death ratios in China [5]. Further, one more study has been carried out on regression analysis of COVID-19 using machine learning algorithms by Gambhir et al. in which they adequately explain the use of machine learning tools in analysis and also analyze the COVID-19 impact in India [6]. The research was carried out on key predictions of diverse countries, predominantly focused on South Asia. This typical analysis claimed the self-quarantine, proper hygiene and lockdown can limit the spread but cannot obliterate the rare disease.

After the study of these papers, we analyzed that all these published studies acquire their significance and limitations, then we concluded that all these studies are different from each other as we found that numerous research with various parameters and relations has already has been done based on our paper but still this research has its

individuality, and it's completely different from other research; therefore, prediction for such countries might be helpful for the policymakers and healthcare service providers in curbing the gradual spread of this dreaded disease.

This analysis has aimed at carefully tracing a key trend related to the total number of death counts per million and the total number of cases per million expected at the diverse countries of the COVID-19, we have used an amazing Machine learning tool, i.e., WEKA which aided us to obtain the result conveniently. Mainly, this research focuses on two key aspects:

Regression analysis and forecasting and precisely determine the prediction of COVID-19 cases and mortality rate per million among the major five affected countries (India, USA, France, Italy, and Russia).

At that time, we tend toward performed comparative analysis in which we rigorously compared that among the above five countries, which country can have expected maximum and a minimum number of "total deaths per million" and "new cases per million."

2   Methodology

The process of collecting information, as per the researcher is addressed as data, only indicates the beginning of the research process. In today's attainable scenario, the foremost sensational topic in common is COVID-19 that has trembled the whole world. The globe generally suffers several vast and terrifying difficulties; however, the coronavirus is threatening among all of them, so in this study, we'll eagerly discuss the problems related to counting on COVID-19 cases. Through the process of data mining, we analyze founded data. Data mining accurately represents the process of going through data to information, information to data that can provide us the useful predictions.

Based on statement problem demand predictive analysis, so we tend to enforce linear regression to elucidate the matter and so a comparative analysis has been done on that, supported some of its key parameters, we came to understand that within the approaching days, at which country expected a total number of deaths and new cases per million will increase and reduces accordingly. For this method to be done, we found ingenious machine learning tool "Weka" which is an open-source package tool, and it's developed at the University of Waikato; Weka explains the elementary principles of many customary algorithms. As determined by knowledge points are based on many key attributes of five elite countries and every country has 453 instances. Therefore, we tend to are inundated with our determined data which is why Weka helps to turn up the result concisely. The required data conspicuously utilized for the present study are directly out there in the GitHub repository at this official site.

The data contain info on a routine basis in over 230 countries worldwide. OWID is a website that ponders on global threats, such as poverty, certain diseases, severe famine, potential climate change, potential war, survival risk, and inequality. OWID

is a global modification knowledge laboratory, a non-profit organization registered in England and Wales; it truly owns and adequately maintains the website and also the data tools and was based by Max Roser. The trustworthy data comprised in OWID help to makes progress against the world's vastest problems. In this study, we've collected case data reports for the five elite countries. The data set contains relevant parameters like Continent, Location, Date, Latitude, Longitude, total cases, new cases, total deaths, new deaths, total cases per million, and total deaths per million of each day. However, we have selected some parameters to examine the data from the overall dataset for analysis of the COVID-19 outbreak.

The data used in the current study are based on the numerical values which contain 453 instances for each country so, as the popularity curve gradually increases; we specifically looked at the data of five countries (India, the USA, Russia, Italy, and France). In these countries, from February 1, 2020, to April 28, 2021, we processed "new cases," "total cases," "total new cases per million," "total deaths," "total deaths per million." As data can be imported from documents in numerous compositions: ".arff," ".csv," ".c4.5."

2.1 Steps to Convert .xls to .arff

1. Convert Excel files with collected data to.csv-format (Values are separated by commas); this is a structured table format.
2. Use a.csv file to change the format of the key algorithm to .arff.

Next, enter the data in Weka.

Figure 1 indicates uploaded data of India. In this data, it is visible that the relation is India new; total instances are 453, total attributes are nine, and missing value is 0%, etc.

This study applies regression methods by actually using real-time global COVID-19 data to determine the probability of successful COVID-19 cases within 5 days based on previous new cases per million and the total number of deaths per million in key countries. Linear regression is fragile and well-founded mathematical technique.

3. In statistics, linear regression is a linear approach to look at the connection between dependent and freelance variables.
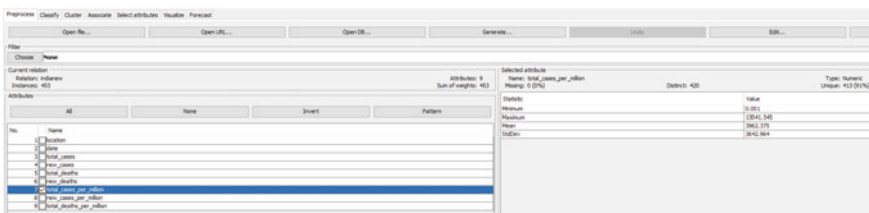

Fig. 1 Data of India

4. In prediction analysis: regression toward the mean is that the most widely used modeling technique for predicting the future.

2.2 The Specific Formula Used as Shown in Eq. (1)

$$X = x^1 + \alpha x^2 + \beta \tag{1}$$

It is represented by an Eq. 1; where x^1 and x^2 are the intercept, α is the slope of the line, β is the error term, and X is the predictive value. This equation can be used to accurately predict the possible value of the target variable based on given predictor variables.

2.3 Weka Using Time Series Forecasting Package

1. Reload all files "indianew.arff," "francenew.arff," "russianew.arff," "italynew.arff," and "USAnew.arff" individually.
2. Go to the Forecast panel → advanced configuration option → linear regression.
3. Go to basic configuration → According to Data Frequency-here < Detect automatically comprised of given options according to which data is predicted > the year, quarter, month, week, day, and hour data are displayed.

3 Evaluation Process

Estimate the number of steps required to generate an accurate forecast. Forecast 5 units ahead, to predict the future after 5 active days (the dotted line, the circle mark on the chart provides a general forecast of the future). Then estimate and select the total number of deaths per million. And activate the evaluation, and then we can accurately get the prediction required for the next 5 days, as well as the stepwise absolute error and root, mean square value. The same procedure usually applies to new cases per million. Look at the prediction of the training set. Each country has two graphs where the dependent variable is "total death per million" or "new cases per million" and the independent variable is "time."

4 Results and Analysis

Weka's time series forecasting package determines the evaluation on 5-step-ahead "number of time units to forecast" [7] classify the linear regression and also sets out "perform evaluation" for mean absolute error and root mean squared error step-wise.

Figure 2 determines the graph of the data of France "new cases per million" where the training set is indicated by a square and predicted values, or the test set is indicated by a circle. Here, in the test set, the value is decreasing through which we can determine that in the future there are chances that the "number of new cases in France" might be decreased.

Figure 3 shows the graph of France's total deaths rate per million, and the test set indicates a hike in the prediction of the total number of deaths in France in the upcoming days.

Figure 4 indicates that India's new cases per million and the outcome appears to increase in the future.

Figure 5 is India's total death rate per million and the test set indicates that India's mortality rate is least than the other countries, and the deaths rate may also fall in the future.

Figure 6 shows new cases per million in Italy which may decrease gradually.
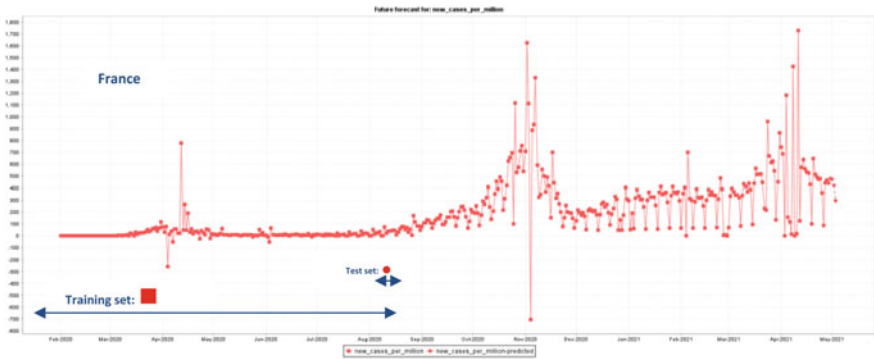
Fig. 2 France, new cases per million

Fig. 3 France, total deaths per million
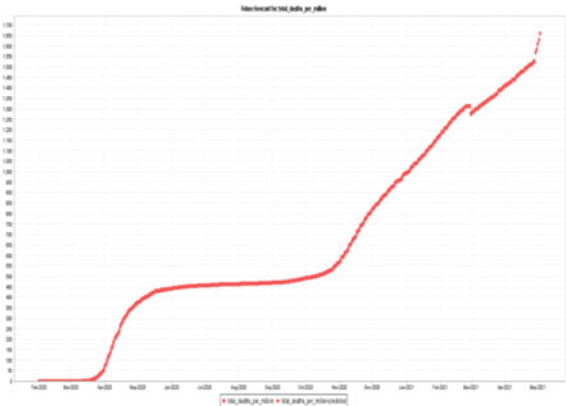
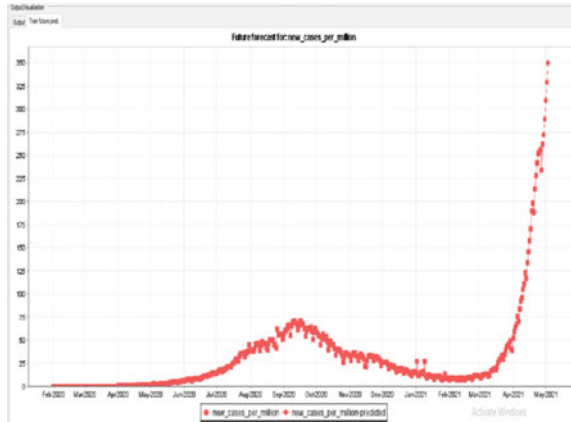**Fig. 4** India, new cases per million

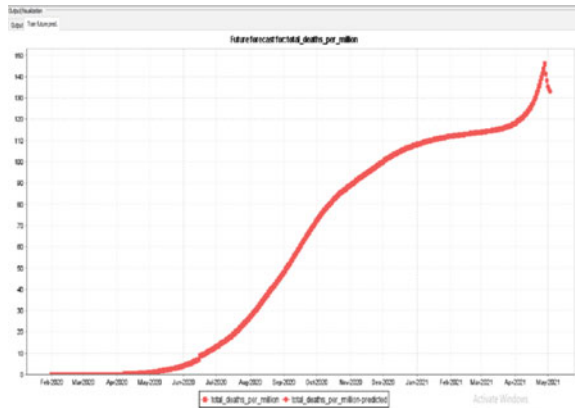
Fig. 5 India, total deaths per million

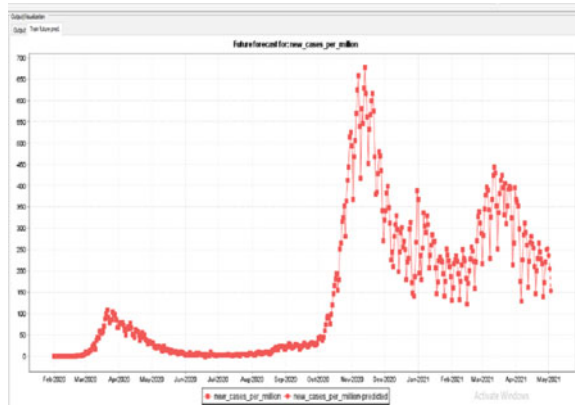**Fig. 6** Italy, new cases per million
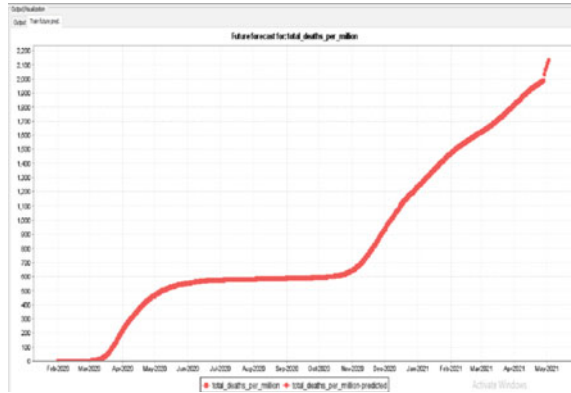
Fig. 7 Italy, total deaths per million

Figure 7 shows total deaths per million in Italy which is increasing at a faster pace in the future.

Figure 8 shows Russia new per million which although in a moderate state but may decrease in the future. Despite this, the rate of new cases is very low compared to other countries.

Figure 9 shows total deaths per million in Russia which may increase gradually.

Figure 10 suggests that new cases per million in the USA may decline sharply in the future.

Figure 11 indicates that the total death rate per million in the USA could increase by a good amount.
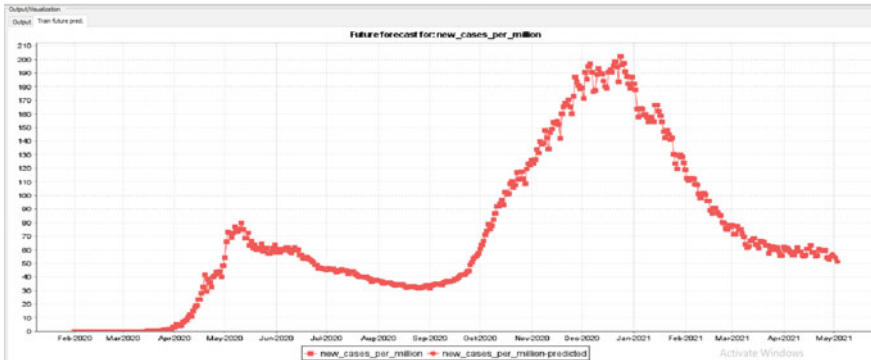

Fig. 8 Russia, new cases per million
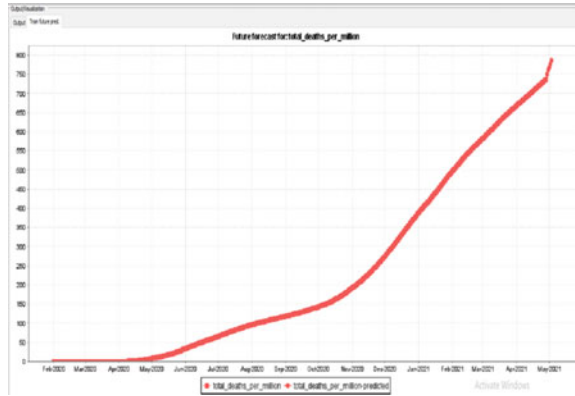
Fig. 9 Russia, total deaths per million
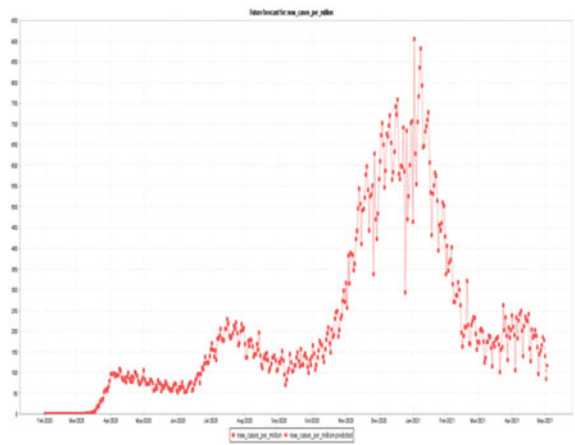

Fig. 10 USA, new cases per million

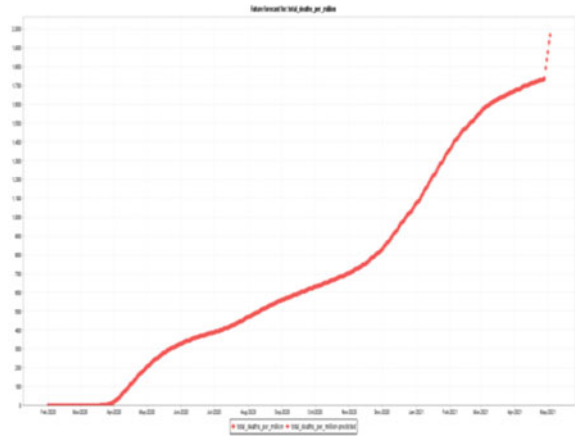**Fig. 11** USA, total deaths per million

Table 1 Comparative analysis

No. of max/min cases	Cases/deaths	
	Total deaths per million	New cases per million
Minimum	India	Russia
Maximum	Italy	France

5 Conclusion

From the investigation and the previous enumeration, in conclusion, the predicted minimum "total number of deaths per million" takes place in India, and the expected maximum number of "total deaths per million" takes place in Italy as given in Table 1.

Whereas a predicted minimum "total number of new cases per million" takes place in Russia, and the expected maximum number of "total new cases per million" takes place in France as given in Table 1. It is due to the fact that the governments of various countries have taken various measures, i.e., lockdown in the most effective areas, increasing tests and increasing health infrastructure, etc., to prevent the COVID cases. Therefore, the predicted values for deaths and new COVID cases are more in comparison with actual count in deaths and new COVID cases.

References

1. The data used for the current study is freely available in the GitHub repository at the site given. https://covid.ourworldindata.org/data/owid-covid-data.xlsx
2. Alafif T, Alotaibi R, Albassam A, Almudhayyani A (2021) On the prediction of isolation, release, and decease states for COVID-19 patients: a case study in South Korea. IS Trans, 5 Jan 2021 (In Press)
3. Qasim M, Ahmad W, Yoshida M, Gould M (2020) Analysis of the worldwide corona virus (COVID-19) pandemic trend; a modelling study to predict its spread
4. Singh RK, Rani M, Shah R et al. (2020) Prediction of the COVID-19 pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (ARIMA) model, vol 6, no 2, 13 May 2020
5. Anastassopoulou C, Russo L et al. (2020) Data-based analysis, modeling and forecasting of the COVID-19 outbreak, 31 Mar 2020
6. Gambhir E, Jain R, Gupta A et al. (2020) Regression analysis of COVID-19 using machine learning algorithms, IEEE, 10–12 Sep 2020
7. The algorithms used for the current study https://drive.google.com/file/d/1JFNC3jzU9TP_n-ExVM2HX91HfFOG32zu/view?usp=sharing