ORIGINAL ARTICLE



A data analytics approach for COVID-19 spread and end prediction (with a case study in Iran)

Arman Behnam¹ · Roohollah Jahanmahin¹

Received: 7 October 2020 / Accepted: 2 January 2021 / Published online: 30 January 2021 © The Author(s), under exclusive licence to Springer Nature Switzerland AG part of Springer Nature 2021

Abstract

World is now experiencing the new pandemic caused by COVID-19 virus and all countries are affected by this disease specially Iran. From the beginning of the outbreak until April 30, 2020, over 90,000 confirmed cases of COVID-19 have been reported in Iran. Due to socio-economic problems of this disease, it is required to predict the trend of the outbreak and propose a beneficial method to find out the correct trend. In this paper, we compiled a dataset including the number of confirmed cases, the daily number of death cases and the number of recovered cases. Furthermore, by combining case number variables like behavior and policies that are changing over time and machine-learning (ML) algorithms such as logistic function using inflection point, we created new rates such as weekly death rate, life rate and new approaches to mortality rate and recovery rate. Gaussian functions show superior performance which is helpful for government to improve its awareness about important factors that have significant impacts on future trends of this virus.

Keywords COVID-19 · Trend prediction · Machine learning · Data analytics

Introduction

On 31 December 2019, the World Health Organization [WHO] was informed of a cluster of cases of pneumonia which related to coronavirus family. At first, a novel coronavirus was identified in Wuhan, Hubei Province, China causing severe respiratory disease including pneumonia. It was originally named Novel Coronavirus and this virus causing the infection has been named by The World Health Organization (WHO)—severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (The faculty of Intensive Care Medicine, 2019). SARS-CoV-2 is spreading between people globally and WHO (2019) has announced that on August 16, 2020, there have been over 21 million confirmed cases of COVID-19, including over 761,000 deaths. The COVID-19 pandemic is far more than a health crisis: it is adversely affecting societies and economies of countries across the globe such as increasing the unemployment rate and decreasing production. According to the results of a survey study,

The World is suffering from the new pandemic caused by the COVID-19 virus affecting most countries including Iran. In mid-February, Iran became the second-largest country with highest prevalence of coronavirus in the world after China (The New York Times 2020). The first case of COVID-19 was reported in Qom on February 19, 2020. Because of the high prevalence of the virus, almost all cities in Iran became infected. According to World Health Organization (WHO 2020) as of August 16, 2020, the number of confirmed cases of COVID-19 in Iran was



conducted by Imam Sadiq University, in relation to the beginning of the outbreak, 65% of the economic enterprises declared that they cut back their production, while only 4% stated that their activities increased (Iranian Students' News Agency [ISNA], 2020). According to an another survey study by ISPA (Iranian Student Opinion Polling Center) in Tehran, about 70% of the respondents said that their household incomes had declined, due to the unfavorable effects of the coronavirus on the economy (Iranian Students' News Agency [ISNA], 2020). Without urgent socio-economic responses, global suffering will escalate, jeopardizing lives and livelihoods for years to come. Development trajectories in the long-term will be affected by the choices countries make now and the support they receive." (United Nations Development program, 2020).

Arman Behnam arman_behnam@ind.iust.ac.ir; r_jahanmahin@ind.iust.ac.ir

Industrial Engineering Department, Iran University of Science and Technology (IUST), Narmak, 16846-13114 Tehran, Iran

341,070 and the number of deaths was 19,492. Tehran, the capital of this country, has the highest prevalence among Iranian cities (World meter 2020).

According to recent studies, a number of factors may influence the spread of COVID-19 such as climatology parameters (Ahmadi et al. 2020), tobacco smoking (Ahmed et al. 2020), and air pollution (Wu et al. 2020). For example, Ahmadi et al. (2020) showed that areas with a little amount of wind, humidity, and solar radiation support the virus's survival. Ahmed et al. (2020) indicated that smoking tobacco can be a possible mode of transmission for the virus for both active and passive smokers. The authors argued that COVID-19 transmits through salivary droplets and causes severe lung pneumonia, thus, tobacco smokers are at high risk of infection Wu et al. (2020) noted that a small increase in long-term exposure to air pollution leads to a large increase in the COVID-19 death rate.

To the best of our knowledge, no effort has been made to model this prevalence by ML methods. In this study, we aim on predicting death, recovered, and confirmed cases in different periods of time and trends of analytical parts of healthcare with statistics and machine-learning tools. Our data format is time-series since our data are gathered by the number of cases per day. Our study area is categorized into two parts due to variables: (1) first, we desire to predict the number of cases with machine-learning algorithms as much as accurate. (2) Second, using variables to define helpful rates and use them to control conditions by statistics approximation and data mining tools and also check whether the predictions will become quite real is intended. We implemented substantial tools used in the trend calculation and dynamic rates are used to analyze encounters obstacles in fields of explaining results more accurately and meaningfully. Findings of this study can help decision makers to control the pandemic using our analysis.

Materials and methods

Study area

Before predicting trends of this disease throughout the world, we need to get acquainted with the current pandemic situation. Therefore, in Fig. 1, a heat map of most affected countries is plotted according to the number of confirmed cases.

Since Iran has a large portion of COVID-19's spread compared to other countries, a comparison is undertaken to realize Iran's share related to the whole world. Iran's number of confirmed cases is 215,096 out of 9,344,686 global cases. It is equal to 2.30% of all confirmed cases.

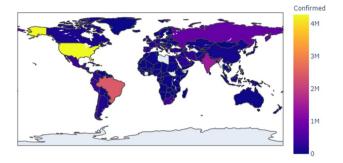


Fig. 1 The most affected areas by geography

Data collection and preparation

Variables in our research include number of newly found cases, new death cases, and newly found recovered cases in Iran. All information was collected and categorized from reputable sources such as WHO (World Health Organization). These variables are chosen to utilize in our prediction methods due to their numerical nature.

The high prevalence rate of COVID-19 and the need for estimated calculations necessitate collecting the required data sets from reliable sources including WHO and worldometer. Research data including observation data are obtained from a collection of samples' reports in three parts (i.e. death, confirmed, and recovered). This countrywide daily information (database mendeley, 2020) is confirmed by the WHO. It should be noted that the relevant data was collected between February 19th and June 25th, 2020.

Machine learning

ML methods help us perceive trends and variables behaviors using a wide range of methods which is almost better than human performance. Previous studies used some techniques such as statistical techniques, machine learning, and data mining followed by a discussion on segmentation strategies including Distributed, Parallel Processing, and Statistical Inference and classification of data mining applications in Operation Management studies into two categories including forecasting and risk analysis in operations management. To calculate the approximate growth distribution and the delayed disease transmission, methods of Log-normal distributions, Weibull, and Gamma were utilized (Li et al. 2020). The laboratory, hospital, and epidemic characteristics of patients with coronavirus were found out and used to identify patients in and out of the ICU using the screening method of patients with pneumonia in Wuhan (Huang et al. 2020). The type of COVID-19 was determined using four virus detection methods by examining three patients with pneumonia, centrifugation, electron microscopy, and viral



genome sequencing, leading to the discovery of a new coronavirus through seafood (Zhu et al. 2020). Drug screening based on the 2019-nCov_main_protease model structure was studied (Zhang et al. 2020). A preliminary screening model to differentiate COVID-19 pneumonia from influenza viral pneumonia and healthy cases with pulmonary CT images using deep learning techniques were developed (Xiaowei et al. 2020). A neural network algorithm with the least prediction error called PNN was developed (James et al. 2020). VHP (virus-host prediction) was introduced to predict the potential host of viruses using deep learning algorithms (Guo et al. 2020). 2019-nCoV was classified using MLDSP and MLDSP-GUI. Also, Machine learning and Digital Signal Processing (DSP) were used as alignment methods for genome analysis (Randhawa et al. 2016).

The following workflow is extracted to implement this research as depicted in Fig. 2. As the figure shows, we aim to predict the trend of COVID-19 using novel machine-learning techniques for Iran's case. For this purpose, after collecting the dataset which carried out on the global data and the data related to Iran, we determined and investigated death, recovered and confirmed rates and after that we predicted these rates with predicting tools including cumulative distribution, SVM, linear regression, Gaussian and prophet. Finally, we have predicted the trend of virus infection in Iran in short-term, mid-term, and long-term. This analysis is conducted on the case of Iran to predict the spread and end of this disease.

Python software is used to implement operations such as data cleaning, data manipulation, and machine-learning

algorithms. Some operations such as cleaning data, joining, summarizing, defining functions, building, and changing data frames are applied. "Numpy", "Pandas", "datetime", "sklearn", "math", "seaborn", "matplotlib", "statsmodels" and "scipy" packages are utilized in this analysis.

Descriptive analytics

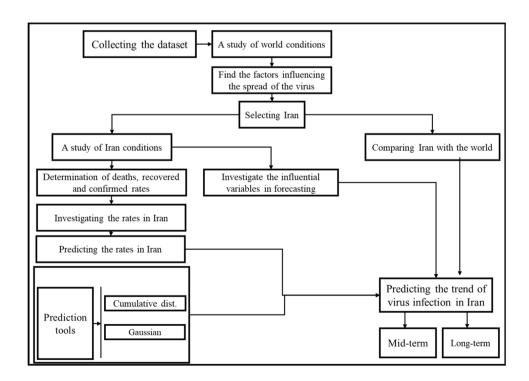
Data analytics tools are categorized in three classes: descriptive analytics, predictive analytics, and prescriptive analytics (Hazen et al. 2016). Descriptive analytics, extracts information from raw data appeared in business and management reports regarding sales, customers and operations. It helps organizations in grasping the reasons of the events happened in the past. Combined with simple graphical analysis, they shape the quantitative data analysis basis (Choi et al. 2018).

In this section, descriptive analysis is done to review the position. Choosing the death rates and analyzing them more accurately is another way to survey them as the main factor and make new rates. Therefore, the mortality rate which is calculated as Death/Recovered cases, is smoothing, as well as Death/Confirmed cases. This is mainly because of a drastic increase in the number of deaths at first and appropriate reactions to them in countries like Iran.

Prescriptive analytics

Prescriptive analytics phase provides studies with adaptive, automated, time-dependent, and optimal decisions. In general, prescriptive analytics is predictive analytics that

Fig. 2 The proposed prediction framework





prescribes one or more courses of action and shows the likelihood of outcome/influence of each action. Prescriptive analytics is purely built on the "what-if" scenarios. The main tools of prescriptive analytics are optimization, simulation, and evaluation methods. Simply, it provides advice based on predictions (Choi et al. 2018).

Based on the literature, analysis in this section provides suggestions for future behaviors and policies with attention to rates instead of cases. Some known rates such as recovery and death are limited in usage. Thus, two new rates are considered and defined here.

Rates are the best indexes to predict the future and end of disease, some useful rates are as follows:

- The "Weekly recovered rate" is the ratio of the average weekly number of people have cured over the average weekly number of people infected with COVID-19.
- The "Weekly death rate" is the ratio of the average weekly number of dead people over the average weekly number of people have infected with COVID-19.

rate or sharpness of the curve. The logistic function is just a solution for the following first-order, non-linear ordinary differential equation where f(0) = 0.5. The solutions stability is explored. Since it may not be conducive to predict approved cases, growth criteria have been used.

• The Analysis Growth Factor, Growth Ratio, Growth Rate, 2nd Derivative are growth metrics being used to gain insight into which countries may have already hit their inflection points. For example, whether Iran's growth factor has stabilized around 1.0, which is a sign this country has reached its inflection point. Thus, curve fitting is used to fit a logistic curve to the number of confirmed cases, helping predict if Iran has hit its inflection point, and when it will reach a possible maximum number of confirmed cases.

The growth factor just shows the curvature of the data. Whether the cases are growing at an accelerating or decelerating rate is quite important. The derivative test is used to test concavity and find saddle points. The inflection point is where the curve changes concavity, as follows:

Growth factor on day
$$N = \frac{\text{Confirmed cases on day } N - \text{Confirmed cases on day } (N-1)}{\text{Confirmed cases on day } (N-1) - \text{Confirmed cases on day } (N-2)}$$

 The "Life rate" increase or decrease is the difference between the weekly recovered rate and the weekly death rate. It shows an increase when it is positive, while it indicates a decrease when it is negative.

The infectious disease spread is better to be modeled using a logistic curve rather than an exponential curve. The growth starts exponentially, but slows down after the point called the "inflection point" which is in the midpoint of the spread. The number of confirmed cases using a logistic curve is modeled. At first, three things are needed to get defined: (1) an equation for such a curve, (2) a differential equation for which this curve is a solution, (3) the graph of the curve are.

 Logistic Function A logistic function or "logistic curve" is an equation of the form where: ×0=the inflection point, N=the curve's maximum value, and k=growth

Growth ratio on day
$$N = \frac{\text{Confirmed cases on day } N}{\text{Confirmed cases on day } (N-1)}$$
.

Predictive analytics

Predictive analytics phase incorporates the descriptive analytics output as well as some ML algorithms and simulation techniques to build accurate models that predict the future trends. Predictive analytics assists studies in identifying future opportunities and likely risks by distinguishing specific patterns over the historical data. Some outstanding techniques which are utilized in this phase are data mining (DM), text/web/media mining, and forecasting approaches such as regression, SVM, K-Nearest neighbors and Random forest (Choi et al. 2018).

In the following, a question arises regarding where and when COVID-19 will be eradicated.

Cumulative cases prediction As predicting the future of outbreak is something nearly impossible, a great approach toward several each case including confirmed, death, and recovered, is enforced through using density distribution for fitting cumulative amounts of each category. Due to the cumulative nature



of these numbers, the cumulative distributions under consideration should not follow the normal distribution.

Gaussian function The Gaussian distribution fitting of the data is one great method for predicting the trend when variables and functions are defined. To implement it, the function should be built. In one dimension, the Gaussian function is the probability density function of the normal distribution, sometimes also called the frequency curve (in below).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(X-\mu)^2/(2\sigma^2)},$$

The full width at half maximum (FWHM) for a Gaussian is found by finding the half-maximum points x_0 . The constant scaling factor can be ignored, so we must solve

$$e^{-(X_0 - \mu)^2/(2\sigma^2)} = \frac{1}{2}f(x_{\text{max}}).$$

In two dimensions, the circular Gaussian function is the distribution function for uncorrelated variates X and Y having a bivariate normal distribution and equal standard deviation $\sigma = \sigma_x = \sigma_y$. The corresponding elliptical Gaussian function:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-[((x-\mu_x)^2/(2\sigma_x^2)) + ((y-\mu_y)^2/(2\sigma_y^2))]}.$$

In conclusion, we will use the following function in our calculations:

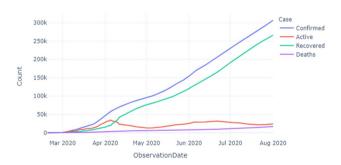


Fig. 3 Number of confirmed, active, recovered and dead cases

$$f(x,l,k,x_0) = \frac{l}{1+e^{-2k(x-x_0)}} + 1.$$

This function has variables with initial values including: L (maximum number of confirmed cases) = (80,000 taken) from figures above), k (growth rate) = 0.2 (approximated value from most of the countries), $\times 0$ (the day of the inflexion) = 50 approximated.

Results

Descriptive analytics

Based on our dataset, active cases are derived. Figure 3 reveals the number of confirmed, recovered, and death cases per day. A tremendous rise in the case number is obvious though freshly the active cases number is decreasing. Death cases trend has a gentle slope.

A case variety review in this country is executed in the following to assess the extent of the spread. However, one way to go through if this trend is amplified or not is related to instances ratio assessing. As shown in Fig. 4, numbers of all three categories are rising daily and so maybe the peak of disease is near (X-axis is date and Y-axis is number of cases). Besides, recovered cases are getting raised sharply. Rate of recovery growth being more than death growing rate is a hopeful point here.

Considering a significant factor is the main point about the COVID-19 prevalence in the world is that has been a hot topic everywhere; age is considered in comparison with recovered cases and death cases as shown in Fig. 5. It is obvious most of the confirmed cases that are not recovered are about 35–65, most of the recovered people are about 30–50 and some of them are kids. Thus, age is an important factor in recovering procedures based on average difference. In death cases, it is found out that age has the most effect on being dead, and among the people that have died, it seems to be mainly older people, above the age 35, between 55 and 85.

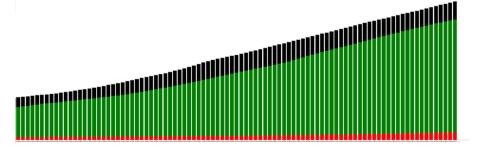




Fig. 5 Age in relation to recovery and death

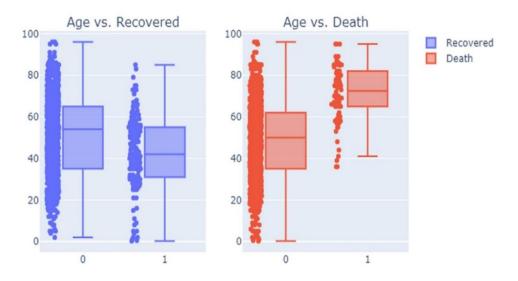
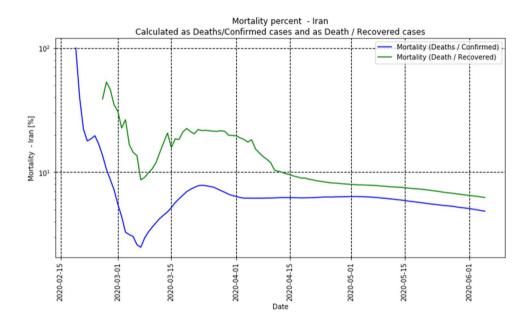


Fig. 6 Mortality percent for Iran

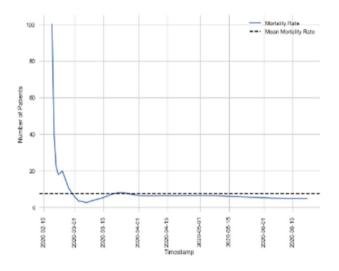


Some relevant information due to these graphs are all policy that Iranian authorities have implemented have enabled them to control COVID-19 in some time intervals. The growth or decline of COVID-19 disease will depend on the effectiveness of the policy put in place to combat this disease. Therefore, the curve does show us the correlation that exists.

As shown in Fig. 6, a strange pattern is observed starting from March 7th, which seems to indicate a certain correlation between Confirmed and Recovered cases. Also, an important issue is mortality rate oscillation which is obvious in the disease onset.

In some other approaches, the global mortality and recovery rates should be considered in comparison with





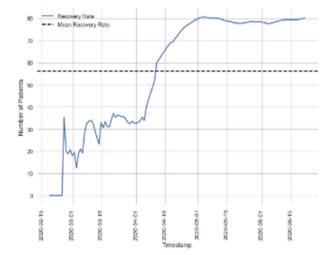
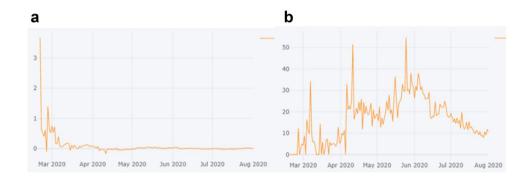


Fig. 7 Mortality and recovery rates in Iran versus the mean global rate

Fig. 8 Rates in Iran per day



Iran to understand the situation more realistically. Iran's mortality rate has come under the global mean which is represented on the dashed line in Fig. 7. It resembles to continue under this line. It is better to keep this rate under the global mean. Also, the recovery rate in Iran had some problems till March 15th, and after that had a lot of variation around the mean rate. Unlike the past, from April 6th, it started to increase exponentially causing many hopes for combating the disease.

In a conclusion, Iran's case has a growing number of confirmed and death cases and some solutions should be proposed to take the situation under control. These are a good sign of advancing therapeutic goals and healing people, according to daily recovered cases number. A case variety review in this country is executed in the following to assess the extent of the spread. However, one way to go through if this trend is amplified or not is related to instances ratio assessing.

Prescriptive analytics

Computing the active confirmed cases growth rate in Iran is as follows:

Active confirmed growth rate on day $N = \frac{\text{Active confirmed cases on day } N - \text{Active confirmed cases on day } (N-1)}{\text{Active confirmed cases on day } (N-1)}$



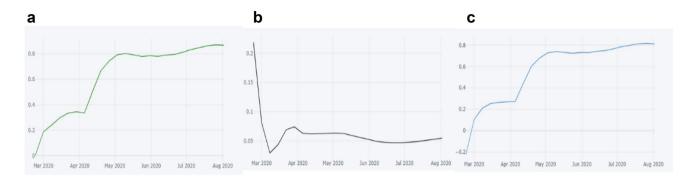


Fig. 9 Weekly rates of Iran

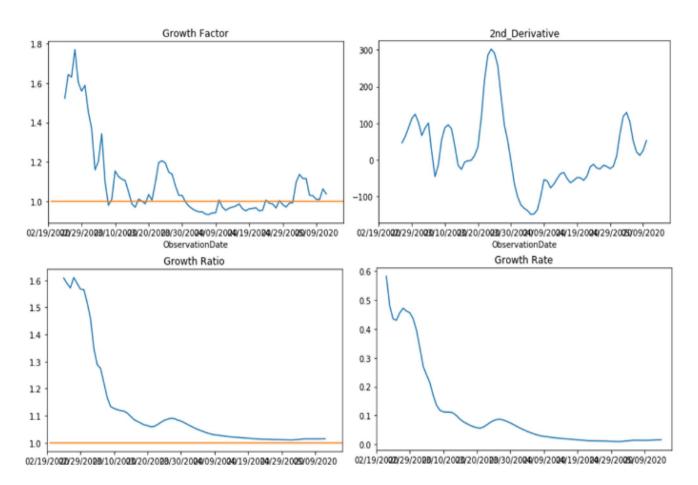


Fig. 10 Mortality and recovery rate in Iran versus mean global rate

Figure 8a shows that the infection rate is getting under control overtime since the usage of appropriate policies to qualify the disease growth. After April 6th, this growth rate is about 0 and keeps constant, which is a good sign.

Another vital rate is recovered-death rate (not being an active case) which does not show whether the patient is exactly recovered or dead, indeed it helps to determine these people are not able to transfer the virus to others, as follows:



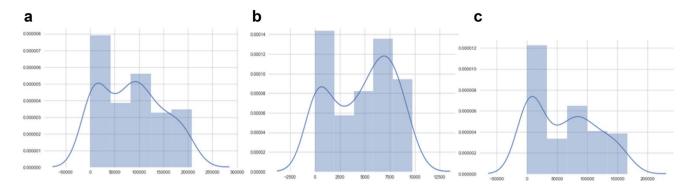


Fig. 11 Density Distribution Plot for Iran Cases

Recovered – Deaths rate on day $N = \frac{\text{Recovered cases on day } N - \text{Recovered cases on day } (N-1)}{\text{Death cases on day } N - \text{Death cases on day } (N-1) + 1}$

A large variation of this rate is obvious in Fig. 6, which shows recovering methods, isolation policies, and available hospital facilities are always in conflict with each other, in terms of influence. Therefore, there is a correlation between the growth rate and types of mitigation plans. The eradication of COVID-19 in a country depends on its policy for combating the disease and how this policy is implemented immediately. Variation in the number of recovered and death cases is a very odd pattern. There have been two periods of little oscillation and after that a great rise that it will happen later too, which is shown in Fig. 8b.

Figure 9a, including two identical patterns with April 6th as a junction point, displays that the weekly recovery rate is rising over the time because of certainty in recovering methods in each of two intervals. Recently, amazing progress is obvious in recovery process about 0.8. From a psychological point of view, the weekly death rate is related directly to isolation and remote working condition. According to Fig. 9b, Iran's weekly death rate was unpredictable until March 29th. But it will be accurately continuing at 0.07 and below it in the future. Life rate is the best rate that helps us to control the disease. Figure 9c shows that this

rate looks like weekly recovered rate. Besides, factors with major impact can change the trend of this variable, telling us that situation takes time to improve in terms of improving health conditions.

Iran's confirmed curve and recovered curve resemble a logistic curve, represented in Fig. 10. A dramatic decrease in Iran's growth ratio and growth factor, until it stabilizes around 1.0 for growth ratio and comes under 0.1 for growth rate.

Predictive analytics

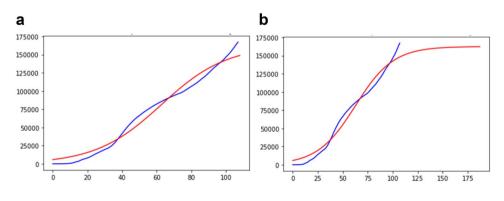
Cumulative cases prediction

With notice to Fig. 11a, we forecast that it would be 200,000 confirmed cases at maximum and worst-case scenario. Note that isolation policies will change conditions through time.

As shown in Fig. 11b, death cases distribution is skewed by some anomalies that occurred. It predicts the zenith of death would be 10,000 cases cumulated with the least possibility. Most of the cumulative numbers are about two intervals: 0–2000, and 6000–8000.

Most of the recovered cases are between 0 and 35,000. Since in comparison with scale in confirmed cases, a greater

Fig. 12 Number of cases (blue) and prediction (red) for Iran





amount of cumulative high-numbered recovered cases is needed (about 0–50,000), represented in Fig. 11c. In the greatest amount, it could be around 150,000 recovered cases in Iran. (Estimated by the distribution of data).

Gaussian functions

Mid-time prediction Forecasting in this area needs building time intervals and analyzing each of them. In the mid-time area, our predictor has worked well to fit the real data and it will be useful to predict either near or far future, which is presented in Fig. 12a.

Variables calculation results at the 108th day are: Predicted L: 162,442, Predicted k: 0.052598, and Predicted $\times 0$: 62.

Long-time prediction Predicting long-time periods, fatalities and deaths caused by the disease in Iran is possible by comparing actual values with logistic fitted values. It needs to compare actual values with logistic fitted values. As shown in Fig. 12b, logistic function helps us predict the peak of the outbreak peak 150 days after the onset of the outbreak. The period of decline will be about this period as well, which means that the end of the outbreak will be about 290–320 days from the start time (meaning November 2020).

Discussion

This study focused on predicting the trend of COVID-19 prevalence and find the best approximate fitting pattern to predict the peak and end in different periods of time including short-term, mid-term and, long-term, and also defining and analyzing mortality rate, recovery rate, death rate, weekly rates, life rate, and infection rate. Descriptive and prescriptive analytics is done on these rates and then predictive analytics was performed with applying machine-learning algorithms. In our results, Gaussian functions are the best method to predict the outbreak since it has enough power to estimate its curve, peak and end. These analytics are needed to find out how effective were policies for prohibiting COVID-19 spread by the government including isolation or closing markets have been made to diminish commuting and traffic have been chosen.

This work is inconsistent with other studies for two reasons; first, we made many comparisons between ML methods which best of them performed for revealing outbreak peak, end and best one was utilized. Second, using results together with rates, help us understand reality alongside analyses to perceive situations accurately and how much can we trust predictions because great or poor rates. Other works like (Fanelli and Piazza 2020; Anastassopoulou et al.

2020), used algorithms like simple susceptible- infected-recovered-deaths (SIRD), and scenario-based analysis may not be the best view for forecasting situations just because of time-series datatype. Hence, we have to do more to justify our predictions. In the descriptive phase, this research h is trying to survey recovered deaths, active confirmed growth rate, and define new weekly rates like the life rate. Analysis of mortality and recovery rates are discussed and the peak of disease is determined. In the predictive analytics phase, not only the cumulative distribution of cases is discussed to predict the peak and end of the disease, but also some functions are defined to resemble the case numbers curve.

There are a lot of drawbacks for predicting the trend of COVID-19 disease and analyzing countries' conditions to explain the situation exactly such as lack of information about social and political measurements and reactions to illness spread, state categorized data is not available for the public completely, many algorithms do not work in this restricted dataset due to its time-series nature.

For future research, some environmental rates could be defined and many other ML algorithms may have higher accuracy and better output in forecast procedures. Some parameters including Sociological, psychological, and climatic parameters will help model this prevalence trend better. Also with a larger dataset, predictions will have higher accuracy on the dataset.

All Governments and health decision-making body organizations can rely on rate analysis along with working on rates to keep them at an appropriate level and zoom on weak points discussed in the article. Additionally, our predictions based on ML algorithms can be used to plan for healthcare programs and also take advantage of this situation to improve conditions.

Conclusion

Weekly recovery rate is getting better over time and it will continue its way not only because of certainty in recovering methods but also people will deal with isolation situations. Life rate is getting more, the meaning situation is getting better, and increasing the mortality rate by both metrics and death cases will cause a big worry. Iran's mortality rate has come under the global mean and is continuing under this line. Also, the recovery rate in Iran has some problems as said before, and has a lot of variation around the mean rate. Toward our prediction, with most probability, the peak of the outbreak is 150 days after the start and it will resist the same amount of days till the end meaning 290–320 days and Gaussian function works well for the long-term phase.

Acknowledgements In this section, you can acknowledge any support given which is not covered by the author contribution or funding



sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Funding This research received no external funding.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Ahmadi M, Sharifi A, Dorosti Sh, Jafarzadeh Ghoushchi S, Ghanbari N (2020) Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Science of the Total Environment, 720
- Ahmed N, Maqsood A, Abduljabar T, Vohra F (2020) Tobacco smoking a potential risk factor in transmission of covid-19 infection. Pak J Med Sci 36:104–107
- Anastassopoulou C, Russo L, Tsakris A, Siettos C (2020) Data-based analysis, modelling and forecasting of the COVID-19 outbreak: PLOS ONE, 15, 1–2
- Choi T-M, Wallace SW, Wang Y (2018) Big data analytics in operations management. Prod Oper Manag 27:1868–1883
- Database: https://data.mendeley.com/datasets/hdw2zhmw4f/1
- Fanelli D, Piazza F (2020) Analysis and forecast of COVID-19 spreading in China, Italy and France: Chaos, Solitons and Fractals, Nonlinear Science, and Nonequilibrium and Complex Phenomena, 134, 109761
- Guo Q, Li M, Wang C, Wang P, Fang Z, Tan J, Wu S, Xiao Y, Zhu H (2020) Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm
- Hazen BT, Skipper JB, Boone CA, Hill RR (2016) Back in business: operations research in support of big data analytics for operations and supply chain management. Ann Oper Res 270:201–211

https://www.isna.ir/news/

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497–506
- James S, Fong GL, Dey N, Crespo RG, Herrera-Viedma E (2020) Finding an Accurate Early Forecasting Model from Small Dataset a Case of 2019-nCoV Novel Coronavirus Outbreak. Int J Interact Multim Artif Intell

- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z (2020) Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 382:1199–1207
- Randhawa GS, Soltysiak MPM, Roz HE, De Souza CPE, Hill KA, Kari L (2016)
- The faculty of Intensive Care Medicine, The Intensive Care Society, The Association of Anaesthetists and The Royal College of Anaesthetists reported at 2019. www.icmanaesthesiacovid-19.org/background
- The New York Times. Recipe for a Massive Viral Outbreak': Iran Emerges as a Worldwide Threat. Wana News Agency, via Reuters 2020. www.nytimes.com/2020/02/24/world/middleeast/coron avirus-iran.html
- United Nations Development program. The UN's framework reported at 2020. www.undp.org/content/undp/en/home/coronavirus/socio-economic-impact-of-covid-19.html
- World Health Organization. Global Situation Report 209–2020. www. who.int/docs/default-source/coronaviruse/situation-reports/20200 816-covid-19-sitrep-209.pdf?sfvrsn=5dde1ca2 2
- World Health Organization. Global Tuberculosis Report 2020. www.covid19.who.int/. Accessed 16 Aug 2020
- World meter. World Population report 2020. www.worldometers.info Wu X, Nethery R-C, Sabath M-B, Braun D, Dominici F (2020) Exposure to air pollution and COVID-19 mortality in the United States: a nationwide cross-sectional study. J Chem Inf Model 53:1689–1699
- Xiaowei Xu1, M. X. J., Md, Chunlian Ma3, Md; Peng Du4; Xukun Li4;
 Shuangzhi Lv5, Md; Liang Yu1, Md; Yanfei Chen1, Md; Junwei Su1, Md; Guanjing Lang1, Md; Yongtao Li1, Md; Hong Zhao1, Md; Kaijin Xu1, PHD Md; Lingxiang Ruan5, Md; Wei Wu1, PHD Md. (2020). Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia. Applied Intelligence.
- Zhang H, Saravanan KM, Yang Y, Hossain MT, Li J, Ren X, Wei Y (2020) Deep learning based drug screening for novel coronavirus 2019-nCov, 19
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B,
 Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao
 GF, Tan W, China Novel Coronavirus, I. & Research, T (2020)
 A Novel Coronavirus from Patients with Pneumonia in China, (2019). N Engl J Med 382:727–733

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

