

Highlights

An intelligent method for prediction of COVID-19 fatalities in Iran

Roozbeh Badiee, Amir H.D. Markazi

- Examining Some Inherent Characteristics of COVID-19
- Exploring Various Prediction and Modeling Methods for COVID-19
- Based on the available data, the discrete logistic model has been selected for prediction in Iran.
- Two methods have been introduced for using discrete logistic regression in predicting COVID-19
- The first parameter estimation method demonstrated superior accuracy in its predictions, whereas the second method, is more user-friendly but offers lower accuracy.

An intelligent method for prediction of COVID-19 fatalities in Iran

Roozbeh Badiie^a, Amir H.D. Markazi^a

^a*Mechanical Engineering Department, Iran University of Science and Technology, , 16844, Tehran, Iran*

Abstract

We conducted a study in Iran to develop an improved approach for predicting COVID-19 fatalities. The study analyzed various factors that contribute to the spread of the virus, such as transmission rates, vaccine efficacy, and climate impact. Several prediction techniques were evaluated, and the logistic model showed promise due to its practicality and accuracy. To enhance the forecasting of COVID-19 mortalities, two methodologies were introduced. The first approach integrated predictions from other models using a neural network to refine the logistic model's parameters. The second approach relied solely on available data and initial guesses for logistic model parameters. These techniques aim to assist in combating the pandemic by facilitating better planning and resource allocation in Iran.

Keywords: Discrete logistic model, SIRDH model, COVID-19, Iran, LSTM

1. Introduction

The precise forecasting of COVID-19 mortality rates is of utmost importance in order to make well-informed policy choices. Focusing on fatalities rather than just cases is more significant in representing the virus's overall impact on public health. Making predictions about mortality offers policymakers significant information that may be used to allocate resources and ensure public health protection efficiently.

Numerous studies have been conducted on COVID-19 modeling, including the research study by Yazdani et al. Their research examined the effectiveness of machine learning algorithms in predicting the survival outcomes of COVID-19 patients in Iran. The findings suggest that machine learning models have the potential to accurately forecast the survival rate of COVID-19

patients, which can help healthcare professionals provide more individualized and effective care, especially to those at high risk[1]. Yajada et al. conducted a study in which they formulated a mathematical model to predict the incidence of COVID-19 cases and mortality rates in three states, including Iran, the United States, and South Korea. The research investigated six distinct models, namely Fourier, Interplant, Gaussian, Polynomial, Sum of Sine, and Smoothing Spline, in order to identify the optimal model with the lowest error rate. According to the study, the smoothing spline model was found to be the most efficacious in forecasting COVID-19 mortality rates across all three nations[2]. Ehsan Badfar et al. conducted a study on enhancing the control of the nonlinear epidemiological model of COVID-19 by designing a robust sliding mode controller. Their proposed model incorporates vaccination, social distancing, facial masks, and medical treatment as control inputs. To estimate the unknown parameters of the system, the authors employed the long short-term memory (LSTM) algorithm. They successfully developed a robust sliding mode controller using Lyapunov stability analysis. Simulation results demonstrated the effectiveness of the proposed controller in mitigating the spread of COVID-19 within society. However, the authors acknowledged that several challenges remain in eradicating the pandemic, such as the emergence of new variants of the virus and varying levels of vaccine effectiveness against these variants. In conclusion, this research offers valuable insights for policymakers in implementing effective measures to combat the ongoing COVID-19 crisis[3].

The use of the SIR model and compartmental models has been extensively adopted by a multitude of researchers for the goal of anticipating the spread of COVID-19. The authors of the study, Marina Medvedeva et al., have presented a novel approach for determining the parameters of the SIR model in their research [4]. In other investigations conducted by Chozbeh C. Calafiore, Carlo Novara, and Corrado Possieri, they aimed to comprehend the spread of the coronavirus in Italy [5]. Cooper, Mondal, and Antonopoulos conducted a study to evaluate the efficacy of modeling techniques in predicting the spread of epidemic outbreaks over a period of time[6].In a study conducted by Shiva Moein et al., the predictability of the COVID-19 pandemic using the SIR model in the Isfahan province is investigated. The results of this study demonstrate that the SIR model has the capability to predict the disease's prevalence in a short-term time frame but cannot accurately forecast the onset of the disease in a community over the long term [7].

Despite the inherent difficulty in predicting COVID, the utilization of

Long Short-Term Memory (LSTM) represents a viable approach to interpreting and strategizing against this infectiousillness. Several research studies have been carried out to explore the application of Long Short-Term Memory (LSTM) in forecasting COVID-19. Chandra et al. conducted a study aimed at proposing a framework for COVID prediction in India utilizing LSTM. The study's findings suggest that the ED-LSTM model, when combined with random data splitting and the incorporation of a data variable for neural network training, exhibits superior performance compared to other models[8]. Vinay Kumar Chimmula and colleagues performed COVID prediction using data from affected individuals in Canada, the United States, Italy, and Spain, with a focus on Canada. They encountered challenges such as small data volumes, unknown variables, and external factors that complicate COVID prediction [9].

A study by Shahid et al. investigated various predictive models, including ARIMA, SVR, LSTM, Bi-LSTM, and GRU, for forecasting COVID-19 cases. The study found that the Bi-LSTM model demonstrated superior predictive abilities in forecasting COVID-19 cases in countries with large populations. Specifically, the Bi-LSTM model achieved the lowest MAE and RMSE values for mortality in China when compared to other models. Given its high accuracy and reliability, the Bi-LSTM model could be effectively utilized by public health officials and policymakers in their efforts to control the transmission of COVID-19 [10]. Nurul Absar et al. used LSTM models to predict COVID-19 cases, fatalities, and recoveries in Bangladesh. The LSTM models showed better results in simulation and prediction than prior research [11].

Kafieh et al. conducted a study utilizing deep learning models, specifically multilayer perceptron, random forest, and various LSTM versions, to forecast the prevalence of COVID-19. The study drew upon data from multiple countries, with a particular emphasis on Iran. The study's findings indicate that the multivariate LSTM (M-LSTM) model was the most effective in predicting COVID outcomes, as it incorporated data on the number of infected, deceased, and recovered individuals as inputs. Subsequent analysis indicated that the utilization of data that spans six to eight days prior resulted in more accurate prognostications [12]. Long Short-Term Memory (LSTM) networks have been utilized for the purpose of predicting the occurrence of COVID-19. The reliability of LSTM predictions for long-term forecasts may be compromised by various factors, including the intricate nature of infection spread, insufficient data collection, and unreliable reporting.

2. COVID-19

COVID-19 is an infectious respiratory disease that began in December 2019 in Wuhan, China. COVID-19 can cause respiratory symptoms similar to cold, flu, or pneumonia. This disease may be mild, but some people, particularly the elderly and those with special medical conditions, become seriously ill. Vaccines against COVID-19 are effective in preventing severe disease and decreasing the disease's incidence.

2.1. *COVID-19 variants*

Novel variants of the SARS-CoV-2 virus surfaced in the spring of 2020. Various strains of the virus emerged, each possessing distinct capabilities and characteristics, ultimately replacing their predecessors. The emergence of novel variants of this virus has been observed to exhibit heightened transmissibility. The Alpha (B.1.1.7), Beta (B.1.351), and Gamma (P.1) variants were identified in the United Kingdom, South Africa, and Brazil, respectively. The Delta type (B.1.617.2), which appeared in India in the summer of 2021, gradually replaced these old variants. In late 2021, another type of the COVID-19 virus, known as B.1.1.529 or Omicron, emerged in Africa and rapidly gained global prevalence due to its increased transmissibility. The present virus exhibits more than fifty mutations, resulting in the ineffectiveness of numerous neutralizing antibody responses that were triggered by vaccination or prior infection with earlier strains of the COVID-19 virus.

2.1.1. *The Omicron variant*

In late November 2021, the first variant of Omicron (BA.1) was identified in Botswana and South Africa, and it quickly spread to other countries. By December of that same year, Omicron had caused a significant increase in daily cases in the United States, resulting in over one million infections. As of August 2022, one of its sub-lineages, BA.5, was responsible for over 88% of cases in the US, making it the most prevalent variant. Another related variant of this species is BA.4, which accounted for approximately 8% of reported cases at that time[13].

. The Omicron variant is more contagious than the Delta variant. Its effective reproduction number is 3.8 times higher, and its basic reproduction number is 2.5 times higher. The average effective reproduction number for the Omicron variant is 8.2, and the basic reproduction number is 3.6[13]. This

increased contagiousness is attributed to the presence of more than thirty mutations on the spike protein of the Omicron variant, potentially heightening the likelihood of infection. Despite its higher transmissibility, the primary Omicron variant shows lower virulence compared to previous variants, as reported by the Centers for Disease Control and Prevention (CDC)[14].

2.1.2. The Delta variant

The Delta variant, also referred to as B.1.617.2, was initially discovered in India in December 2020, and it led to a severe surge of COVID-19 infections in April 2021. It was first identified in the United States in March 2021 and became the primary cause of COVID-19 worldwide until other variants started spreading[15]. Based on the investigations conducted by the Food and Drug Administration of the United States, the Delta variant of the previous strains of the COVID-19 virus exhibits a higher capability to hospitalize individuals, with a transmission rate of approximately 40% to 60% greater than previous variants, such as Beta. The National Health Service of Britain has announced that the Delta variant possesses approximately 90% higher transmission power compared to earlier strains. Further examinations have also demonstrated that this variant leads to a 108% increase in hospitalizations, a 235% increase in intensive care unit admissions, and a 133% higher likelihood of mortality if infected with this variant of the virus [14, 16].

2.1.3. The Gamma Variant

The P.1, or Gamma variant, was discovered in Brazil in December 2020 and was reported in the United States in January 2021. Although there is no conclusive evidence that this variant causes a more severe form of the disease, it is believed to be more contagious than the Alpha and Beta variants. The increased contagiousness of the Gamma variant makes the disease it causes more lethal compared to less infectious variants. This poses a more significant threat to society as faster-spreading variants worsen the challenges faced by public health and local healthcare systems. As of the latest available information, there have been no widespread reported cases of this variant in Iran [17].

2.1.4. The Beta Variant

The Beta variant also referred to as B.1.351, is characterized by multiple mutations. It was initially identified in South Africa in October 2020 and played a significant role in this country's second wave of COVID-19. The

Beta variant was later detected in the United States in late January 2021. Studies conducted by Kevin Linka and Mathias Perlink demonstrate that the B.1.351 variant is between 50% and 56% more transmissible than other variants. These findings highlight how even small changes in transmissibility can substantially impact the number of people falling ill or succumbing to the disease [18].

2.1.5. The Alpha variant

Towards the end of 2020, the B.1.1.7 variant of SARS-CoV-2, also referred to as the Alpha variant, was discovered in the United Kingdom. Research indicates that this strain of Coronavirus is more infectious than the initial variant [19, 20]. Research has shown that the Alpha variant is responsible for a significant increase in transmission, ranging from 43% to 90% compared to previous strains. This increased transmissibility has been observed in the United Kingdom and other countries such as Denmark, Switzerland, and the United States.

. Following is a table that, on the basis of a review, provides a summary of the transferability of each species.

Table 1: Transferability of Different COVID-19 Variants Compared to the Wild Type

Variant	Transferability Ratio compared to the Wild Type
Alpha	55%
Beta	50%
Gamma	60%
Delta	95%
Omicron	230%

2.2. COVID-19 vaccines

The widespread distribution of vaccines is crucial in preventing the spread of infectious diseases. Iran administered its first vaccine injection on February 9, 2021. However, as of July 3, 2023, there have been 7,612,728 confirmed cases of COVID-19 and 146,204 deaths in Iran. On a positive note, Iran has administered 155,429,149 vaccine doses, utilizing 14 different vaccines. The majority of these vaccines come from Sinopharm and AstraZeneca for COVID-19 vaccination [21].

2.2.1. AstraZeneca vaccine

A study involving 11,636 individuals aged 18 to 55 found that those who received two standard doses experienced a relief effect of 70.4% [22]. Additionally, the vaccine has shown an efficacy of approximately 90% when the first dose is administered as a half dose and the second dose as a full standard dose [22]. According to Shinde et al., the vaccine demonstrates a 51% efficacy rate against the Delta variant. [23]. Moreover, Andrews et al. concluded that individuals who received two doses of the vaccine had a vaccine efficacy of 62.4% against the Omicron B.1.1.529 variant [24].

2.2.2. Sinopharm BBIBP-CorV vaccine

The Beijing Institute of Biological Products developed the BBIBP-CorV. The efficacy of this vaccine was examined through various phases of the study. The initial stage of the study, as carried out by Xia et al., involved an examination of the effectiveness and potential adverse reactions associated with the vaccine [25]. The vaccine in question is an inactive virus administered to individuals in two separate doses with a 12-day interval between them. During the third phase of the clinical trial, the vaccine's clinical efficacy was determined to be 79% [26]. The study conducted in the United Arab Emirates revealed that the Sinopharm vaccine demonstrated a 62% efficacy against the Delta strain [27]. In a study conducted on 612,597 people, it was found that the vaccine against omicron species has an effectiveness of 86% [28].

2.3. The effect of climate on COVID-19 transmission

In a review article by Alvaro Breez-Redon and Angel Serrano-Aruca, the impact of various weather factors on the transmission of COVID-19 was examined. After analyzing multiple studies, the researchers concluded that there is no clear correlation between environmental and meteorological variables and the rate of transmission of COVID-19. Only a few researchers consider factors such as temperature to be indirect contributors to the transmission rate [29]. In another study, the effect of effective weather parameters in Iran on the rate of COVID-19 transmission was evaluated, and a result similar to Alvaro's review article was obtained. They also discovered that population density has a more significant effect on transmission rate [30].

Table 2: Effectiveness of Vaccines Used in Iran Against Different Variants

Vaccine Type	AstraZeneca Vaccine	Sinopharm Vaccine (<i>BBIBP-CorV</i>)
Vaccine Effectiveness in Initial Results	90.00%	90.00%
Vaccine Effectiveness against Delta Variant	51.0%	62%
Vaccine Effectiveness against Omicron Variant	62.4%	86%

3. Data Sources

We utilized data from two primary sources: primarily from [31], and in cases where data was unavailable, supplementary information was extracted from [32]. These datasets were employed to train a variety of models for our analysis. The collected data encompasses several key metrics: total patient count, new cases, daily and cumulative fatality counts, as well as the overall count of administered vaccines.

However, it's worth noting that the complete dataset lacked certain entries, particularly for the total count of administered vaccines. To address this gap, we employed linear regression techniques to estimate and fill in this missing values.

4. Genetic Algorithm

John Holland and his team developed the genetic algorithm during the 1960s and 1970s, drawing inspiration from biological evolution and Darwin's theory of natural selection. Holland investigated the application of crossover, mutation, and selection in the study of adaptive and artificial systems, and he developed the genetic algorithm as a problem-solving approach as a result of his research. Since then, several genetic algorithms have been created and used for a wide range of optimization issues, including pattern recognition

and multi-objective engineering optimization. Genetic algorithms have advantages over traditional optimization algorithms in handling complex issues and parallelization.

The early convergence problem is one of the most notable benefits of the genetic algorithm. This occurs when the population of possible solutions gets too homogeneous, and the algorithm can no longer explore the search space efficiently. Premature convergence may occur for several reasons, such as lack of diversity in the initial population or over-reliance on a single operator, such as mutation or crossover. However, this approach is a useful tool for a variety of issues, such as curve fitting and parameter estimation for ode systems [33].

. One of the advantages of genetic algorithms is their ease of implementation. Unlike traditional optimization algorithms, which often require complex mathematical models and assumptions, genetic algorithms can be applied to a wide range of problems without requiring extensive domain knowledge or mathematical expertise. This is because genetic algorithms rely on simple principles of evolution and natural selection that are easy to understand and implement using basic programming techniques. In addition, many software packages and libraries are available that provide pre-built genetic algorithm implementations, simplifying the implementation process. In general, the ease of implementation of genetic algorithms makes them a practical and accessible tool for solving a variety of optimization problems.

5. Different method for COVID-19 prediction

5.1. Autoregressive moving average or ARIMA

Due to the global impact of COVID-19, it is important to develop effective planning techniques for managing the illness. One such technique involves using mathematical models, like the Autoregressive Integrated Moving Average model, to predict the number of COVID-19 deaths. However, it is essential that these forecasts are based on sound reasoning. The ARIMA model has successfully predicted past outbreaks, such as SARS and H1N1, to a certain extent [34].

. ARMA, or Autoregressive Moving Average Model, combines the moving average model and the Autoregressive process. A moving average model describes the current value of a time series as a linear function of previous values and random errors from a normal distribution. The ARMA Equation can be expressed as follows:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (1)$$

In Equation 1, p is the number of past terms; similarly, q shows the number of error terms used in the ARMA process, and C is constant. Note that the ARMA model can be used only on stationary data.

Autoregressive Integrated Moving Average (ARIMA). The ARIMA model is similar to the ARMA model but is used when the data is non-stationary. In some cases, non-stationary data can be transformed into stationary data differencing. Although the differenced data may not be useful, it can be forecasted using the ARMA model. Therefore, ARIMA involves predicting the differenced time series. This model is usually represented as $ARIMA(p, d, q)$, where p represents the order of the autoregressive model, q represents the order of the moving average model, and d represents the number of Differencing applied to the raw data.

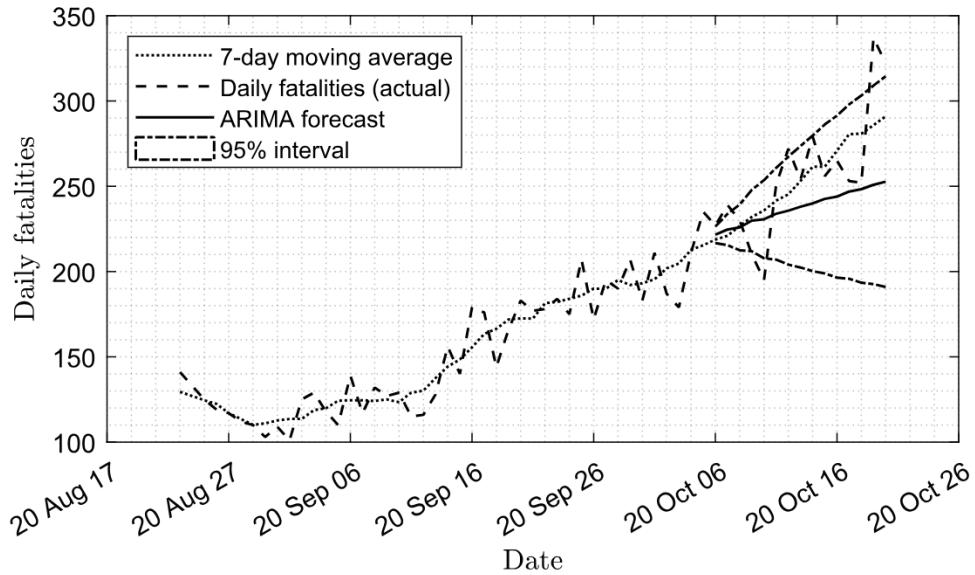


Figure 1: Forecasting deceased population for the third wave of COVID-19 in Iran using ARIMA

5.1.1. Forecasting fatalities in Iran using ARIMA

For effective ARIMA forecasting, it's crucial that a time series or its differences should be stationarity. After examining the available data and relevant research sources [35, 36, 37], we have determined that the first or second difference of the COVID-19 data demonstrates stationarity over a brief timeframe.

Upon further investigation, we have deduced that within approximately 3 to 4 weeks, the *ARIMA(5, 2, 3)* model holds promise for predicting the number of deaths over short intervals, such as ten days.

. As seen in figure 1, ARIMA can only be used for the short-term prediction of COVID-19. Using a non-linear version of ARIMA is better because our data is not stationary, and the model cannot model and predict COVID for a period longer than ten days[38]. Given that we can assume that our data is stationary at short intervals, specifically at the beginning of each wave of COVID-19, we should be able to utilize this model in combination with another model to forecast the number of deaths caused by COVID-19 in Iran.

5.2. Discrete logistic model

Discrete logistic model is a mathematical model that depicts the growth of a population over time at discrete or discontinuous distances. Discrete logistic models are a discrete variant of the continuous logistic model, which is typically employed to model the growth of populations in natural systems. The discrete logistic model Equation is shown in Equation 2. In this equation, M and r are called the population's growth rate and carrying capacity.

$$P_{n+1} = r \times P_n \left(1 - \frac{P_n}{M} \right) \quad (2)$$

The discrete logistics model had wide-ranging applications, including modeling population growth in ecosystems, understanding disease prevalence, and simulating the behavior of financial markets. In addition to this, they are utilized in a wide variety of sectors, including engineering and computer science, and they are capable of modeling the behavior of complex systems.

The discrete logistic model has its own set of limitations, one of which is that it assumes a constant natural growth rate, which may not always be the case in real-world populations. In addition, the model does not account for factors such as migration or genetic mutations of COVID-19, which may

substantially affect population growth. Despite this, a number of people have used this model to make COVID-19 predictions. Wang and colleagues, for instance, have evaluated the parameters of this model using the Facebook artificial intelligence Prophet model [39]. Zhou and his colleagues have studied this model and compared it to the SERI model, and they have discovered that the logistic model performs quite well in predicting the COVID-19 wave[40].

5.2.1. discrete logistic model for the perdition of COVID-9 death in Iran

We proposed, in order to forecast COVID-19, to make the following adjustments to the discrete logistic model and express the equations of this model as follows:

$$p_{n+1} = r \times p_n(1 - p_n) \quad (3)$$

p is defined in 4 as follows:

$$p = \frac{P}{M} \quad (4)$$

and is calculated as so:

$$M = 81160000 \times 0.7 \times 0.25 \times M' \quad (5)$$

p The normalized cumulative population is the number of people who died, and M equals the number of people likely to get sick and die. And M itself comes from the Equation 5, where M' is the ratio of the number of people likely to be at risk of developing COVID-19 disease. It's better to use M' instead of M because the change in M' is less than M , so it's easier to predict. We can calculate M by trial and error.

Two methods can be used to calculate the parameters of the logistic model. However, it is essential to note that in this study, the similarity of the two waves is determined by how close the average number of fatalities and the number of new cases are within the first 10 to 14 days of each wave. If necessary, the value of r can be computed using Equation 6 and compared with other waves to identify the most similar one to the wave under investigation. It is also worth mentioning that we predict the cumulative population using the logistic model in our research. However, we convert the cumulative predictions to daily predictions for examination and visualization purposes.

first method

To calculate r , we used Equation 3 and obtained the relation for r as shown in equation 6.

$$r = \frac{p_{n+1}}{p_n(1 - p_n)} \quad (6)$$

where

$$P = d - d_{offset} \quad (7)$$

In Equation 7, d is the population of the deceased population(cumulative cases), and population d_{offset} is equal to the average population of the deceased population at the beginning of each wave. The value of r can be determined by utilizing the equation 6.

Note that to calculate r using Equation 6, we have to use smooth data to get acceptable results. Smooth data is the moving (rolling) average of our data.

. The proposed method involves initial data adjustment of each wave using Equation 7, followed by determining the value of r by utilizing Relation 6. Finally, the value of M' prime is derived via a trial-and-error process. Notably, the trial and error approach utilized in estimating M' is the most challenging aspect of this method. Furthermore, The calculation of r can be challenging due to the need for sufficient data to make accurate predictions.

. The graph in Figure 2 shows the predicted outcomes for the third wave of COVID-19 in Iran. The value of M' remains constant for all predictions and is determined through trial and error. One curve is based on data until October 14, 2020, while the other is based on data until October 7, 2020. Evidently, this graph highlights the sensitivity of the logistic model to the quantity of data employed for parameter estimation. Even a minor divergence in data can significantly influence the outcomes due to the model's susceptibility to initial conditions and parameters.

. To address the issue of estimating r , the ARIMA prediction method can be used by applying equation 6 to the prediction of ARIMA. The estimated r obtained through this process is r_{arima} . Figure 2 shows the prediction result, which was generated using r_{arima} , closely matching the 7-day moving average of data. However, it's important to highlight that despite progress in addressing the issue of estimating r , the challenge of estimating M' persists even with integrating the ARIMA model with logistics.

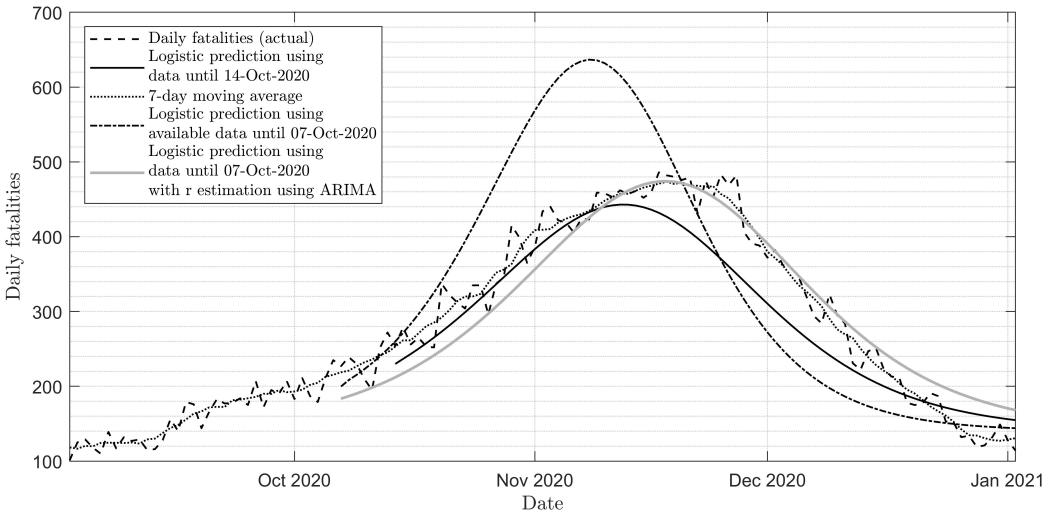


Figure 2: Prediction of the third wave of COVID-19 in Iran using the logistic first method and combined with ARIMA

second method

This approach involves using genetic algorithms to calculate system parameter values. To determine these values, the mean square error (MSE) is utilized as a suitable performance measure. The cost parameter is indicative of the MSE between the data at hand and the logistic model's forecast. By utilizing genetic algorithms based on the MSE criteria; the model parameters can be efficiently optimized, leading to an improved alignment between the logistic model's predictions and the actual data.

. This method involves making predictions at the start of each wave by using the initial conditions of diseased individuals (d_0) as parameters. In addition to r and M' , we must estimate these parameters to ensure accurate predictions. To utilize the logistic model effectively, When utilizing a genetic algorithm, it is necessary to estimate both lower and upper parameters.

We can employ the first method to obtain an initial approximation for the lower and upper bounds of variable r . Specifically, the upper bound can be determined as 1.5 times the value of r obtained through the first method, while the lower bound can be calculated as 0.5 times the value of r obtained through the first method. For the variable d_0 , its lower and upper bounds can be estimated by analyzing the data from the earlier wave. Similarly, the

variable M' necessitates the adoption of a similar methodology.

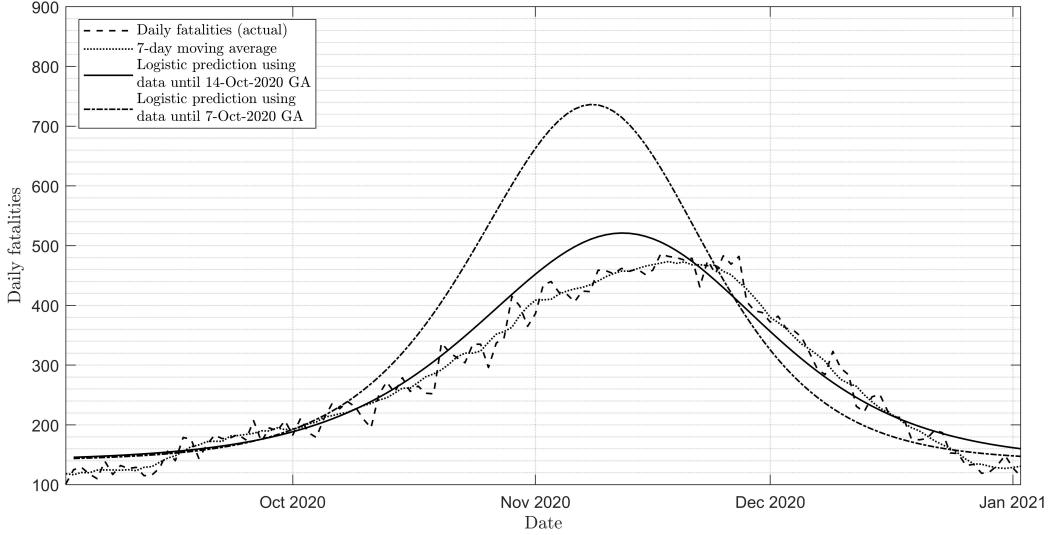


Figure 3: Result of estimation of logistic model parameters using Genetic algorithm

The prediction outcomes of the logistic model, which were computed via the utilization of the genetic algorithm for parameter estimation, are illustrated in Figure 3. Based on COVID-19 data until October 14, 2020, the forecasted curve shows that the logistic model has correctly predicted the third wave of COVID-19 in Iran. However, upon utilizing a reduced data set spanning from September 2, 2020, to October 7, 2020, It is clear that the model's forecast deviates from expected results and from reported deaths in Iran. This prediction is considered credible as the logistics model has exhibited its capacity to depict the overall trajectory of COVID-19 growth in Iran accurately. The necessity to periodically update prediction models in response to changing conditions and new data arises from their limitations in accurately forecasting societal behavior. Therefore, incorporating predictive models should be considered an additional tool in the decision-making process for COVID-19 and other diseases.

5.2.2. Remarks on the logistic model

We conducted a comparison of two methodologies to determine the parameters of the logistic model and assessed their predictive capabilities. Our findings showed that the first method, which involves parameter computation using ARIMA model integration, provides better results. However,

estimating the parameter M' is challenging, but we can approximate its range. Therefore, we recommend using the second approach to identify system parameters. We have developed an improved methodology based on this technique. It's worth noting that calculating r with ARIMA can introduce errors, so it's better to use a method that considers all factors when estimating the logistic model's parameters. The genetic algorithm is an easier way to calculate these parameters, and we used it to estimate different wave results, as shown in Table 3.

Table 3: Calculated Parameters for the Logistic Model

Wave Number	First Day	Last Day	Date first day of Forecast	M'	r	r_{arima}	r_{ga}	M'_{ga}	$d_{0_{ga}}$
1	February 20, 2020	May 23, 2020	March 14, 2020	0.001	1.2351	1.162972	1.0944	0.0071	100
2	May 23, 2020	September 2, 2020	June 28, 2020	0.0059	1.0982	1.092817	1.1038	0.0058	93
3	September 2, 2020	January 2, 2021	October 14, 2020	0.015	1.0835	1.0825	1.0825	0.017	207
4	March 15, 2021	June 23, 2021	April 16, 2021	0.0075	1.1405	1.1470	1.1205	0.0108	150
5	June 21, 2021	October 30, 2021	July 31, 2021	0.019	1.0912	1.0814	1.1093	0.0178	171
6	January 17, 2022	April 9, 2022	February 11, 2022	0.00235	1.1777	1.1676	1.1268	0.0055	72
7	July 10, 2022	September 8, 2022	July 30, 2022	0.0014	1.11369	1.134	1.1149	0.0011	71
8	February 28, 2022	May 26, 2023	March 19, 2023	0.00025	1.125481	1.143271	1.05965	0.0019	70

5.3. LSTM

The LSTM (Long Short-Term Memory) networks were developed by Hochreiter and Schmidhuber as a solution to the constraints of conventional RNNs. These networks belong to the class of recurrent neural networks. LSTM networks regulate data transmission within the long-term memory or cell state by incorporating three gates: the input gate, the output gate, and the forget gate. The cell state is a fundamental element of LSTM networks, serving as a repository of information that is continually refreshed and sustained across temporal iterations. This feature enables the network to preserve knowledge from prior time steps. The gating mechanisms, namely the input and forget gates, are responsible for regulating the inclusion or exclusion of information in the cell state. On the other hand, the output gate plays a crucial role in determining the extent to which long-term memory should be transmitted to the output. Long Short-Term Memory (LSTM) networks have gained significant popularity in various domains such as natural language processing, speech recognition, and time series prediction[41].

5.3.1. LSTM formulation

The Forget Gate: Within a Long Short-Term Memory (LSTM) cell, the forget gate mechanism is responsible for managing the data stored in the memory cell. This mechanism allows the LSTM to forget any irrelevant

information that is not needed for the current task. The forget gate works by applying a sigmoid function to both the hidden state and input, which produces a numerical output ranging from 0 to 1 for each component of the cellular state. This numerical output represents the degree of importance of each component. Using this system, the LSTM decides which information to retain and which information to forget. A value of 1 means that the information is retained, while a value of 0 means that it is forgotten. In other words, the forget gate is a mechanism that helps the LSTM to choose which information is useful and which information can be disregarded.

The forget gate is represented by f_t in a Long Short-Term Memory (LSTM) cell. The weight matrix and bias vector, which are associated with each gate in each of these equations, are here denoted by the symbols W and b . The symbols h_{t-1} and x_t , respectively, indicate the prior hidden state and the current input, while *sigma* denotes the sigmoid function.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (8)$$

The Input Gate:: The function of the input gate within a Long Short-Term Memory (LSTM) cell is to regulate the flow of new information into the cell state. The Long Short-Term Memory (LSTM) model is capable of selectively incorporating relevant information into the cell state C_t at a given time step t . The variable i_t in the equation of an LSTM cell denotes the input gate. The calculation is derived through the utilization of the subsequent formula:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$

The function of the input gate is to incorporate new data into the cell state (C_t) through the process of multiplying the candidate cell state (\tilde{C}_t) with the output of the sigmoid function. The formula representing the candidate cell state (\tilde{C}_t) is presented below:

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10)$$

Cell State. In an LSTM architecture, the cell state is a vector used to store information that is deemed important for the model. It is denoted as C_t . The computation of the cell state C_t can be derived from the explanations presented in the preceding section.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (11)$$

output gate. The utilization of the output gate output, represented as o_t , is an important part of the LSTM architecture, as it serves to determine the ultimate output derived from the current cell state. The output gate is a set of numerical values ranging from 0 to 1, which regulates how much output data is derived from the cell state. The computation of the output gate is determined by the subsequent formula:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (12)$$

The final hidden state of the LSTM at time step t , denoted as h_t , is calculated as:

$$h_t = o_t \cdot \tanh(C_t) \quad (13)$$

Here, C_t represents the cell state at time step t . The \tanh function is applied element-wise to the cell state, and the resulting value is multiplied by the output gate to obtain the final hidden state. Then we can use the hidden state to calculate x_{t+1} .

5.3.2. Bidirectional LSTM

The utilization of Long Short-Term Memory (LSTM) entails the derivation of the subsequent token predicated on the token from a singular direction. However, certain scenarios could justify the inspection of both directions of the sequential data. Bi-LSTM is employed for this objective by using a pair of LSTM layers. The layer integrates the output of two LSTM layers and subsequently generates a prediction based on the merged output [42].

5.3.3. COVID-19 prediction using LSTM in Iran

The Iran COVID-19 data was used to evaluate the effectiveness of LSTM (Long Short-Term Memory) recurrent neural networks in predicting COVID-19. To conduct the investigation, it is important to consider different factors related to the network architecture and data usage. Previous studies have shown that LSTM and bi-LSTM networks have demonstrated promising results in the field of COVID-19 prediction, as supported by other relevant research [12, 10]. Therefore, we decided to analyze these two networks.

. It's possible to use the recently updated data on deceased populations alone or include additional information, such as newly reported cases and administered vaccinations, to make predictions. LSTM models can be used to forecast multiple variables or focus on a single variable, such as the number

Table 4: RMSE Results for Different Network Types

Network Type	Mean RMSE	Lower 95% Confidence Interval	Upper 95% Confidence Interval
LSTM Network (Training)	20.50	20.51	22.48
LSTM Network (Testing)	40.51	38.65	42.36
LSTM Network (Validation)	39.06	36.84	41.28
Bi-LSTM Network (Training)	2.55	2.34	2.76
Bi-LSTM Network (Testing)	6.84	5.73	7.94
Bi-LSTM Network (Validation)	6.18	5.33	7.02

Confidence intervals are calculated based on forty sets of training networks.

of deceased individuals. Multivariate LSTM models are suitable for predicting time series with multiple features, as they can capture complex patterns in the data and provide more accurate predictions. However, predicting long-term outcomes involving multiple factors may be challenging in this scenario[43]. So, in order to forecast the number of fatalities over the next seven days, we decided to incorporate a variety of variables from the previous 10 days. Considering that the network performance improved when the data were randomly split between testing and validation, we employed this approach for training [8]. Finally, we considered raw data, first derivatives of deceased individuals, and moving averages of raw data as inputs for the LSTM network.

. To pick the most suitable network, we standardized the data using mean and standard deviation. The data was then randomly partitioned into three subsets for training, validation, and testing, with 80%, 10%, and 10% allocation, respectively. We trained two types of networks based on the available data - the LSTM architecture with two LSTM layers of dimensions 64 and 16, respectively, connected to a Dense layer, and the bi-directional Long Short-Term Memory (bi-LSTM) architecture, which employs a bidirectional layer instead of a unidirectional layer. To compare the two networks, we used the Root Mean Square Error (RMSE) and partitioned the data into groups of forty times for each network. Table 4 shows the results of our analysis.

Upon analyzing Table 4, it is evident that the Bi-LSTM model has outperformed other LSTM network in terms of prediction accuracy. Based on our research and other literature, we suggest utilizing the Bi-LSTM framework for COVID-19 forecasting [10]. Figure 4 illustrates the third wave of COVID-19 prediction using the Bi-LSTM network. A comparative analysis between Bi-LSTM and ARIMA shows that Bi-LSTM is a suitable option

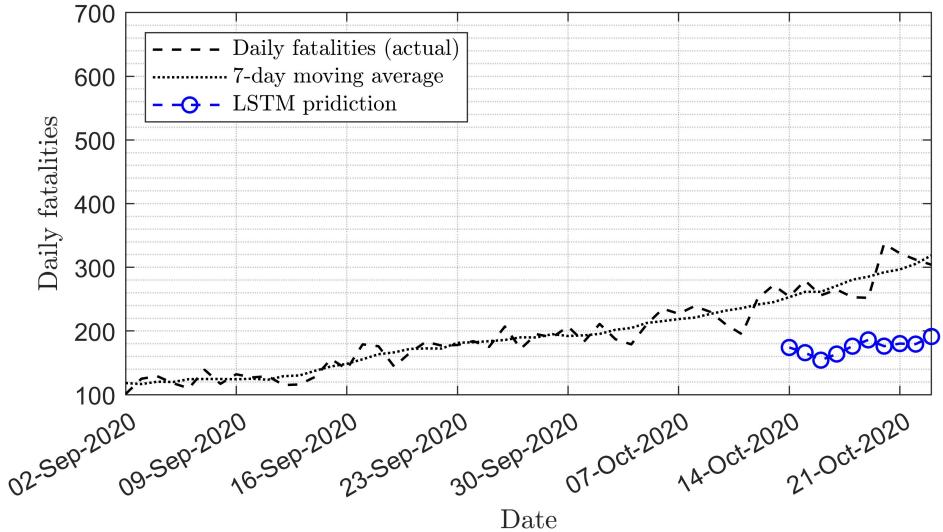


Figure 4: COVID-19 Prediction Using Bi-LSTM

for short-term forecasting, typically covering one week or ten days. Nevertheless, models like SIR and logistics exhibit better performance for longer periods. Bi-LSTM has limitations when it comes to certain types of sequential data, such as irregular patterns, which may affect its long-term prediction accuracy. Furthermore, various factors like limited data availability, population density, social behavior, cultural practices, lifestyle, and governmental policies can hinder an accurate assessment of COVID-19 spread. The combination of these factors and the constraints of Bi-LSTM on specific types of sequential data may lead to suboptimal outcomes in long-term forecasting.

5.4. Compartmental models

Compartmental models are a prevalent type of mathematical model employed in the field of epidemiology to investigate infectious diseases. Compartmental models divide the population into different categories based on disease and societal status. For instance, a society can be divided into three different population groups, such as susceptible (S), infected (I), and recovered (R). The aforementioned models employ differential equations to depict the movement of individuals among various compartments over a period of time, utilizing parameters such as transmission rate and recovery rate.

5.4.1. SIR model

The fundamentals of compartmental models, which elucidate the spread of infectious diseases, were established in a trilogy of studies published by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933. The SIR model represents a case of the model introduced by Kermack and McKendrick in 1927, which forms the foundation for our analysis of epidemic models [44]. The SIR model comprises equations that can be expressed as follows:

$$\frac{dS}{dt} = -\beta \times S \times I \quad (14)$$

$$\frac{dI}{dt} = \beta \times S \times I - \gamma \times I \quad (15)$$

$$\frac{dR}{dt} = \gamma \times I \quad (16)$$

These equations represent the number of susceptible population (S), infected individuals (I), and recovered people (R). The corresponding temporal change rates for each variable are denoted as $\frac{dS}{dt}$, $\frac{dI}{dt}$, and $\frac{dR}{dt}$. The infection rate passed from sick individuals to susceptible population is represented by β , while γ represents the recovery rate for sick individuals [45].

Effective Reproduction Number. The effective reproduction number, commonly represented as R_e , serves as a critical parameter that governs the dynamics of infectious diseases within a population, determining whether the disease will experience a rapid decline or lead to an epidemic outbreak. The relationship for observing the effective reproduction number is expressed by Equation 17 for the SIR model:

$$R_e = \frac{S(0) \times \beta}{\gamma} \quad (17)$$

where $S(0)$ represents the initial number of susceptible individuals, β denotes the transmission rate, and γ corresponds to the recovery rate.

5.4.2. SEIR

The main difference between the SIR and SEIR models is the addition of an "exposed" compartment in the SEIR model. The exposed compartment includes individuals who are infected but not yet infectious due to the pathogen's incubation period.

Advantages of SEIR over SIR. : The SEIR model offers a notable benefit over the SIR model in that it affords a more precise depiction of illnesses characterized by an incubation period, given its incorporation of the latency period. The mentioned capacity holds significant value in the modeling of COVID-19, a disease that exhibits an incubation period that can extend up to a maximum of 14 days.

Trade-off of SEIR over SIR. : The SEIR model exhibits a higher degree of complexity than the SIR model, posing challenges in terms of analysis and interpretation. The SIR model may be better suited for brief epidemics, where the assumption of a constant overall population is justifiable.

. The SEIR model is a flexible framework utilized in modeling the transmission of infectious diseases in various populations. Researchers can customize the model by adding or removing compartments and adjusting parameters to suit specific disease dynamics. For instance, a study conducted by Shaobo He and colleagues proposed a modified SEIR model for COVID-19, which incorporated infectious and quarantine classes to account for quarantine and hospitalization effects in Hubei province, China [46]. The SEIRS model expands upon the SIER model by including additional compartments and flows such as incubation period, loss of immunity, births, and deaths, allowing for a more accurate simulation of epidemic spread [47]. This flexibility makes SEIR models valuable tools for studying infectious diseases and assessing control strategies.

5.4.3. Modified SIRD model

As previously stated, the SIR compartmental models have additional compartments added as needed. The SIRD model includes a compartment for the population who have died, allowing for the prediction of the total number of fatalities from the disease. This model has been used to forecast the spread of COVID-19 in various countries. [48]. In certain instances, the utilization of the time-varying transmission rate has been employed to enhance the accuracy of COVID-19 forecasting. Specifically, the *beta* function has been expressed in either an exponential form, $\beta = \beta_0 e^{(\beta_1 \times t)}$, or a linear form, $\beta = \beta_0 + \beta_1 \times t$ [49].

. Predicting COVID-19 using complex compartmental models, such as Shaobo He's modified SEIR model, is extremely difficult due to heavy dependence on actual conditions in a specific region of interest. Additionally, available

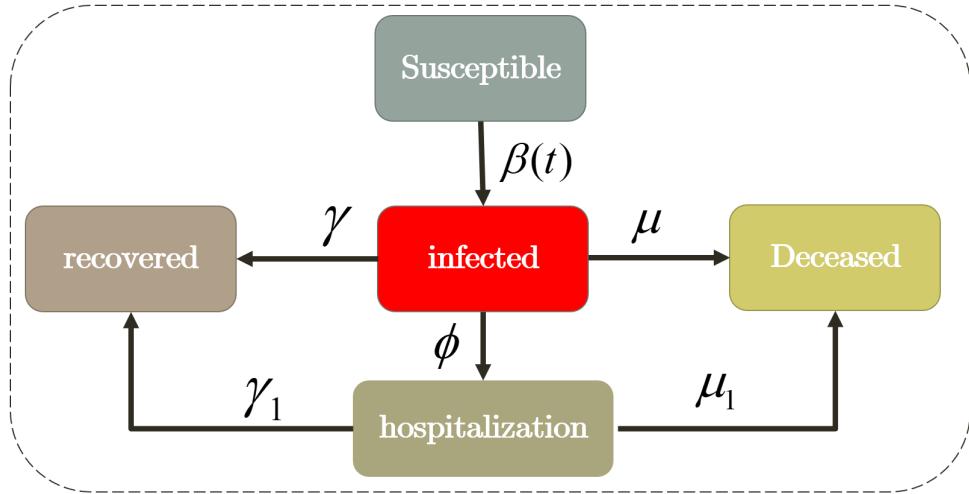


Figure 5: modified SIRD model diagram

data may be insufficient or unreliable, leading to unacceptable outcomes. To address these challenges, we simplified the modified SEIR equations to meet our needs, as demonstrated in Equation 18. We aim to estimate the number of individuals who require hospitalization. To achieve this, we have formulated equations as shown in Equation 18. It is important to note that our goal is not to make accurate predictions of hospitalizations using these equations. Rather, we aim to analyze the pressure on the healthcare system during each wave of COVID-19 by utilizing the data available.

$$\begin{aligned}
\frac{dS}{dt} &= \underbrace{-\beta(t) \frac{SI}{N}}_{\text{infection}} & (18) \\
\frac{dI}{dt} &= \underbrace{\beta(t) \frac{SI}{N}}_{\text{infection}} - \underbrace{\gamma I}_{\text{recovery from infection}} - \underbrace{\mu I}_{\text{Deceased from infection}} - \underbrace{\phi I}_{\text{hospitalization}} \\
\frac{dR}{dt} &= \underbrace{\gamma I}_{\text{recovery from infection}} + \underbrace{\gamma_1 H}_{\text{recovery from hospitalization}} \\
\frac{dD}{dt} &= \underbrace{\mu I}_{\text{Deceased from infection}} + \underbrace{\mu_1 H}_{\text{Deceased from hospitalization}} \\
\frac{dH}{dt} &= \underbrace{\phi I}_{\text{hospitalization}} - \underbrace{\gamma_1 H}_{\text{recovery from hospitalization}} - \underbrace{\mu_1 H}_{\text{Deceased from hospitalization}}
\end{aligned}$$

where $\beta(t) = \beta_0 e^{(\beta_1 t)}$

Description for modified SIRD model:. These equations include five population groups. In these equations, the variables S , I , R , D , and H represent the number of susceptible, infected, recovered, deceased, and hospitalized people, respectively.

1. $\beta(t)$ denotes the transmission rate, which is associated with the rate at which a disease spreads among individuals. To enhance the modeling process, we have regarded it as a variable with an exponential function that exhibits temporal variation.
2. γ is the recovery rate that shows how much people have recovered.
3. μ is the death rate, which shows how many people died due to the disease.
4. γ_1 represents the recovery rate of hospitalized patients. In other words, this coefficient shows how many hospitalized patients recover and are discharged from the hospital at any time.
5. μ_1 represents the mortality rate of hospitalized patients. In other words, this coefficient shows how many hospitalized patients die due to the disease at any given time.
6. ϕ represents the hospitalization rate.

5.4.4. Using the modified SIRD model for forecasting in Iran

To deploy this model in Iran, we need to develop a method for obtaining the model's parameters in Iran based on fatality and new infection data. The most acceptable technique for computing the parameters of modified SIRD model equations is to apply an optimization algorithm such as the genetic algorithm; in this method, we should utilize the appropriate cost function, which is introduced in the Equation 19.

$$MSE_{\text{total}} = \frac{2 \times D_{\text{error}} + I_{\text{error}}}{N_{\text{samples}}} + n_h \times H_{\text{error}} \quad (19)$$

$$D_{\text{error}} = \sum_{i=1}^{N_{\text{samples}}} |\text{deaths}(i) - D(i)| \quad (20)$$

$$I_{\text{error}} = \sum_{i=1}^{N_{\text{samples}}} |\text{newcases}(i) - I(i)| \quad (21)$$

$$H_{\text{error}} = \max(H) - H_{\text{Guess}} \quad (22)$$

Equation 19 consists of three components. The initial constituent, denoted as $\frac{2 \times D_{\text{error}}}{N_{\text{samples}}}$, is designed to ensure an adequate level of concordance between the predicted fatality predicted by the modified SIRD model and the observed data. The second component, $\frac{I_{\text{error}}}{N_{\text{samples}}}$, aims to make the predicted number of infections closer to reality. Unfortunately, data regarding hospitalized individuals were not readily available. Therefore, we estimated the maximum number of hospitalized individuals by examining news reports. Then, by defining $H_{\text{error}} = \max(H) - H_{\text{Guess}}$, we attempted to estimate the number of hospitalized individuals. In Equation 22, H_{Guess} represents the initial guess for the maximum number of hospitalized individuals using available news. To determine H_{Guess} , we initially make an educated guess and adjust it based on the accuracy of the predicted trend of hospitalized individuals, employing trial and error if necessary. The coefficient n_h represents the weighting factor for H_{error} in the cost function.

results. In order to further investigate and compare the prediction of this model with other models, we have made a prediction using the obtained parameters for the third wave of COVID-19 in Iran, and the results of this prediction are shown in Fig 6 to 9. Figure 6 shows the forecast number for the

number of hospitalized people. In this diagram, the number of hospitalized cases decreased until October 7. One of the reasons for this decrease is the decrease in the number of sick people in the previous wave. Another factor could be related to the delay between the trends of hospitalized and sick people. After October 7, an increase in the number of hospitalized people was seen. To evaluate, we know that the number of new people admitted to the hospital on December 10, 2020, was equal to 1448 people, and our model predicted a value of 1200 for us; Therefore, by using this model, he was able to estimate the number of hospitalized people. In addition to this, the model has been able to predict cumulative death well; however, around 12,000 people predicted the peak of new cases, whereas the actual number stands at approximately 14,000 people.

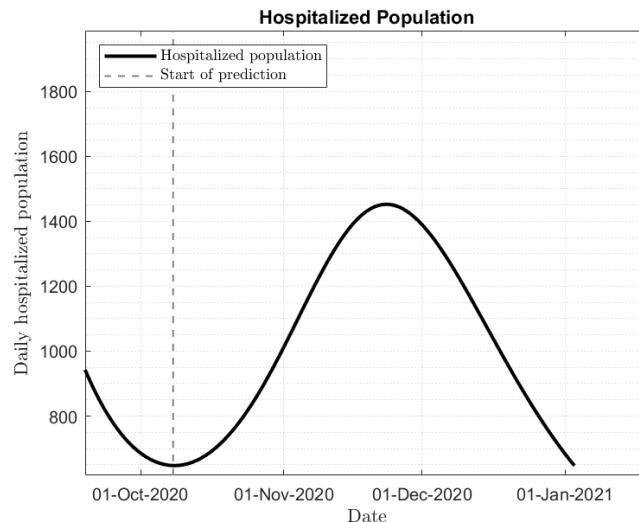


Figure 6: Prediction of the number of hospitalized people

In contrast to the logistic model, the modified susceptible-infected-recovered-deceased (SIRD) model necessitates a greater amount of data and more intricate computations. Furthermore, the model exhibits a high degree of sensitivity to parameters and initial conditions. Even minor inaccuracies in the computation of these values can result in inaccurate predictions. Moreover, the identification of the boundaries of said parameters through the utilization of the genetic algorithm presents certain challenges. In order to accurately predict the intricate dynamics of COVID-19 using the modified SIRD

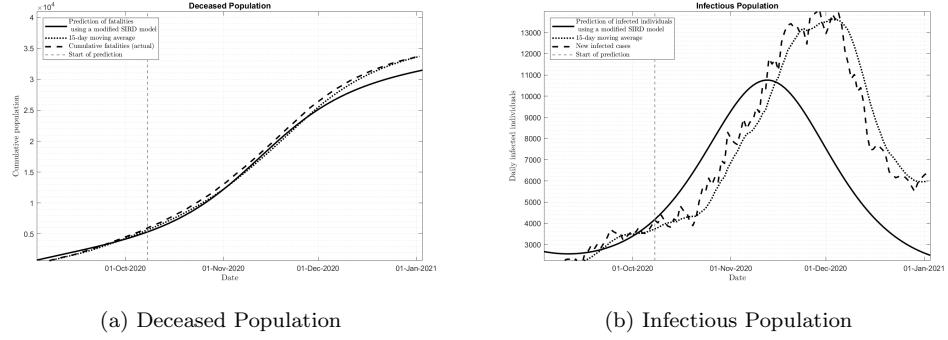


Figure 7: Prediction of the number of fatalities and new infectious cases

model, a comprehensive understanding of the necessary data and parameters is imperative. The complexity of COVID-19 dynamics poses a challenge in utilizing this model for predictive purposes. Accurate parameter determination necessitates access to information concerning prior COVID-19 waves that bear resemblance to the targeted wave, should this model be employed.

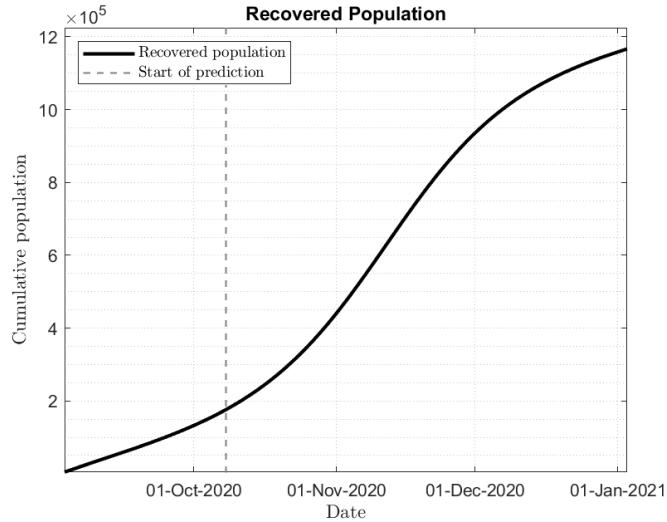


Figure 8: Recovered population

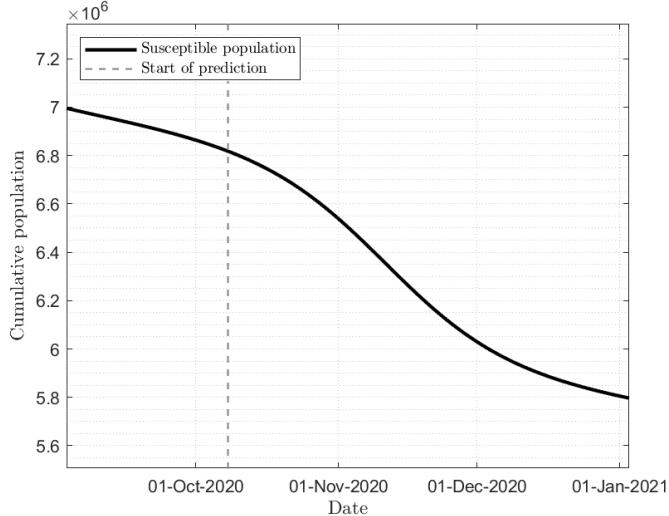


Figure 9: Susceptible population

6. predicting the number of COVID-19 deaths in Iran

Various studies, including ours, have indicated that the ARIMA model is not very effective and should only be used for predicting the next seven days. Although LSTM networks perform better than other methods over short periods, their errors tend to accumulate, making them unsuitable for predictions beyond 7-10 days. Moreover, due to the complexities of COVID-19 and other external factors, this method cannot be relied upon for long-term predictions. The modified SIRD model has the ability to estimate the number of new cases of infection, hospitalized patients, and deceased individuals. However, it is a complex model that requires different data types to be effective. In our use case, we have found that it is best to use this model to compare the waves of COVID-19 and assess the pressure on the healthcare system in Iran. The discrete logistic model can accurately predict the number of deaths if the parameters are set properly, making it a viable option for COVID-19 predictions in Iran, where data is limited.

6.1. Proposed method for predicting the number of COVID-19 deaths in Iran

After conducting an analysis, we have concluded that predicting COVID-19 fatalities is a more accurate way of measuring the virus's impact in Iran than predicting infections. We used the discrete logistic model to forecast

the fatalities. Using a genetic algorithm for parameter estimation is more straightforward and easier to implement, but determining its upper and lower bounds presents a challenge. We aim to establish a methodology for determining these boundaries to ensure accurate forecasting.

6.1.1. Boundaries of r

Our investigation shows that using the prediction of the ARIMA forecasting technique in conjunction with equation 6 produces the most correct assessment of the r value. The result of this evaluative technique is symbolized by the symbol r_{arima} . The results reported in Table 3 strongly support the association between r_{arima} and r_{ga} . After thoroughly examining the available data for determining system parameters, we determine that r_{arima} closely resembles the optimum value for the logistic curve. Instead of relying on actual data and the accompanying r value, we can instead use r_{arima} to set upper and lower bounds for r as follows:

$$r_{\text{lower band}} = 0.8 \times r_{arima} \quad (23)$$

$$r_{\text{upper band}} = 1.2 \times r_{arima} \quad (24)$$

6.1.2. Boundaries of d_0

All available data from previous waves of the outbreak were analyzed to determine the acceptable bonds for d_0 to determine the range of initial conditions for fatalities. In every study of COVID-19 waves in Iran, d_0 ranged from 50 to 230 individuals. Therefore, we can establish the highest and lowest possible prediction limits as follows:

$$d_{0_{\text{lower band}}} = 50 \quad (25)$$

$$d_{0_{\text{upper band}}} = 230 \quad (26)$$

. After careful examination, it has been noticed that predictions can be improved in some cases when $d_{0_{\text{lower band}}} = 70$. Therefore, it is suggested to consider both scenarios for predicting fatalities, one with 50 and the other with 70. Then, the scenario that delivers better predictive results should be chosen as $d_{0_{\text{upper band}}}$.

6.1.3. Boundaries of M'

Estimating the number of death cases using the discrete logistic model presents several challenges. However, the main obstacle is determining the value of M' for each wave. Since we are utilizing a genetic algorithm, it is necessary to establish the upper and lower bounds using a methodology.

. In Figure 3, we can see how the behavior of the logistic curve is affected by the number of data points used for fitting it. Adding an additional week's worth of data can improve our ability to make predictions and achieve a more accurate projection compared to not utilizing this data for forecasting. As a result, if we analyze data patterns and identify changes in M' , we can get a better estimate of the upper and lower limits.

There are multiple ways to analyze data trends and estimate M' . One highly effective method is utilizing machine learning techniques, particularly neural networks, to predict the trend of M' by adding new data to the dataset and performing parameter estimation. By analyzing past data, we can create a model that can forecast the trajectory of M' and then use this method to determine its upper and lower limits.

To get a more accurate count of fatalities, we can use a combination of a discrete logistic model, genetic algorithm, and machine learning techniques. This study aims to predict changes in M' using a two-layer MLP and establish the boundaries for M' for the genetic algorithm. This approach can help overcome existing challenges more efficiently.

selection of input features for MLP. In order to predict how M' will change, we need to take into account the input features used for the neural network. These variables can be divided into two categories.

The first category provides information on the transmission rates of various COVID-19 strains in comparison to the original variants during a particular study period. To access this data, please refer to Table 1. This category also includes an evaluation of the efficacy of COVID-19 vaccines in Iran, as well as climate data related to urban areas and the Iranian population during the same study period.

The second set of input features consists of various model predictions or parameters. The initial features in this category are denoted as r and M' , derived from reviewing previous waves, serving as the initial guesses for these values. The next variable is r_{arima} Using the SIRDH model, we made

predictions for COVID-19 in Iran. We calculated the average populations predicted by the model, represented as S_{mean} , I_{mean} , D_{mean} , H_{mean} , and R_{mean} , which serve as inputs to the neural network.

We use the Bi-LSTM model to predict the next seven days and include these predictions in the input feature. The final set of input features consists of predictions for the next 10 days, which are made using initial parameter estimates for the logistic model and actual data of deceased individuals from 5 days prior to the prediction.

Calculating Training Data. A sequential approach is used to gather the necessary training data to train a neural network. This involves separating data from previous waves from the current wave and performing distinct actions for each set.

Calculating Data for Previous Waves. In order to train our network, we gather relevant data on transmission rates, vaccine efficacy, and meteorological factors for each wave. We also choose a forecast day for each wave, as specified in Table 3. Using data from the first wave up until the selected forecast day, we calculate a second type of input variable for each wave based on this initial data.

Calculating for Current Wave . To generate the training data, we utilize the most recent wave data. Our process involves the following steps to calculate the different types of input features:

1. We start by collecting data for the first type of input features, which will be used for training purposes.
2. Next, we determine the second type of input features by following these steps:
 - (a) choosing the main forecast day corresponds to the actual forecasting day.
 - (b) separates a specific set of data from the beginning of the COVID-19 wave up until 10 days before the main forecast day.
 - (c) Calculate the second type of input features using this selected data.
 - (d) We repeat steps b and c, but now with a new set of days that include data from the first day of the wave up to 9 days prior to the main forecast day.

- (e) This iterative process continues by incorporating data from nine days prior, followed by eight days prior, and so on until we reach the forecast day.

Following these steps, we can effectively gather the necessary training data based on the current wave information.

the process of obtaining the upper and lower bounds of M' .

In summary, the process of obtaining the upper and lower bounds of M' is described as follows:

1. In the first step, we need to calculate and collect the input features of previous waves according to the method described in section 6.1.3.
2. Then, using the obtained parameters for the logistic model in previous waves, we make an initial guess for the bounds of the genetic algorithm. We set the upper bound to $1.5 \times M'_{\text{guess}}$ and the lower bound to $0.5 \times M'_{\text{guess}}$.
3. Next, using the guessed bounds and the current waveform data, we calculate the second-type variable according to the method mentioned in section 6.1.3 and collect the first-type features.
4. Next, we proceed with training the neural network mentioned earlier using the data provided. Once the training is complete, we then calculate the network's output for the Main prediction day, which is represented as M'_{nn} .
5. Finally, using M'_{nn} and equations 27 and 28, we determine the upper and lower bounds and make the final prediction.

$$M'_{\text{lower band}} = 0.9 \times M'_{nn} \quad (27)$$

$$M'_{\text{upper band}} = 1.2 \times M'_{nn} \quad (28)$$

We utilize the upper and lower bounds for the perimeter to predict mortality rates in Iran by employing a genetic algorithm and logistic model.

6.2. Simplified method for predicting the number of COVID-19 deaths in Iran

In the previous method for predicting COVID-19 fatalities, there were challenges associated with estimating many parameters and using multiple models simultaneously. However, this section has introduced an easier

method to facilitate COVID-19 prediction. By reducing the number of parameters and utilizing simpler techniques, this new method offers greater flexibility and ease in predicting COVID-19 fatalities.

This method is similar to the previous approach in that it utilizes a logistic model. However, it introduces a simpler approach to determine the upper and lower bounds of the parameters r , d_0 , and M' . In order to improve predictions for COVID-19 fatalities, it is important to predict the trend of M' by incorporating new data into the data set and performing parameter estimation based on initial estimates. Unlike the previous approach, which used a neural network to combine different models, this simpler method employs a linear approach and reduces the number of variables for the sake of simplicity in order to determine the bounds of M' , this method also necessitates determining the bounds of r and d_0 .

6.2.1. Determining the Upper and Lower Bounds of r and d_0

We use data from COVID-19 fatalities to determine the upper and lower bounds. Equation 6 is used to calculate the value of r for the available data prior to the main prediction day, and the average is taken, denoting the calculated value as r_{data} . For the upper bound, we consider $r_{\text{upper}} = 1.2 \times r_{\text{data}}$, and for the lower bound, $r_{\text{lower}} = 0.9 \times r_{\text{data}}$. Additionally, the value of d_0 is based on two constraints: $50 < d_0 < 230$ or $70 < d_0 < 230$, as stated earlier. Predictions are made for both cases, and the scenario that provides a better prediction is chosen based on the results.

6.2.2. Process of determining the Upper and Lower Bounds of M' and COVID-19 Prediction

In this method, similar to the previous approach, we start with an initial estimate for the value of M' and attempt to find the trend of changes by adding data.

1. Start by initializing an initial approximation for the value of M' through a search for a waveform that closely resembles the current wave. Utilize the parameters from this waveform (refer to Table 3) to estimate the value of M' . This estimated value can be represented as M'_{guess} .
2. To improve M'_{guess} , divide the data set into segments as follow:
 - (a) Select the main prediction day that aligns with the actual prediction day.
 - (b) Separate a specific set of data from the beginning of the COVID-19 wave up to 10 days before the main prediction day.

- (c) Repeat step (b), each time with a new set of days, including data from the first day of the wave up to 9 days before the main prediction day. This process continues successively, using data from 9 days before, then 8 days before, and so on, until reaching the main prediction day.

note: for each segments dataset, the genetic algorithm is employed to determine the value of M' , with the upper limit set at $1.5 \times M'_{\text{guess}}$ and the lower limit set at $0.5 \times M'_{\text{guess}}$. It's important to mention that in this section, the boundaries for r are defined as $1.5 \times r_{\text{data}} < r < 0.5 \times r_{\text{data}}$, and the constraints for d_0 are established based on Section 6.2.1.

- 3. Fit a line to the values of M' that were obtained from step 2 and the prediction day. This fitted line will be used to calculate M' on the main prediction day, which we'll denote as M'_{line} . Set the bounds for this calculation as follows

$$M'_{\text{lower bound}} = 0.9 \times M'_{\text{line}} \quad (29)$$

$$M'_{\text{upper bound}} = 1.2 \times M'_{\text{line}} \quad (30)$$

To predict COVID-19 in Iran, utilize these boundaries and a genetic algorithm to determine the parameters of a logistic model.

7. Results

In order to comprehensively evaluate our methods and determine their effectiveness, we conducted an analysis of the data from Iran regarding the seventh and eighth occurrences of COVID-19.

7.1. Results of the Proposed Method (Neural Network) for Predicting COVID-19 Fatalities in Iran

We used the first prediction method and available data to estimate the number of COVID-19 patients and fatalities in Iran for the seventh wave. Initially, we set the M' value to 0.0011. Next, we calculated the value of r_{arima} for this wave to be 1.346. Using these values and the mentioned methods, we obtained the value of M'_{nn} as 1.5359×10^{-3} . Finally, we used the genetic algorithm to determine the final value of M' for the seventh wave, which was 1.6834×10^{-3} .

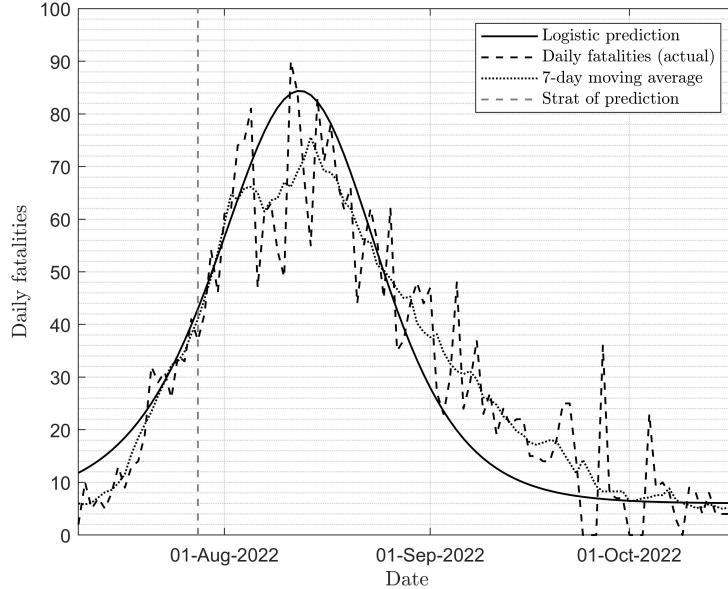


Figure 10: Prediction of the seventh wave using the proposed method (Neural Network)

. The seventh wave of COVID-19 in Iran was predicted utilizing our method, as illustrated in Figure 10. The prediction made by our model exhibited a high degree of concordance with the moving average, aligning closely until August second. Notably, it accurately projected the number of fatalities, successfully capturing the peak fatality count. Despite encountering discrepancies in specific sections of the logistic curve, we maintain that our overall prediction exhibited satisfactory performance.

. To predict the eighth wave, we used similar methods and available data. Initially, we set M' to be 0.0025 and determined r_{arima} to be 1.143. We used a neural network to calculate M'_{nn} , which resulted in 1.644×10^{-3} . Finally, we found that the final value of M' for the eighth wave was 1.4933×10^{-3} . Upon conducting an analysis of Figure 11, it was observed that our prediction exhibited a congruence with the moving average until April 5th. Nevertheless, due to an abrupt surge in the data, our model lacked precision in accurately forecasting the highest fatalities. The number of fatalities was 42, whereas our projected estimate was approximately 27. However, this prediction was regarded as being acceptable.

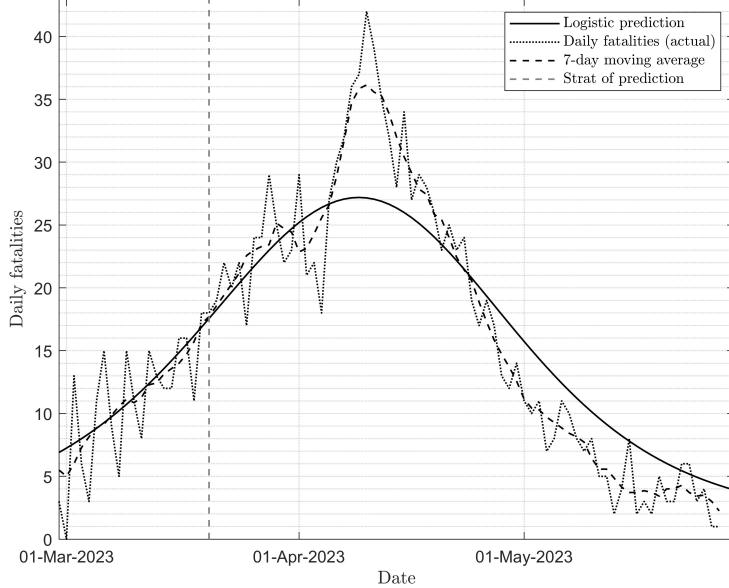


Figure 11: Prediction of the eighth wave using the proposed method (Neural Network)

7.2. The results of the simplified method for predicting the number of COVID-19 deaths in Iran

We used the seventh wave data in Iran to predict the number of fatalities and our initial estimated M' to be 0.0011. Following this, we created multiple data sets (as discussed in the previous section) and calculated the value of M' using these sets, then fitted a line based on these calculated values. In Figure 12, the outcome of this procedure is illustrated. The label in the figure indicates which data was utilized to compute the value of M' . For instance, if "July 26" is the label, it implies that the data from the start of the wave up until July 26 was employed. Through this process, we determined that the value of M'_{line} would be 0.00171 on the main prediction day. In table 3, the value of r_{data} is indicated by r . Figure 13 shows the prediction outcome made using this method. Although our method did not accurately replicate the smoothed data like the previous prediction method, it did correctly predict the wave peak. While this method may not yield better results than the previous one, the result is still a satisfactory prediction. Additionally, this method is user-friendly and requires less knowledge about COVID-19.

To forecast the number of fatalities in Iran during the eighth wave, we first estimated the value of M' to be 0.0025 and then proceeded to calculate

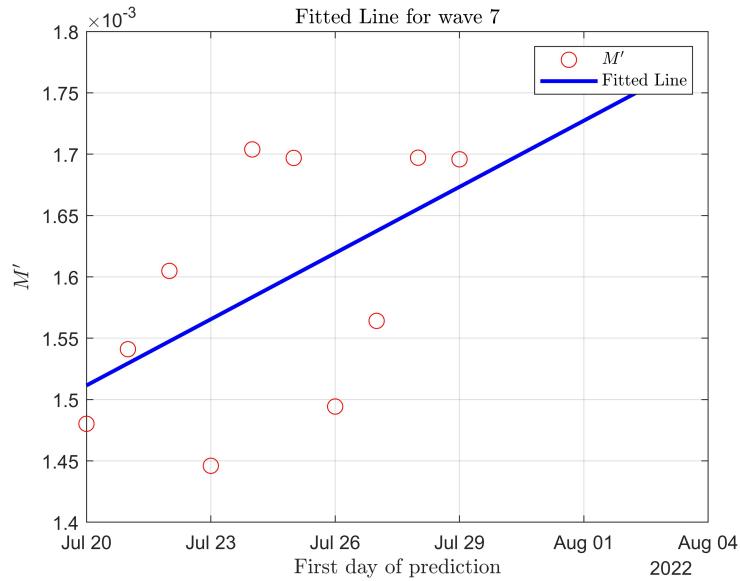


Figure 12: The trend of M' based on the day's prediction for the seventh wave

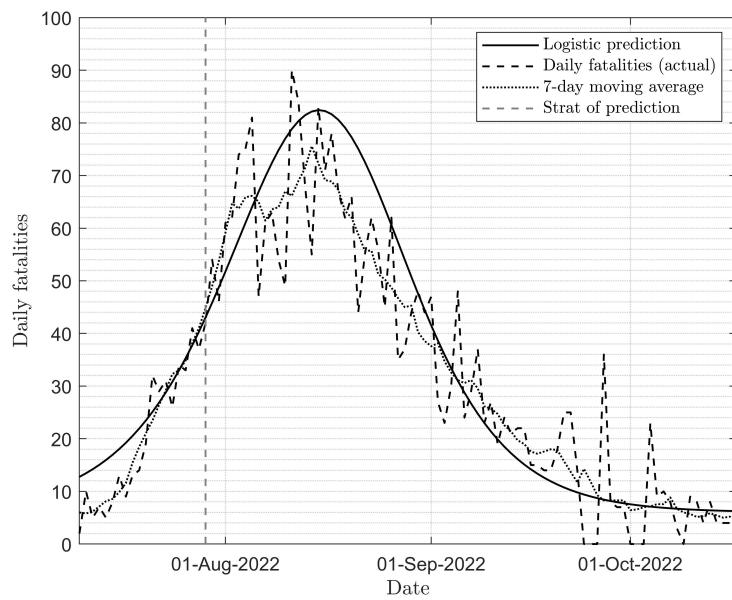


Figure 13: Prediction of the seventh wave using the simplified method

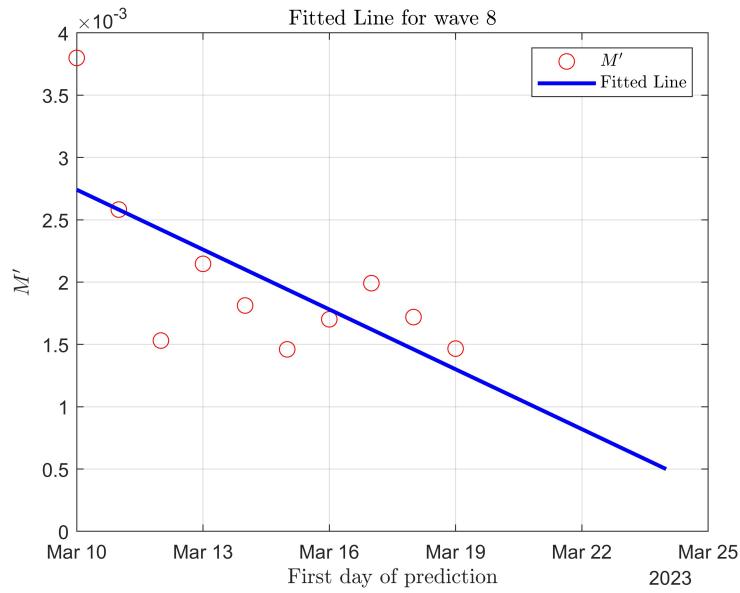


Figure 14: The trend of M' based on the day's forecast for the eighth wave

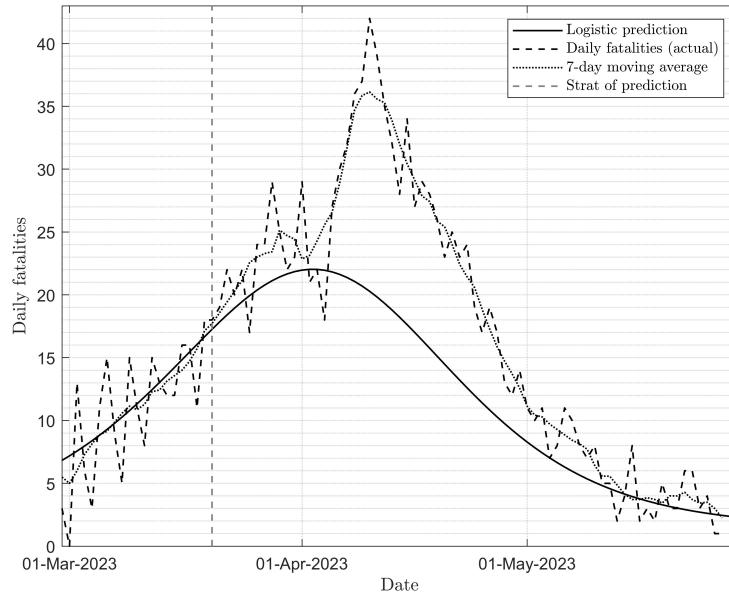


Figure 15: Eighth wave prediction using the simplified method

the value of M' for each data set. By using these data points, we were able to create a best-fit line (refer to Figure 14) and calculate M'_{line} as 0.0011 for the primary prediction date. The prediction process utilized the r value specified in Table 3. Our predictive trend remains valid within a broader context despite not accurately predicting the wave's peak, as evident in Figure 15. The discrepancy could potentially be attributed to abrupt data discontinuity. Nevertheless, this approach, while not surpassing the effectiveness of the previous method, still yields an acceptable forecast.

7.3. Comparison of Results from These Two Methods

The genetic algorithm is employed to determine the logistic model parameters in both methods. You can find the parameter values in Table 5 and Table 6. Furthermore, the RMSE (Root Mean Square Error) and R (Correlation Coefficient) values for forecasting wave 7 and wave 8 using two different techniques are presented in Table 7 and Table 8, respectively.

Table 5: Logistic Curve Parameters Using the Proposed Method (Neural Network)

Wave	M'	r	d_0
Wave 7	1.6834×10^{-3}	1.1230	54.0039
Wave 8	1.4933×10^{-3}	1.0722	80.8119

Table 6: Logistic Curve Parameters Using the Simplified Method

Wave	M'	r	d_0
Wave 8	9.3098×10^{-4}	1.0806	72.457
Wave 7	2.0399×10^{-3}	1.1080	70.010

. Table 7 shows the results of using the neural network method to predict waves 8 and 7 for different time periods. For the eighth wave, the root mean square error (RMSE) values were 2.732, 2.5972, and 4.1925 when predicting for 7 days, 10 days, and 30 days respectively. During the 10-day prediction period, there was a robust positive correlation between the predicted and actual values, with a correlation coefficient value of 0.95998. As for wave 7, the corresponding root mean square error (RMSE) values were 8.4164, 7.3146, and 12.238, while the correlation coefficient value was determined to be 0.93448, indicating a strong positive correlation.

On the other hand, the simplified approach for predicting 10 days and 30 days resulted in higher RMSE values of 3.2864, 3.5, and 13.2793, 11.6608, respectively, with corresponding correlation coefficient values of 0.94757 and 0.75685 for waves 8 and 7. It is evident that the neural network method yields better RMSE and correlation coefficient values compared to the simplified method. The neural network method consistently achieves lower RMSE values, indicating better accuracy in predicting actual values. Additionally, the neural network method's correlation coefficient values are consistently higher, indicating a stronger correlation between predicted and actual values.

Overall, the results demonstrate that the neural network method is a more reliable approach for predicting waves 8 and 7. The method's ability to achieve high accuracy and strong correlation values makes it a valuable tool for predicting ocean wave behavior.

Table 7: RMSE and R (Correlation Coefficient) Values for Wave 7 and Wave 8 Using the Proposed Method (Neural Network)

Prediction Period Length	RMSE (Root Mean Square Error)			R (Correlation Coefficient)
	7 days	10 days	30 days	
Wave 8	2.732	2.5972	4.1925	0.95998
Wave 7	8.4164	7.3146	12.238	0.93448

Calculations are based on moving averages of raw data with a window size of seven days.

Table 8: RMSE and R (Correlation Coefficient) Values for Wave 7 and Wave 8 Using the Simplified Method

Prediction Period Length	RMSE (Root Mean Square Error)			R (Correlation Coefficient)
	7 days	10 days	30 days	
Wave 8	3.2864	3.4878	9.3565	0.94757
Wave 7	13.2793	11.6534	13.6608	0.75685

Calculations are based on moving averages of raw data with a window size of seven days.

8. Conclusion

Our proposed methods have proven to be effective in predicting COVID-19-related deaths in Iran. While they can predict one epidemic wave, they cannot predict multiple waves. Using models like ours can assist policymakers in understanding the impact of COVID-19 on the community and making informed decisions.

The neural network method provides higher accuracy and correlation in predicting actual values than the simplified method. Consistently lower RMSE values and higher correlation coefficient values were achieved through the neural network, indicating better accuracy and stronger correlation in predicting actual values.

Further analysis and evaluation are required to determine the statistical significance of these differences and evaluate the overall reliability and stability of both methods. The simplified method is advantageous in ease of use and less complexity, requiring less expertise than the neural network.

Overall, the neural network approach can be a suitable option for predicting COVID-19-related deaths, considering its higher accuracy and correlation in predicting actual values, while the simplified method is easier to use. However, implementing this method in other countries and examining the obtained results would be beneficial for a more comprehensive evaluation.

Statement: During the preparation of this work, the author(s) used Grammarly go in order to improve writing. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

- [1] A. Yazdani, S. K. Bigdeli, M. Zahmatkeshan, Investigating the performance of machine learning algorithms in predicting the survival of COVID-19 patients: A cross section study of iran, *Health Science Reports* 6 (2023). URL: <https://doi.org/10.1002/hsr2.1212>. doi:10.1002/hsr2.1212.
- [2] M. Yajada, M. K. Moridani, S. Rasouli, Mathematical model to predict COVID-19 mortality rate, *Infectious Disease Modelling* 7 (2022) 761–776. URL: <https://doi.org/10.1016/j.idm.2022.11.005>. doi:10.1016/j.idm.2022.11.005.
- [3] E. Badfar, E. J. Zaferani, A. Nikoofard, Design a robust sliding mode controller based on the state and parameter estimation for the nonlinear epidemiological model of covid-19, *Nonlinear Dynamics* 109 (2021) 5–18. URL: <https://doi.org/10.1007/s11071-021-07036-4>. doi:10.1007/s11071-021-07036-4.
- [4] M. Medvedeva, T. E. Simos, C. Tsitouras, V. Katsikis, Direct estimation of SIR model parameters through second-order finite differences, *Mathematical Methods in the Applied Sciences* 44 (2020) 3819–3826. URL: <https://doi.org/10.1002/mma.6985>. doi:10.1002/mma.6985.
- [5] A. Godio, F. Pace, A. Vergnano, SEIR modeling of the italian epidemic of SARS-CoV-2 using computational swarm intelligence, *International Journal of Environmental Research and Public Health* 17 (2020) 3535. URL: <https://doi.org/10.3390/ijerph17103535>. doi:10.3390/ijerph17103535.
- [6] I. Cooper, A. Mondal, C. G. Antonopoulos, A SIR model assumption for the spread of COVID-19 in different communities, *Chaos, Solitons & Fractals* 139 (2020) 110057. URL: <https://doi.org/10.1016/j.chaos.2020.110057>. doi:10.1016/j.chaos.2020.110057.

- [7] S. Moein, N. Nickaeen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javanmard, J. Ghaisari, Y. Gheisari, Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of isfahan, *Scientific Reports* 11 (2021). URL: <https://doi.org/10.1038/s41598-021-84055-6>. doi:10.1038/s41598-021-84055-6.
- [8] R. Chandra, A. Jain, D. S. Chauhan, Deep learning via LSTM models for COVID-19 infection forecasting in india, *PLOS ONE* 17 (2022) e0262708. URL: <https://doi.org/10.1371/journal.pone.0262708>. doi:10.1371/journal.pone.0262708.
- [9] V. K. R. Chimmula, L. Zhang, Time series forecasting of COVID-19 transmission in canada using LSTM networks, *Chaos, Solitons, and Fractals* 135 (2020) 109864. URL: <https://doi.org/10.1016/j.chaos.2020.109864>. doi:10.1016/j.chaos.2020.109864.
- [10] F. Shahid, A. Zameer, M. Muneeb, Predictions for COVID-19 with deep learning models of LSTM, GRU and bi-LSTM, *Chaos, Solitons & Fractals* 140 (2020) 110212. URL: <https://doi.org/10.1016/j.chaos.2020.110212>. doi:10.1016/j.chaos.2020.110212.
- [11] N. Absar, N. Uddin, M. U. Khandaker, H. Ullah, The efficacy of deep learning based LSTM model in forecasting the outbreak of contagious diseases, *Infectious Disease Modelling* 7 (2022) 170–183. URL: <https://doi.org/10.1016/j.idm.2021.12.005>. doi:10.1016/j.idm.2021.12.005.
- [12] R. Kafieh, R. Arian, N. Saeedizadeh, Z. Amini, N. D. Serej, S. Minaee, S. K. Yadav, A. Vaezi, N. Rezaei, S. H. Javanmard, COVID-19 in iran: Forecasting pandemic using deep learning, *Computational and Mathematical Methods in Medicine* 2021 (2021) 1–16. URL: <https://doi.org/10.1155/2021/6927985>. doi:10.1155/2021/6927985.
- [13] Y. Liu, J. Rocklöv, The effective reproductive number of the omicron variant of SARS-CoV-2 is several times relative to delta, *Journal of Travel Medicine* 29 (2022). URL: <https://doi.org/10.1093/jtm/taac037>. doi:10.1093/jtm/taac037.
- [14] K. KATELLA, Omicron, delta, alpha, and more: What to know about the coronavirus, 2023. URL: <https://www.yalemedicine.org/news/covid-19-variants-of-concern-omicron>.

- [15] S. Siddiqui, H. W. S. Alhamdi, H. A. Alghamdi, Recent chronology of COVID-19 pandemic, *Frontiers in Public Health* 10 (2022). URL: <https://www.frontiersin.org/articles/10.3389/fpubh.2022.778037>.
- [16] S. Shiehzadegan, N. Alaghemand, M. Fox, V. Venketaraman, Analysis of the delta variant b.1.617.2 COVID-19, *Clinics and Practice* 11 (2021) 778–784. URL: <https://doi.org/10.3390/clinpract11040093>. doi:10.3390/clinpract11040093.
- [17] C. Danner, What we know about the new p.1 strain of the coronavirus, 2021. URL: <https://nymag.com/intelligencer/article/what-we-know-about-the-p1-variant-of-the-coronavirus.html>.
- [18] K. Linka, M. Peirlinck, A. Schäfer, O. Z. Tikenogullari, A. Goriely, E. Kuhl, Effects of b.1.1.7 and b.1.351 on COVID-19 dynamics: A campus reopening study, *Archives of Computational Methods in Engineering* 28 (2021) 4225–4236. URL: <https://doi.org/10.1007/s11831-021-09638-y>. doi:10.1007/s11831-021-09638-y.
- [19] M. Cascella, M. Rajnik, A. Aleem, et al., Features, evaluation, and treatment of coronavirus (covid-19), 2022. URL: <https://www.ncbi.nlm.nih.gov/books/NBK554776/>.
- [20] E. Volz, , S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O'Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen, R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, N. M. Ferguson, Assessing transmissibility of SARS-CoV-2 lineage b.1.1.7 in england, *Nature* 593 (2021) 266–269. URL: <https://doi.org/10.1038/s41586-021-03470-x>. doi:10.1038/s41586-021-03470-x.
- [21] WHO, Iran (islamic republic of): WHO coronavirus disease (COVID-19) dashboard with vaccination data, 2023. URL: <https://covid19.who.int>.

- [22] M. D. Knoll, C. Wonodi, Oxford–AstraZeneca COVID-19 vaccine efficacy, *The Lancet* 397 (2021) 72–74. URL: [https://doi.org/10.1016/s0140-6736\(20\)32623-4](https://doi.org/10.1016/s0140-6736(20)32623-4). doi:10.1016/s0140-6736(20)32623-4.
- [23] V. Shinde, S. Bhikha, Z. Hoosain, M. Archary, Q. Bhorat, L. Fairlie, U. Laloo, M. S. Masilela, D. Moodley, S. Hanley, L. Fouche, C. Louw, M. Tameris, N. Singh, A. Goga, K. Dheda, C. Grobbelaar, G. Kruger, N. Carrim-Ganey, V. Baillie, T. de Oliveira, A. L. Koen, J. J. Lombaard, R. Mngqibisa, A. E. Bhorat, G. Benadé, N. Laloo, A. Pitsi, P.-L. Vollgraaff, A. Luabeya, A. Esmail, F. G. Petrick, A. Oommen-Jose, S. Foulkes, K. Ahmed, A. Thombrayil, L. Fries, S. Cloney-Clark, M. Zhu, C. Bennett, G. Albert, E. Faust, J. S. Plested, A. Robertson, S. Neal, I. Cho, G. M. Glenn, F. Dubovsky, S. A. Madhi, Efficacy of NVX-CoV2373 covid-19 vaccine against the b.1.351 variant, *New England Journal of Medicine* 384 (2021) 1899–1909. URL: <https://doi.org/10.1056/nejmoa2103055>. doi:10.1056/nejmoa2103055.
- [24] N. Andrews, J. Stowe, F. Kirsebom, S. Toffa, T. Rickeard, E. Gallagher, C. Gower, M. Kall, N. Groves, A.-M. O’Connell, D. Simons, P. B. Blomquist, A. Zaidi, S. Nash, N. I. B. A. Aziz, S. Thelwall, G. Dabrera, R. Myers, G. Amirthalingam, S. Gharbia, J. C. Barrett, R. Elson, S. N. Ladhani, N. Ferguson, M. Zambon, C. N. Campbell, K. Brown, S. Hopkins, M. Chand, M. Ramsay, J. L. Bernal, Covid-19 vaccine effectiveness against the omicron (b.1.1.529) variant, *New England Journal of Medicine* 386 (2022) 1532–1546. URL: <https://doi.org/10.1056/nejmoa2119451>. doi:10.1056/nejmoa2119451.
- [25] S. Xia, Y. Zhang, Y. Wang, H. Wang, Y. Yang, G. F. Gao, W. Tan, G. Wu, M. Xu, Z. Lou, W. Huang, W. Xu, B. Huang, H. Wang, W. Wang, W. Zhang, N. Li, Z. Xie, L. Ding, W. You, Y. Zhao, X. Yang, Y. Liu, Q. Wang, L. Huang, Y. Yang, G. Xu, B. Luo, W. Wang, P. Liu, W. Guo, X. Yang, Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBIBP-CorV: a randomised, double-blind, placebo-controlled, phase 1/2 trial, *The Lancet Infectious Diseases* 21 (2021) 39–51. URL: [https://doi.org/10.1016/s1473-3099\(20\)30831-8](https://doi.org/10.1016/s1473-3099(20)30831-8). doi:10.1016/s1473-3099(20)30831-8.
- [26] J. Venkadapathi, V. K. Govindarajan, S. Sekaran, S. Venkatapathy, A minireview of the promising drugs and vaccines in pipeline for the

- treatment of COVID-19 and current update on clinical trials, *Frontiers in Molecular Biosciences* 8 (2021). URL: <https://doi.org/10.3389/fmolb.2021.637378>. doi:10.3389/fmolb.2021.637378.
- [27] M. Mousa, M. Albreiki, F. Alshehhi, S. AlShamsi, N. A. Marzouqi, T. Alawadi, H. Alrand, H. Alsafar, A. Fikri, Similar effectiveness of the inactivated vaccine BBIBP-CorV (sinopharm) and the mRNA vaccine BNT162b2 (pfizer-BioNTech) against COVID-19 related hospitalizations during the delta outbreak in the UAE, *Journal of Travel Medicine* 29 (2022). URL: <https://doi.org/10.1093/jtm/taac036>. doi:10.1093/jtm/taac036.
 - [28] Z. Huang, S. Xu, J. Liu, L. Wu, J. Qiu, N. Wang, J. Ren, Z. Li, X. Guo, F. Tao, J. Chen, D. Lu, X. Sun, W. Wang, Effectiveness of inactivated and ad5-nCoV COVID-19 vaccines against SARS-CoV-2 omicron BA. 2 variant infection, severe illness, and death, *BMC Medicine* 20 (2022). URL: <https://doi.org/10.1186/s12916-022-02606-8>. doi:10.1186/s12916-022-02606-8.
 - [29] Á. Briz-Redón, Á. Serrano-Aroca, The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modelling techniques, *Progress in Physical Geography: Earth and Environment* 44 (2020) 591–604. URL: <https://doi.org/10.1177/0309133320946302>. doi:10.1177/0309133320946302.
 - [30] M. Ahmadi, A. Sharifi, S. Dorosti, S. J. Ghoushchi, N. Ghanbari, Investigation of effective climatology parameters on COVID-19 outbreak in iran, *Science of The Total Environment* 729 (2020) 138705. URL: <https://doi.org/10.1016/j.scitotenv.2020.138705>. doi:10.1016/j.scitotenv.2020.138705.
 - [31] E. Mathieu, H. Ritchie, L. Rods-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, M. Roser, Coronavirus pandemic (COVID-19), *Our World in Data* (2023). URL: <https://ourworldindata.org/coronavirus>.
 - [32] ISNA, ISNA news agency, ISNA (2023). URL: <https://en.isna.ir>, publisher: ISNA News Agency.

- [33] X.-S. Yang, Nature-Inspired Optimization Algorithms, Elsevier, 2014.
URL: <https://www.sciencedirect.com/science/article/pii/B9780124167438000166>. doi:10.1016/B978-0-12-416743-8.00016-6.
- [34] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan, F. S. Al-Anzi, On the accuracy of ARIMA based prediction of COVID-19 spread, Results in Physics 27 (2021) 104509.
URL: <https://doi.org/10.1016/j.rinp.2021.104509>. doi:10.1016/j.rinp.2021.104509.
- [35] G. Ding, X. Li, F. Jiao, Y. Shen, Brief analysis of the ARIMA model on the COVID-19 in italy, (2020). URL: <https://doi.org/10.1101/2020.04.08.20058636>. doi:10.1101/2020.04.08.20058636.
- [36] A. K. Sahai, N. Rath, V. Sood, M. P. Singh, ARIMA modelling & forecasting of COVID-19 in top five affected countries, Diabetes & Metabolic Syndrome: Clinical Research & Reviews 14 (2020) 1419–1427. URL: <https://doi.org/10.1016/j.dsx.2020.07.042>. doi:10.1016/j.dsx.2020.07.042.
- [37] T. Tran, L. Pham, Q. Ngo, Forecasting epidemic spread of sars-cov-2 using arima model (case study: Iran), Global Journal of Environmental Science and Management 6 (2020). URL: <https://doi.org/10.22034/GJESM.2019.06.SI.01>. doi:10.22034/GJESM.2019.06.SI.01.
- [38] F. M. Khan, R. Gupta, ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in india, Journal of Safety Science and Resilience 1 (2020) 12–18. URL: <https://doi.org/10.1016/j.jnlssr.2020.06.007>. doi:10.1016/j.jnlssr.2020.06.007.
- [39] P. Wang, X. Zheng, J. Li, B. Zhu, Prediction of epidemic trends in COVID-19 with logistic model and machine learning techniques, Chaos, Solitons & Fractals 139 (2020) 110058. URL: <https://doi.org/10.1016/j.chaos.2020.110058>. doi:10.1016/j.chaos.2020.110058.
- [40] X. Zhou, X. Ma, N. Hong, L. Su, Y. Ma, J. He, H. Jiang, C. Liu, G. Shan, W. Zhu, S. Zhang, Y. Long, Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model, medrxiv (2020). URL: <https://doi.org/10.1101/2020.03.26.20044289>. doi:10.1101/2020.03.26.20044289.

- [41] Y. Yu, X. Si, C. Hu, J. Zhang, A review of recurrent neural networks: LSTM cells and network architectures, *Neural Computation* 31 (2019) 1235–1270. URL: https://doi.org/10.1162/neco_a_01199. doi:10.1162/neco_a_01199.
- [42] A. Zhang, Z. C. Lipton, M. Li, A. J. Smola, Dive into deep learning, arXiv preprint arXiv:2106.11342 (2021).
- [43] S. Chen, R. Paul, D. Janies, K. Murphy, T. Feng, J.-C. Thill, Exploring feasibility of multivariate deep learning models in predicting COVID-19 epidemic, *Frontiers in Public Health* 9 (2021). URL: <https://doi.org/10.3389/fpubh.2021.661615>. doi:10.3389/fpubh.2021.661615.
- [44] F. Brauer, P. van den Driessche, J. Wu (Eds.), *Mathematical Epidemiology*, Springer Berlin Heidelberg, 2008. URL: <https://doi.org/10.1007/978-3-540-78911-6>. doi:10.1007/978-3-540-78911-6.
- [45] P. Zhang, K. Feng, Y. Gong, J. Lee, S. Lomonaco, L. Zhao, Usage of compartmental models in predicting COVID-19 outbreaks, *The AAPs Journal* 24 (2022). URL: <https://doi.org/10.1208/s12248-022-00743-9>. doi:10.1208/s12248-022-00743-9.
- [46] S. He, Y. Peng, K. Sun, SEIR modeling of the COVID-19 and its dynamics, *Nonlinear Dynamics* 101 (2020) 1667–1680. URL: <https://doi.org/10.1007/s11071-020-05743-y>. doi:10.1007/s11071-020-05743-y.
- [47] O. N. Bjørnstad, K. Shea, M. Krzywinski, N. Altman, The SEIRS model for infectious disease dynamics, *Nature Methods* 17 (2020) 557–558. URL: <https://doi.org/10.1038/s41592-020-0856-2>. doi:10.1038/s41592-020-0856-2.
- [48] J. Fernández-Villaverde, C. I. Jones, Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities, *Journal of Economic Dynamics and Control* 140 (2022) 104318. URL: <https://doi.org/10.1016/j.jedc.2022.104318>. doi:10.1016/j.jedc.2022.104318.
- [49] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, D. Liu, Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model, *IEEE Access* 8

(2020) 51761–51769. URL: <https://doi.org/10.1109/access.2020.2979599>. doi:10.1109/access.2020.2979599.