A comparison of Covid-19 cases and deaths in Turkey and in other countries

Oğuzhan Çağlar1 · Figen Özen1

Received: 2 August 2022 / Revised: 28 September 2022 / Accepted: 11 October 2022 / Published online: 27 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract
In this study, the characteristics of the Covid-19 pandemic in Turkey are examined in terms of the number of cases and deaths, and a characteristic prediction is made with an approach that employs artificial intelligence. The number of cases and deaths are estimated using the number of tests, the numbers of seriously ill and recovered patients as parameters. The machine learning methods used are linear regression, polynomial regression, support vector regression with different kernel functions, decision tree and artificial neural networks. The obtained results are compared by calculating the coefficient of determination ($R^2$), and the mean absolute percentage error (MAPE) values. When $R^2$ and MAPE values are compared, it is seen that the optimal results for cases in Turkey are obtained with the decision tree, for deaths with polynomial regression method. The results reached for the United States of America and Russia are similar and the optimal results are obtained by polynomial regression. However, while the optimal results are obtained by neural networks in the Indian data, linear regression for the cases in the Brazilian data, neural network for the deaths, decision tree for the cases in France, polynomial regression for the deaths, neural network for the cases in the UK data and decision tree for the deaths are the methods that produced the optimal results. These results also give an idea about the similarities and differences of country characteristics.

Keywords Artificial neural network · Covid-19 · Decision tree · Linear regression · Polynomial regression · Support vector regression

1 Introduction
Covid-19, which emerged at the end of 2019, became a pandemic in a short time and continues to be a pandemic today. The cause of Covid-19 disease is the virus named SARS-CoV-2. It was discovered at the end of 2019 and it caused a pandemic in a short time all over the world. According to the figures announced by the World Health Organization, there are 611,421,786 confirmed cases and 6,512,438 deaths as of September 23, 2022. Covid-19 has been effective since the first months of 2020 in Turkey. In the statistics of the World Health Organization, the number of confirmed cases for Turkey on 23 September 2022 is 16,852,382 and the number of deaths is 101,068. Turkey ranks 11th in the world in terms of the number of cases and 19th in terms of the number of deaths (World Health Organization 2022). According to estimates, the pandemic is not likely to come to an end in the near future.

Covid-19 has caused various difficulties in the health systems of countries. Many proved to be insufficient for such a large-scale epidemic. The lessons to be learned from this epidemic will be an important guide both in determining what needs to be done in the later stages of this ongoing epidemic and in determining the strategy for new epidemics that may be encountered in the future.

When the number of cases and deaths are compared on the basis of countries, differences emerge. The response of societies to the epidemic has been shaped in the light of many economic, social and cultural phenomena.

We think that a study on this subject will be useful in understanding the nature of the pandemic and making emergency plans accordingly for the future. The subject of this paper is the estimation of the total number of cases and deaths caused by Covid-19 in Turkey, using machine
learning algorithms and finding the optimal way to represent the time series model. To compare and contrast the behavior of the two time series' with other countries, the data of the USA, Russia, France, Brazil, UK and India are also used and estimations are done.

The main contribution of this study is to help to understand and formulate the nature of the pandemic in Turkey, as well as to provide a comparison with the pandemic behavior of some other countries where the pandemic left many deaths behind. This study also shows that the total number of cases and total deaths for the countries in question cannot always be expressed in the same way and the methods will vary depending on the specific conditions of the countries.

The paper is organized as follows: literature review is provided on time series estimation of Covid-19 data in various countries in Sect. 2. Machine learning methods employed in this work and performance criteria are described in Sect. 3. Experimental results and discussions are given in Sect. 4. Finally, conclusion is drawn in Sect. 5.

2 Literature review

Machine learning methods are numerous and diverse in scope. The nearest neighbor method, artificial neural networks, decision trees, Bayesian network, clustering and similar other methods are among popular machine learning methods (Ertel 2017). Such methods fall under the artificial intelligence meta-concept. Deep learning, which has been intensively researched recently, is included as a class under machine learning (Goodfellow et al. 2016). The application of machine learning methods in health problems is common. A comprehensive literature review on machine learning in healthcare is provided in Garg and Mago (2021).

Covid-19 also received a lot of attention form the machine learning community. A thorough list of artificial intelligence related Covid-19 studies is given in Khanna et al. (2022). Here, we list a few examples which focus on the time series prediction since the problem of this study is only this aspect of Covid-19 research.

In Obaid et al. (2020), early Covid-19 data are used to predict the trend of the spread using long-short-term memory approach. It is concluded that the situation was worrisome. In Wang et al. (2020), Johns Hopkins data are utilized to predict the trend of Covid-19 in the coming months in Russia, Peru and Iran using a modified version of long short-term memory model. In Khan and Gupta (2020), ARIMA and NAR models are used to predict the Indian trend using the first few months of data; whereas in Salgotra et al. (2020), genetic programming is used for the same purpose.

In Gambhir et al. (2020), the course of the disease in India is analyzed with 154 days of data using polynomial regression and 93% success rate is achieved in predicting the 3-week future. In Mandayam et al. (2020), it is concluded, due to the linear structure of the data, linear regression performed better than support vector regression when estimating with data published by Johns Hopkins University. In Rustam et al. (2020) with the same data, linear regression, LASSO regression, support vector machines and exponential smoothing methods are used to predict a number of new cases, deaths and recoveries within 10 days. Reportedly, the optimal result is obtained with the exponential smoothing algorithm.

In Ramchandani et al. (2020), the features used in prediction of Covid-19 cases through deep learning are analyzed. Twitter data are used to estimate the number of cases in the USA with regression models (Yang and Chen 2020).

In Singh and Dalmia (2020), the number of deaths for India is estimated by linear regression, given the number of cases. In Bhadana et al. (2020), the number of cases, deaths and recoveries in the next 5 days are calculated for India using linear regression, decision tree, LASSO regression, support vector machines and random forest algorithms, polynomial regression giving the optimal results.

Using data from Brazil, India, Peru, Russia and the USA, the spread of Covid-19 is predicted by a wavelet-coupled random vector functional link network (Hazarika and Gupta 2020). In Leon et al. (2021), the number of cases and deaths are estimated with polynomial regression, support vector regression, Holt Winter, ARIMA and Facebook Prophet models using Bangladesh data. The closest result is obtained with the Facebook Prophet model.

In Kumari et al. (2021), using Indian data, the number of cases, deaths and recoveries for 30 days are estimated with multiple regression and autoregression. In Gupta et al. (2021), the number of cases, deaths and recoveries are estimated for India with random forest, support vector machines, decision tree and neural network models. The closest results are obtained with the random forest model. In Yudistira et al. (2021), the factors affecting the spread of Covid-19 are examined with a convolutional long short-term memory model.

In de Oliveira et al. (2021), the number of cases and deaths are predicted using artificial neural networks for Brazil, Portugal and the United States of America. In Ayobi et al. (2021), long short-term memory, convolutional long short-term memory and gated recurrent unit models are used to predict the future trends in Australia and Iran.

In ArunKumar et al. (2021), the number of cases, deaths and recoveries for a 60-day horizon are estimated using the ARIMA and SARIMA models, by selecting 16
countries heavily affected by the pandemic. SARIMA model predictions turn out to be more realistic. Particle swarm optimization and fuzzy time series estimation are combined to estimate the number of cases for 10 countries in Kumar and Susan (2021).

When it comes to Turkey, various AR (autoregressive) models are compared in Koçak (2020). A comparison of the course of the pandemic in Turkey using the current data up to April 10, 2020, is done with 22 countries and the best fitting time series modeling is found to be the Box-Cox method (Ergül et al. 2020). The Poisson regression method using the data of the first two months gave more successful results in some periods in Taşdelen and Yıldırım (2020).

3 Methods

Machine learning methods are numerous. In this study, we chose to use methods that are particularly suitable for time series estimation. In this section, the machine learning methods employed to estimate the time series under consideration are described briefly.

3.1 Linear regression

Linear regression is a basic method for estimation. It assumes that a line can be fit to the given set of time series. It is very easy to implement and works only in the case of simple relationships as described in the Eq. (1):

\[ y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip} + e_i, \]  

where \( y_i \) is the estimated value, \( b_i \)'s are the parameters of the estimator to be determined, \( x_{ij} \) is the inputs to the estimator, \( e_i \) is the error that is caused by the estimator, and \( n \) is the degree of the polynomial used in the estimator.

3.2 Polynomial regression

Polynomial regression is more complicated in comparison to linear regression, it assumes that a polynomial fits the data instead of a line. The degree of the polynomial and, therefore, the complexity of the fit depends on the complexity of the time series. It is described by the Eq. (2):

\[ y_i = b_0 + b_1 x_i + b_2 x_i^2 + \cdots + b_m x_i^m + e_i, \quad i = 1, 2, \ldots, n, \]  

where \( y_i \) is the estimated value, \( b_i \)'s are the parameters of the estimator to be determined, \( x_i \) is the inputs to the estimator, \( e_i \) is the error that is caused by the estimator, and \( n \) is the degree of the polynomial.

3.3 Support vector regression

Support vector regression tries to find a hyperplane that fits as much data as possible. The expression is given by the Eq. (3):

\[ y = \sum_{i=1}^{n} \alpha_i K(x_i, u) + b, \]  

where \( x_i \)'s represent the data points, \( u \) represents the new data, \( \alpha_i \)'s and \( b \) represent the parameters of the regressor, \( K(x_i, u) \) represents the kernel function used in the regressor.

Support vector regressors can be designed with different kernel functions, depending on the complexity of the data, for example linear kernel is given by the Eq. (4):

\[ K(x_i, u) = x_i^T u, \]  

where \( x_i^T \) represents the transpose of \( x_i \).

Polynomial kernel is described by is given by the Eq. (5):

\[ K(x_i, u) = (\beta (x_i^T u) + 1)^m, \]  

where \( \beta \) represents a scaling factor and \( m \) represents the degree of the polynomial.

![Fig. 1 Example of a decision tree](image-url)
Radial basis kernel is given by the Eq. (6)

\[ K(x, u) = \exp\left(-\delta ||x - u||^2\right), \]  

where \(\delta\) represents a scaling factor (Kuhn and Johnson 2016).

### 3.4 Decision tree regression

A decision tree consists of decision nodes and leaves, which are formed according to the criteria set by the problem. It can be used either as a classifier or regressor. To use it as a regressor, a careful partitioning of the space must be done and in this case, training is more involved. An example of part of a decision tree is shown in Fig. 1.

### 3.5 Artificial neural network regression

Artificial neural networks are designed to mimic the biological neural networks in a simpler form. They are used for solving many problems such as image recognition, natural language processing, time series estimation, to name a few. A basic artificial neural network is shown in Fig. 2. The architecture of neural networks become complex when the problem is more sophisticated.

Each branch in the neural network is assigned an unknown weight and solving the problem using artificial neural network means finding all the weights such that the least error is produced at the output.

### 3.6 Performance criteria

Two performance criteria are used to measure the success of the models used for estimation of the time series. These are \(R^2\) and MAPE scores.

#### 3.6.1 \(R^2\)

\(R^2\) should ideally be between 0 and 1.0 indicates no variation around the mean, and 1 indicates complete variation around the mean. Values close to 1 are preferred. Negative \(R^2\) values can also occur where the model does not fit the data well. The formula used to calculate \(R^2\) is as follows:

\[ R^2 = \frac{\text{model variance}}{\text{true variance}}. \]  

#### 3.6.2 MAPE

MAPE (mean absolute percentage error) is the normalized version of the MAE (mean absolute error) value. The closer the result is to 0, the better the learning model. Its formula is given by the Eq. (8):

\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \]  

where \(y_i\) represents the ground truth, \(\hat{y}_i\) represents the estimated value, \(n\) is the number of data points.

### 4 Experimental results and discussion

The data related to Turkey are tabulated in the Covid-19-related web site of Turkish Ministry of Health (Turkish Ministry of Health 2022). These data contain the number of cases, deaths, recoveries, total number of cases and deaths. It was not consistently tabulated, and sometimes, part of the data is missing. Since the pandemic statistics published by the Turkish Ministry of Health do not contain figures for all the information for each period, the time interval of 29 July 2020–3 July 2021, where the data were complete, was used. The machine learning methods used are linear regression, polynomial regression, support vector regression with different kernel functions, decision tree and artificial neural networks.

For simulation studies, 16 GB RAM, Intel Core i7-9750HF processor, NVIDIA GeForce GTX 1650 graphics card were used. Programs were written in Python programming language in Anaconda environment, using Pandas, Numpy, Matplotlib, Scikit-learn, Tensorflow, Keras libraries.

In Fig. 3, the data of total cases and the predictions using the aforementioned machine learning methods are shown.

In Fig. 4, the data of total deaths and the relevant predictions are shown.
The optimal prediction for the total cases are obtained using decision tree and the worst using support vector regression with polynomial kernel. On the other hand, the optimal prediction for the total deaths are obtained using polynomial regression with a polynomial of degree 2 and the worst using support vector regression with polynomial kernel. The performance calculations are shown in Table 1. The decision tree algorithm produces 1.30% error and, therefore, 98.70% accurate for total cases and the polynomial regression approach causes 0.56% error and, therefore, 99.44% accurate in the case of total deaths.

To compare and contrast, the data related to Covid-19 in other countries where the pandemic caused the highest infections and deaths are also taken into account. First, since the USA is one of the most affected, we studied its data. In Fig. 5, the data of total cases and the predictions using the machine learning methods employed in this work are shown.
In Fig. 6, the data of total deaths and the relevant predictions are shown.

The optimal prediction for the total cases are obtained using polynomial regression with a polynomial of degree 4 and the worst using support vector regression with linear and polynomial kernels. On the other hand, the optimal prediction for the total deaths is obtained using polynomial regression with a polynomial of degree 4 and the worst using support vector regression with linear and polynomial kernels. The performance calculations are shown in Table 2. The polynomial regression produces 0.47% error and, therefore, 99.53% accurate for the total cases and the same approach causes 0.30% error and, therefore, 99.70% accurate in the case of total deaths. The predictions in the case of USA turned out to be more accurate than Turkey.

Second, we studied Covid-19 data of Russia. In Fig. 7, the data of total cases and the predictions using the machine learning methods employed in this work are shown.

In Fig. 8, the data of total deaths and the relevant predictions for Russia are shown.

The optimal prediction for the total cases are obtained using polynomial regression with a polynomial of degree 4 and the worst using support vector regression with linear and polynomial kernels. The optimal prediction for the total deaths are obtained using polynomial regression again with a polynomial of degree 4 and the worst using support vector regression with linear and polynomial kernels. The performance calculations are shown in Table 3. The polynomial
regression produces 1.17% error and, therefore, 98.83% accurate for the total cases and the same approach causes 1.29% error and, therefore, 98.71% accurate in the case of total deaths.

Third, we studied Covid-19 data of France. In Fig. 9, the data of total cases and the predictions using the machine learning methods used in this work are shown.

The data of total deaths and its predictions for France are shown in Fig. 10.

The optimal prediction for the total cases is obtained using decision tree and the worst using support vector regression with linear and polynomial kernels. The optimal prediction for the total deaths are obtained using polynomial regression again with a polynomial of degree 4 and the worst using support vector regression with linear and polynomial kernels. The performance calculations are shown in Table 4. The decision tree approach produces 6.21% error and, therefore, 93.79% accurate for the total cases.

### Table 2 Performance figures in the prediction of USA data

| USA      | Method              | $R^2$ | MAPE |
|----------|---------------------|-------|------|
| Total cases | Linear regression | 0.997 | 1.32 |
|          | **Polynomial regression** | **0.999** | **0.47** |
|          | SVR (linear kernel) | 0.894 | 8.62 |
|          | SVR (polynomial kernel) | 0.834 | 8.62 |
|          | SVR (RBF kernel) | 0.933 | 6.95 |
|          | Decision tree | 0.987 | 3.27 |
|          | ANN | 0.912 | 5.54 |
| Total deaths | Linear regression | 0.998 | 0.44 |
|          | **Polynomial regression** | **0.999** | **0.30** |
|          | SVR (linear kernel) | 0.944 | 4.75 |
|          | SVR (polynomial kernel) | 0.858 | 4.75 |
|          | SVR (RBF kernel) | 0.962 | 2.61 |
|          | Decision tree | 0.987 | 1.64 |
|          | ANN | 0.945 | 2.61 |

Best results are given in bold
cases and the polynomial regression causes 1.03% error and, therefore, 98.97% accurate in the case of total deaths.

We studied Covid-19 data of Brazil, as well. In Fig. 11, the data of total cases and the predictions using the machine learning methods used are shown.

The data of total deaths and the relevant predictions for Brazil are shown in Fig. 12.

The optimal prediction for the total cases is obtained using linear regression and the worst using support vector regression with linear and polynomial kernels. The optimal prediction for the total deaths are obtained using an artificial neural network and the worst using support vector regression with linear and polynomial kernels. The performance calculations are shown in Table 5. The linear regression approach produces 0.97% error and, therefore, 99.03% accurate for the total cases and the artificial neural network approach causes 1.39% error and, therefore, 98.61% accurate in the case of total deaths.

We also studied Covid-19 data of the UK. In Fig. 13, the data of total cases and the predictions using the machine learning methods used are shown.

The data of total deaths and the relevant predictions for the UK are shown in Fig. 14.

The optimal prediction for the total cases is obtained using artificial neural network approach and the worst using polynomial regression. The neural networks employed have three hidden layers with five neurons each, three input neurons and one output neuron. The performance calculations are shown in Table 6. The artificial neural network approach produces 2.83% error and, therefore, 97.17% accurate for the total cases and the artificial neural network approach causes 1.66% error and, therefore, 98.34% accurate in the case of total deaths.

We studied Covid-19 data of India for comparison. In Fig. 15, the data of total cases and the predictions using the machine learning methods used are shown.

The data of total deaths and the relevant predictions for India are shown in Fig. 16.

The optimal prediction for the total cases and deaths is obtained using artificial neural network approach and the worst using polynomial regression. The neural networks employed have three hidden layers with five neurons each, three input neurons and one output neuron. The

| Russia | Method          | $R^2$ | MAPE |
|--------|----------------|-------|------|
| Total cases | Linear regression | 0.996 | 3.09 |
|          | **Polynomial regression** | **0.998** | **1.17** |
|          | SVR (linear kernel) | 0.994 | 10.91 |
|          | SVR (polynomial kernel) | 0.927 | 10.91 |
|          | SVR (RBF kernel) | 0.995 | 4.71 |
|          | Decision tree | 0.997 | 2.89 |
|          | ANN | 0.997 | 1.36 |
| Total deaths | Linear regression | 0.991 | 6.65 |
|          | **Polynomial regression** | **0.999** | **1.29** |
|          | SVR (linear kernel) | 0.988 | 10.75 |
|          | SVR (polynomial kernel) | 0.965 | 10.75 |
|          | SVR (RBF kernel) | 0.997 | 4.94 |
|          | Decision tree | 0.996 | 4.36 |
|          | ANN | 0.996 | 1.96 |

Best results are given in bold

![Fig. 9 Prediction of total Covid-19 cases in France](image-url)
The performance calculations are shown in Table 7. The artificial neural network approach produces 1.60% error and, therefore, 98.4% accurate for the total cases and it causes 1.14% error and, therefore, 98.86% accurate in the case of total deaths.

When we compare the predictions of total cases and total deaths in each country, we can see that the values encountered in the USA were the most successfully predicted data. This means that the data used to predict the total cases and deaths were consistent. Table 8 shows the variation of the methods for optimal results. We make our comparisons using $R^2$ and MAPE values, which are common metrics to measure the performance of time series prediction.

To keep the discussion brief, we provide the $R^2$ and MAPE figures only for the total death figures of Turkey as the proof of optimality of the hyperparameter of the model. The values of $R^2$ and MAPE for varying degrees are shown in Table 9. The optimal degree is obtained when both the $R^2$ is the largest and the MAPE is the lowest. From Table 9 it is understood that this occurs when the degree equals two, therefore we employ degree 2 for optimal prediction. The discussion for other countries continue in a similar way.

The similar works predicting the time series use different criteria to measure the success of the algorithms. We compared our work with the works employing similar criteria for the countries considered here. In Gambhir et al. (2020), total cases in India are predicted using polynomial regression. The result is 93% accurate. In Gupta et al. (2021), total cases in India are predicted using random forest algorithm and the result is 83.54% accurate, whereas total death prediction with the same algorithm is 72.79% accurate. In Arora et al. (2020), the best accurate for India total cases is reached using stacked LSTM model and is 96.78%. In Shastri et al. (2020), ConvLSTM algorithm is used to predict both total cases and deaths in India and success rates are 97.82, 96.66%, respectively. ConvLSTM is also used in the USA data and the success rates are 98, 97.5%, respectively. Comparing our results, our success rate for total cases for India is 98.4%, for deaths 98.86%, for the USA 99.53 and 99.70%, respectively.

In Khan and Gupta (2020), the optimal $R^2$ value in predicting the Indian data is reached by the NAR model and is 0.97, whereas our optimal model ANN results in 0.99.
Therefore, in summary, our simulation studies resulted in better rates with respect to the works we could compare our work.

5 Conclusion

In this paper, we presented our results on estimating the time series behavior of Covid-19 data for Turkey and we compared our results with the data from the USA, Russia, France, Brazil, UK and India. The least successful estimation was obtained for the total cases in France, producing a 6.21% error. The USA data yielded the best results, with 0.47% error in the prediction of total cases and 0.30% error in deaths. Depending on the country, the machine learning performance varied significantly.
algorithm producing the optimal results varied. This could be explained by various factors, such as socio-cultural differences, hygienic conditions, quarantine practices during the pandemic, vaccinations, to name a few. In Turkey, the optimal estimation for the number of total cases was possible using decision tree algorithm, whereas deaths were estimated optimally when polynomial regression was used. In general, support vector regressors especially with linear and polynomial kernels did not produce satisfactory results. We aim to obtain a more comprehensive results by modeling the socioeconomic and psychological factors that affect the pandemic behavior in the future. In addition, the effect of the curfews imposed due to the pandemic should also be examined. It is also necessary to examine the effects of reducing the quarantine periods, which were

![Figure 13](https://example.com/fig13.png)

**Fig. 13** Prediction of total Covid-19 cases in the UK

![Figure 14](https://example.com/fig14.png)

**Fig. 14** Prediction of total deaths related to Covid-19 in the UK

| Table 6 | Performance figures in the prediction of the UK data |
|---------|---------------------------------|
| **UK**  | **Method**          | $R^2$   | MAPE  |
| Total cases | Linear regression    | 0.919  | 16.89 |
|           | Polynomial regression | 0.948  | 31.42 |
|           | SVR (linear kernel)  | 0.913  | 61.39 |
|           | SVR (polynomial kernel) | 0.692  | 61.39 |
|           | SVR (RBF kernel)     | 0.992  | 11.10 |
|           | Decision tree        | 0.997  | 4.53  |
|           | **ANN**              | **0.998** | **2.83** |
| Total deaths | Linear regression    | 0.874  | 12.20 |
|           | Polynomial regression | 0.983  | 5.26  |
|           | SVR (linear kernel)  | 0.869  | 26.62 |
|           | SVR (polynomial kernel) | 0.632  | 26.62 |
|           | SVR (RBF kernel)     | 0.989  | 4.18  |
|           | **Decision tree**    | **0.997** | **1.66** |
|           | **ANN**              | **0.995** | **2.26** |

Best results are given in bold.
Fig. 15 Prediction of total Covid-19 cases in India

![Graph showing prediction of total Covid-19 cases in India]

Fig. 16 Prediction of total deaths related to Covid-19 in India

![Graph showing prediction of total deaths related to Covid-19 in India]

Table 7 Performance figures in the prediction of India data

| India       | Method             | $R^2$ | MAPE |
|-------------|--------------------|-------|------|
| Total cases | Linear regression  | 0.992 | 4.99 |
|             | Polynomial regression | 0.970 | 18.47|
|             | SVR (linear kernel) | 0.991 | 17.57|
|             | SVR (polynomial kernel) | 0.905 | 17.57|
|             | SVR (RBF kernel)    | 0.995 | 5.31 |
|             | Decision tree       | 0.992 | 5.72 |
|             | ANN                 | **0.999** | **1.60** |
| Total deaths| Linear regression  | 0.996 | 2.62 |
|             | Polynomial regression | 0.959 | 13.97|
|             | SVR (linear kernel) | 0.993 | 10.07|
|             | SVR (polynomial kernel) | 0.947 | 10.07|
|             | SVR (RBF kernel)    | 0.993 | 5.14 |
|             | Decision tree       | 0.993 | 4.03 |
|             | ANN                 | **0.999** | **1.14** |

Best results are given in bold

Table 8 Comparison of prediction for each country

| Country   | Optimal model | $R^2$ | MAPE |
|-----------|---------------|-------|------|
| Turkey    | Total cases   | Decision tree | 0.999 | 1.30 |
|           | Total deaths  | Polynomial regression | 0.999 | 0.56 |
| USA       | Total cases   | Polynomial regression | 0.999 | 0.47 |
|           | Total deaths  | Polynomial regression | 0.999 | 0.30 |
| Russia    | Total cases   | Polynomial regression | 0.998 | 1.17 |
|           | Total deaths  | Polynomial regression | 0.999 | 1.29 |
| France    | Total cases   | Decision tree | 0.996 | 6.21 |
|           | Total deaths  | Polynomial regression | 0.998 | 1.03 |
| Brazil    | Total cases   | Linear regression | 0.999 | 0.97 |
|           | Total deaths  | ANN | 0.996 | 1.39 |
| UK        | Total cases   | ANN | 0.998 | 2.83 |
|           | Total deaths  | Decision tree | 0.997 | 1.66 |
| India     | Total cases   | ANN | 0.999 | 1.60 |
|           | Total deaths  | ANN | 0.999 | 1.14 |
applied as 14 days at the beginning of the pandemic, to one week after the start of the application of the vaccine, making the use of masks compulsory in public areas and removing them recently. We also wish to examine how the removal of measures after the decrease in deaths affects the course of the pandemic. We plan to apply deep learning models for the pandemic data in our future work. We would like to examine the trade-off between the complexity of the models and the results obtained this way. We think that such studies are important in terms of preparation for future pandemics.

**Funding** The authors did not receive support from any organization for the submitted work.

**Declarations**

**Conflict of interest** The authors declare that they have no competing interest.

**References**

Arora P, Kumar H, Panigrahi BK (2020) Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. Chaos Solit Fractals 139:1–10

ArunKumar KE, Kalaga DV, Sai Kumar CM, Chilkoor G, Kawaiji M, Brenza TM (2021) Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: auto-regressive integrated moving average (ARIMA) and seasonal auto-regressive integrated moving average. Appl Soft Comput 103:107161. https://doi.org/10.1016/j.soft.2021.107161

Ayobi N, Sharifzadeh A, A’ziz S, Baddoodi N (2021) Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. Results Phys 27(March):1–15. https://doi.org/10.1016/j.rinp.2021.104495

Bhadana V, Jalal AS, Pathak P (2020) A comparative study of machine learning models for Covid-19 prediction in India. IEEE 4th conference on information and communication technology (CICT). https://doi.org/10.1109/CICT51604.2020.9312112

de Oliveira LS, Gruetzmacher SB, Teixeira JP (2021) COVID-19 time series prediction COVID-19 time series prediction. Procedia Comput Sci 181:973–980. https://doi.org/10.1016/j.procs.2021.01.254

Ergül B, Yavuz AA, Aşık EG, Kalay B (2020) Dünya ’da ve Türkiye ‘de Nisan Ayları Ibaratı COVID-19 Salgını Verilerinin İstatistiksel Değerlendirilmesi statistical evaluation of the COVID-19 outbreak data as of April around the world and in Turkey. Anadolu Kliniği Tıp Bilimleri Dergisi 25(1):130–141. https://doi.org/10.21673/anadoluklin.719629

Ertel W (2017) Introduction to artificial intelligence, 2nd edn. Springer, New York, Cham, Switzerland

Gambhir E, Jain R, Gupta A, Tomer U (2020) Regression analysis of COVID-19 using machine learning algorithms. 2020 international conference on smart electronics and communication. Icosec, 65–71. https://doi.org/10.1109/ICOSEC49089.2020.9215356

Garg A, Mago V (2021) Role of machine learning in medical research: a survey. Comput Sci Rev 40:1–17. https://doi.org/10.1016/j.cosrev.2021.100370

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge, Massachusetts

Gupta VK, Gupta A, Kumar D, Sardana A (2021) Prediction of COVID-19 confirmed, death and cured cases in India using random forest model. Big Data Min Anal 4(2):116–123. https://doi.org/10.26599/BDMA.2020.9020016

Hazarika BB, Gupta D (2020) Modelling and forecasting of COVID-19 spread using wavelet-coupled random vector functional link networks. Appl Soft Computing J 96:106626. https://doi.org/10.1016/j.asoc.2020.106626

Khan MF, Gupta R (2020) ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. J Safety Sci Resil 1(June):12–18. https://doi.org/10.1016/j.jmsr.2020.06.007

Khanna VV, Chadaga K, Sampathila N, Prabhu S, Chadaga R, Uma-kanth S (2022) Diagnosing COVID-19 using artificial intelligence: a comprehensive review. Netw Model Anal Health Inform Bioinform. https://doi.org/10.1007/s13721-022-00367-1

Koçak M (2020) A comparison of time-series models in predicting COVID-19 cases COVID-19 Vakalarinin Tahmin Edilmesinde. Türkiye Klinikleri Biyoistatistik Dergisi 12(1):89–96. https://doi.org/10.5336/biostat.2020-75402

Kuhn M, Johnson K (2016) Applied predictive modeling. Springer, New York, USA

Kumar N, Susan S (2021) Particle swarm optimization of partitions and fuzzy order for fuzzy time series forecasting of COVID-19. Appl Soft Comput 110:107611. https://doi.org/10.1016/j.asoc.2021.107611

Kumari R, Kumar S, Poonia RC, Singh V, Raja L, Bhatnagar V, Agarwal P (2021) Analysis and predictions of spread, recovery and death caused by COVID-19 in India. Big Data Min Anal 4(2):65–75. https://doi.org/10.26599/BDMA.2020.9020013

Leon MI, Iqbal MI, Azim SM, Al Mamun KA (2021) Predicting COVID-19 infections and deaths in Bangladesh using machine learning algorithms. 2021 international conference on information and communication technology for sustainable development, ICICT4SD 2021-proceedings. 70–75. https://doi.org/10.1109/ICICT4SD50815.2021.9396820

Mandayam AU, Rakshit AC, Siddhesha S, Niranjan SK (2020) Prediction of Covid-19 pandemic based on regression. Proceedings-2020 5th international conference on research in computational intelligence and communication networks, ICRICCN 2020. 1–5. https://doi.org/10.1109/ICRICCN50933.2020.9296175

Obaid OI, Mohammed MA, Mostafa SA (2020) Long short-term memory approach for coronavirus disease prediction. J Inf Technol Manag 12:11–21. https://doi.org/10.22059/jitim.2020.79187
Ramchandani A, Fan C, Mostafavi A (2020) DeepCOVIDNet: an interpretable deep learning model for predictive surveillance of COVID-19 using heterogeneous features and their interactions. IEEE Access 8:159915–159930. https://doi.org/10.1109/ACCESS.2020.3019989

Rustam F, Reshi AA, Mehmood A, Ullah S, On BW, Aslam W, Choi GS (2020) COVID-19 future forecasting using supervised machine learning models. IEEE Access 8:101489–101499. https://doi.org/10.1109/ACCESS.2020.2997311

Salgotra R, Gandomi M, Gandomi AH (2020) Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. Chaos Solit Fractals. https://doi.org/10.1016/j.chaos.2020.109945

Shastri S, Singh K, Kumar S, Kour P, Mansotra V (2020) Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. Chaos Solit Fractals 140:1–11. https://doi.org/10.1016/j.chaos.2020.110227

Singh M, Dalmia S (2020) Prediction of number of fatalities due to Covid-19 using machine learning. 2020 IEEE 17th India council international conference, INDICON 2020. https://doi.org/10.1109/INDICON49873.2020.9342390

Taşdelen B, Yıldırım DD (2020) Türkiye’de COVID-19 Vaka Sayılarının Poisson Regresyon İle Tahmini Ve Alınan Önlemlerin Insidans Hızı Tahminlerine Etkisi Predicting COVID-19 cases in Turkey with Poisson regression and the effect of preventions on incidence rate ratio estimation. Türkiye Klinikleri Biyoistatistik Dergisi 12(3):293–302. https://doi.org/10.5336/biostat.2020-77595

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.