Neural Computing and Applications
https://doi.org/10.1007/s00521-024-09597-y

ORIGINAL ARTICLE

Spread patterns of COVID-19 in European countries: hybrid deep learning model for prediction and transmission analysis

Anıl Utku[1] · M. Ali Akcayol[2]

Received: 20 September 2023 / Accepted: 6 February 2024
© The Author(s) 2024

Abstract
The COVID-19 pandemic has profoundly impacted healthcare systems and economies worldwide, leading to the implementation of travel restrictions and social measures. Efforts such as vaccination campaigns, testing, and surveillance have played a crucial role in containing the spread of the virus and safeguarding public health. There needs to be more research exploring the transmission dynamics of COVID-19, particularly within European nations. Therefore, the primary objective of this research was to examine the spread patterns of COVID-19 across various European countries. Doing so makes it possible to implement preventive measures, allocate resources, and optimize treatment strategies based on projected case and mortality rates. For this purpose, a hybrid prediction model combining CNN and LSTM models was developed. The performance of this hybrid model was compared against several other models, including CNN, k-NN, LR, LSTM, MLP, RF, SVM, and XGBoost. The empirical findings revealed that the CNN-LSTM hybrid model exhibited superior performance compared to alternative models in effectively predicting the transmission of COVID-19 within European nations. Furthermore, examining the peak of case and death dates provided insights into the dynamics of COVID-19 transmission among European countries. Chord diagrams were drawn to analyze the inter-country transmission patterns of COVID-19 over 5-day and 14-day intervals.

Keywords COVID-19 · Deep learning · Machine learning · European countries

1 Introduction

The COVID-19 pandemic affects almost every field, including healthcare, the economy, management, supply chain, manufacturing, and education [1, 2]. Immediately many restrictions were introduced worldwide, such as a curfew, mask requirement, and suspension of international and domestic flights [3]. The COVID-19 virus is spread by droplet contact with a contaminated surface or infected individual [4]. The biggest challenge in controlling the spread of the virus is that infected people with no symptoms can be contagious, and the virus can spread from them to other people [5]. Some studies have been conducted to obtain the virus's genetic structure [6]. However, the rapid and unpredictable spread of the virus has placed a heavy burden on healthcare systems [7]. As of December 15, 2021, there were about 500 million confirmed cases and approximately 6 million deaths [7, 8]. Capacity planning should be performed for hospital demand changes, such as bed availability and the number of personal protective equipment [9, 10]. Predicting the spread of the pandemic is vital in the fight against COVID-19.

Pinter et al. [11] proposed a hybrid machine learning model for COVID-19 case prediction in Hungary. This hybrid model includes the combination of Support Vector Machine (SVM), ARIMA, and Long-Short Term Memory (LSTM). The experimental results demonstrated that the proposed hybrid machine learning model effectively predicts COVID-19 cases.

Babukarthik et al. [12] presented a genetic CNN model for predicting COVID-19. The developed model includes a

✉ Anıl Utku
anilutku@munzur.edu.tr

M. Ali Akcayol
akcayol@gazi.edu.tr

[1] Department of Computer Engineering, Faculty of Engineering, Munzur University, Tunceli, Turkey

[2] Department of Computer Engineering, Faculty of Engineering, Gazi University, Ankara, Turkey

Published online: 08 March 2024

learning process using genetic algorithm. In this process, the model weights are optimized by the genetic algorithm, and the structural features of the model are determined to provide the best performance. The study evaluates the accuracy and performance of the COVID-19 predictions using the GDCNN model. The model makes predictions using data features such as the number of COVID-19 cases, death rates, and test results. Predictions obtained are evaluated by comparing them with actual data. Experimental results demonstrated that the GDCNN model effectively predicts COVID-19 and provides higher accuracy than traditional methods.

Zoabi et al. [13] introduced a model utilizing machine learning techniques to diagnose COVID-19 through symptom analysis. Algorithms predict the likelihood of a patient having COVID-19 by analyzing symptoms. These predictions are validated on the samples in the training dataset, and performance evaluation is made. Algorithms estimate the probability of patients having COVID-19 using the combination of symptoms.

Zhao et al. [14] aimed to diagnose COVID-19 by developing a deep learning model based on computed tomography (CT) images and focused on the knowledge that certain features in the lungs of COVID-19 patients differ from normal lung images. The deep learning model detects these differences and differentiates COVID-19 from other respiratory diseases. Experimental results show that the developed model recognizes COVID-19 with high accuracy.

Ismael and Şengür [15] proposed a deep learning-based approach for detecting COVID-19 using chest X-ray images. The study used a dataset containing both positive and negative cases. Deep learning models have been used to diagnose COVID-19 by identifying and analyzing patterns in these images. Experimental results showed that deep learning models recognize COVID-19 with high accuracy.

Alassafi et al. [16] conducted a comparative analysis of deep learning methods to perform time series prediction of the COVID-19 outbreak. The study used a dataset containing various indicators such as infection numbers, number of hospital admissions, and death rates. Prediction models were created using LSTM and Recurrent Neural Network (RNN) models. Experimental results demonstrated that LSTM has 98.58% accuracy.

Al-Waisy et al. [17] presented a hybrid model for detecting COVID-19 in chest X-ray images. The study aims to develop a model that will help diagnose COVID-19 by using imaging methods such as chest radiography. The dataset used includes images from positive, negative, and normal breasts. Then, a hybrid model, COVID-CheXNet, a combination of deep learning algorithms, was created. CNN is used to extract features from images, while LVQ is used to identify the COVID-19 virus using these features. Experimental results show that COVID-CheXNet effectively detects COVID-19 in chest X-ray images. Experimental results showed that the developed model has high accuracy and sensitivity. In addition, the model minimizes false positive results with its ability to distinguish between normal and COVID-19 negative images.

Aslam and Biswas [18] discuss using machine learning methods to study COVID-19 death cases. Researchers try to identify factors that affect deaths by analyzing a large data set. Using machine learning algorithms, it evaluates data such as patients' demographics, medical history, and symptoms and builds a model to predict the risk of death. The results show that machine learning methods are effective in analyzing COVID-19 death cases and can be used to develop health policies and natural resources.

The motivation of this study is to predict future spread patterns of epidemics such as COVID-19, which have profoundly affected the health systems and economies of countries around the world. Taking preventive measures and planning health services throughout epidemics is possible. By determining the spread patterns of epidemics such as COVID-19, resource allocations can be optimized, measures can be taken to protect public health, and economic strategies can be determined.

This study aims to predict the number of cases and deaths in the top 20 European countries with the greatest caseload. A hybrid prediction model combining CNN and LSTM models was developed and compared extensively with CNN, eXtreme Gradient Boosting (XGBoost), Multilayer Perceptron (MLP), Linear Regression (LR), k-Nearest Neighbors (k-NN), SVM, Random Forest (RF), LSTM. The performance of the models was assessed using R-Squared (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) metrics. In addition, the dates of the highest cases and deaths were used to assess the transmission dynamics of COVID-19 among European countries.

This paper offers the following significant contributions:

- A prediction model based on the combination of CNN and LSTM was developed to predict the spread of COVID-19 in Europe.
- The developed CNN-LSTM model was compared with CNN, k-NN, LR, LSTM, MLP, RF, SVM, and XGBoost.
- Chord diagrams were drawn to analyze the inter-country transmission patterns of COVID-19 over 5-day and 14-day intervals.

2 Material and method

Time series analysis is a statistical method in which data is organized and analyzed as a time-varying series [19]. This analysis is used to understand data patterns, trends, and regularities measured or recorded over time. Time series analysis is widely used in many fields, such as medicine, economics, finance, meteorology, and marketing. Deep learning methods in time series analysis have become very popular recently. Deep learning is a powerful artificial intelligence approach that can perform automatic feature extraction in complex and large data sets and capture relationships over time [18]. Deep learning methods provide automatic extraction of features, the capture of nonlinear relationships, scalability, and long-term forecasting capability in time series analysis.

2.1 Prediction models

k-NN makes predictions for a sample based on the distance between the sample and its Neighbors [20]. The Euclidean distance can be used to calculate the distance between any two points in multi-dimensional space [21]. k-NN requires large memory for storing the entire dataset for prediction.

LR is the most popular multivariable method used in many different application areas. LR analyzes the relationships between variables in the data set and makes predictions. LR predicts the value of dependent variables using a linear equation [22]. RF creates an ensemble model by combining multiple decision trees. Different sub-datasets are created from the dataset by random sampling method. A decision tree is created and trained on each sub-dataset. When constructing decision trees, a random subset of features is chosen at each node, which is used to determine the best split point [23, 24]. SVM creates a decision class to classify data points and aims to best distinguish this class with maximum marginal decomposition [25]. The basic idea of SVM is to find a hyperplane that best separates the classes while representing the data points in a feature space. This hyperplane determines a decision class that provides the best possible classification of data points [26].

XGBoost is a tree-based learning algorithm that makes predictions using decision trees [27]. Combining multiple weak tree models reduces errors and creates a more robust prediction model. This algorithm can be used in various tasks, such as classification and regression [28]. MLP is one of the most basic and widely used types of artificial neural networks [29]. MLP is a feedforward neural network model with at least one hidden layer. Neurons in each layer receive weighted inputs from nodes in the previous layer, apply an activation function, and produce outputs [30].

CNN is a deep learning model that is especially effective in the image and visual data processing problems [31]. CNNs have a particular architecture that can better capture the structure and properties of the data. The convolution layer takes an image or a feature map as input and extracts the feature maps using the convolution process [32]. The activation layer provides non-linearity learning of the model by using activation functions such as Sigmoid, ReLU, and tanh. The pooling layer is used to summarize the outputs of the convolution layer and reduce its size. The fully connected layer provides a unified rendering of features from previous layers [33, 34].

LSTM is mainly used in processing data with sequential structure, such as time series data or language [35]. LSTM can learn longer-term dependencies than RNNs because it can store information from previous timesteps [36]. This is achieved by the cell state and gate mechanisms in the LSTM's buffer. Cell state is the component that represents the memory of LSTM and is used to store information. The cell state is updated at each time step and can preserve information from previous steps [37]. Forget gate controls what information from the cell state needs to be remembered. The input gate controls how new incoming information is added to the cell state. The output gate determines the output from the cell state [38, 39].

2.2 The developed CNN-LSTM-based prediction model

The sliding window method, which is used to transform time series problems into supervised learning problems, divides the time series into small time steps and predicts the target value for each time step. The sliding window method transforms the time series into data points containing an output value and one or more input properties, as shown in Fig. 1. This transformation can predict future values using historical data from the time series [40].

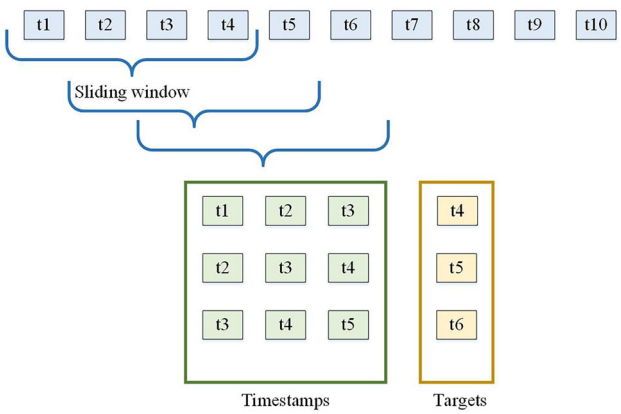

Fig. 1 The sliding window method

Data pre-processing has been performed for missing and erroneous data. The size of the sliding window was selected as 3. In this manner, the input to the sliding window consists of observational data from 3 consecutive time steps, while the output comprises the data from the 4th time step. Min–Max normalization was used to scale the dataset. Min–Max normalization is a scaling method used to transform the values of the data set into a particular range. This method converts the values of the dataset from the original range to the range [0, 1]. Min–Max normalization converts each value in the data set into a proportional value between the minimum and maximum values of the original value. This preserves the distribution of each value in the data set, making the values comparable [41].

The dataset was split, allocating 70% for the train set and 30% for the test set. 10% of the train set was split for validation. The validation data played a crucial role in optimizing the model parameters. Grid search was used for hyperparameter tuning.

Walk-forward validation is used to evaluate time series prediction models, as seen in Fig. 2. In this method, time series data is used sequentially from a specific starting point, and training and testing of the model are performed at each time step. Beginning from the starting point, the data set is progressively divided into training and test sets. A data window size is taken from the starting point determined in the first step and is used as a training set. Then, the new data point from the next time step is added to the test set, and the model's performance is evaluated on this data [42]. Walk-forward validation is shown in Fig. 2.

The CNN-LSTM model was developed to capture time and spatial characteristics of time series data. CNN was used to detect local structures of data over time. The LSTM was used to model the long-term dependencies of the time series. The CNN-LSTM model consists of interconnected CNN layers and successive LSTM layers, as seen in Fig. 3. CNN layers perform convolution operations to capture local features in the time series. LSTM layers are used to model long-term dependencies. The model has an output layer.

The dataset is trained on the generated CNN-LSTM model. In this step, the model's hyperparameters are determined, and the loss function and the optimizer are selected. The model learns to predict target outputs based on input data in the training process. The number of convolution layers, kernel size, activation function, number of filters of convolution layers, and pooling size parameters were optimized for CNN. Hidden unit number, cell number, epoch number, dropout rate, and activation function parameters were optimized for LSTM.

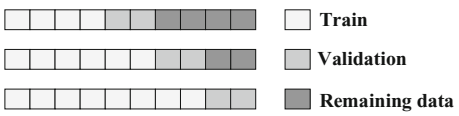

Fig. 2 Walk forward validation

2.3 Dataset

In this study, official COVID-19 statistics from the World Health Organization (WHO) panel were used as a dataset. The dataset consists of eight columns:

- *Date* The date or date range in which the data was recorded.
- *Country* The name of the country to which the data relates.
- *Country Code* A standard code for the country.
- *Region* The country's regional location or continent name.
- *New Cases* Reported new cases to date.
- *Total Cases* Reported total cases to date.
- *New Deaths* Reported new deaths to date.
- *Total Deaths* Reported total deaths to date.

The applied models were tested using data from 20 countries, with the highest cases among European countries. Figures 4 and 5 show the total number of cases and deaths of the 20 countries with the highest number of cases in Europe until April 2, 2022. Figures 4 and 5 present the countries' total number of cases and deaths comparatively. As seen in Fig. 4, the experimental studies were carried out for the top 20 countries.

As shown in Fig. 4, France stands out with the highest number of cases, reaching 25,068,545. After France, Poland has a higher death toll than any other European country. Finland, Belarus, and Slovenia are countries with less than one million deaths.

Figure 5 shows the cumulative number of deaths in the top 20 countries.

Figure 5 shows France has the highest death toll, with 139,272 deaths. Poland and Romania have a higher death toll than any other European country. Belarus, Ireland, and Slovenia reported the lowest number of deaths.

2.4 The evaluation metrics

MSE, RMSE, MAE, and R^2 metrics are commonly used to evaluate the regression models. MSE measures how far the predicted values are from the actual values. It is frequently used, especially in regression analysis and evaluating the performance of machine learning models. As seen in Eq. 1, MSE squares the difference between the estimated and actual values, sums these differences, and divides them by the number of observations to obtain the mean value.

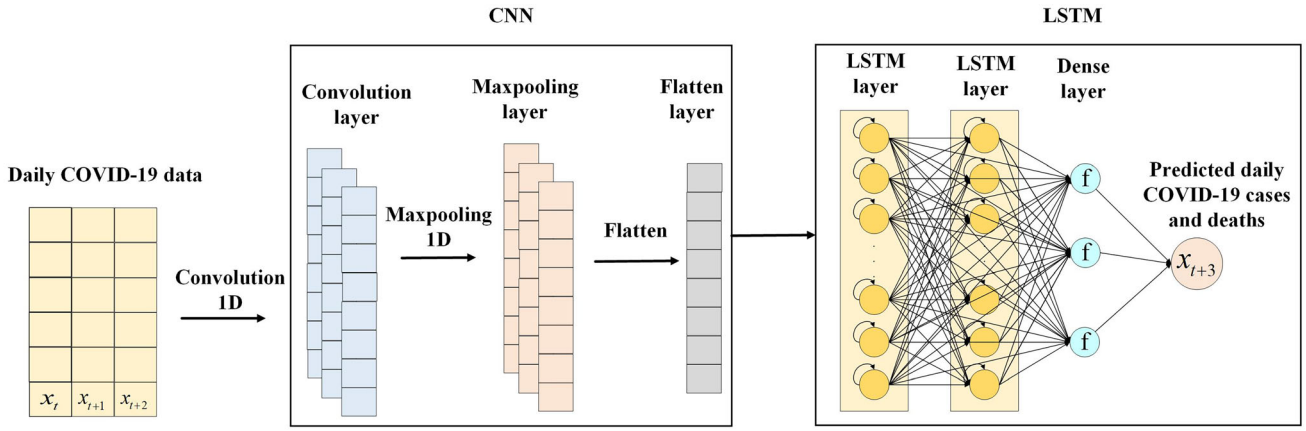

Fig. 3 The developed CNN-LSTM model

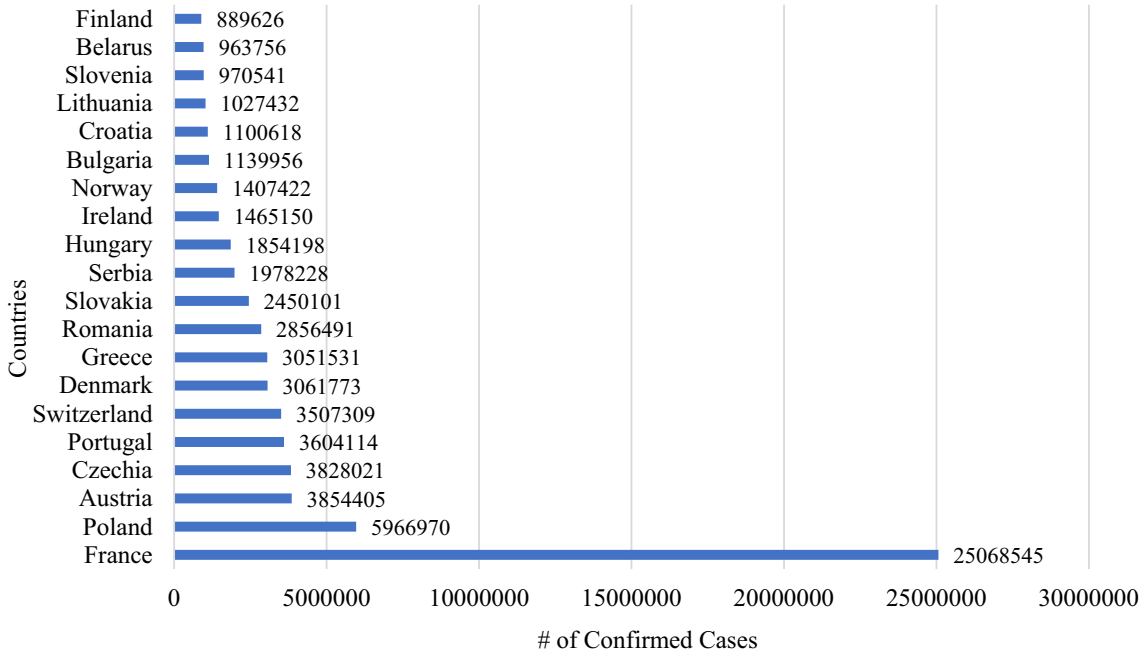

Fig. 4 Top 20 countries with the highest number of cases in Europe

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2 \tag{1}$$

here n denotes the total sample size. $\hat{y}$ represents the predicted values. y represents the true values. The RMSE measures the difference between the true and the predicted values, making the magnitude of errors more visually understandable. Like the MSE, the RMSE is employed to assess the proximity of the predicted values to the actual values. The value of the RMSE indicates the average amount of error to the true values, as seen in Eq. 2, and a lower RMSE value indicates better forecasting performance.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2} \tag{2}$$

MAE is used to measure how far the predicted values are, on average, from the true values, as seen in Eq. 3.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}| \tag{3}$$

R^2 as seen in Eq. 4, expresses how much of the independent variables explain the variance of the dependent variable.

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{(y-\bar{y})^2} \tag{4}$$

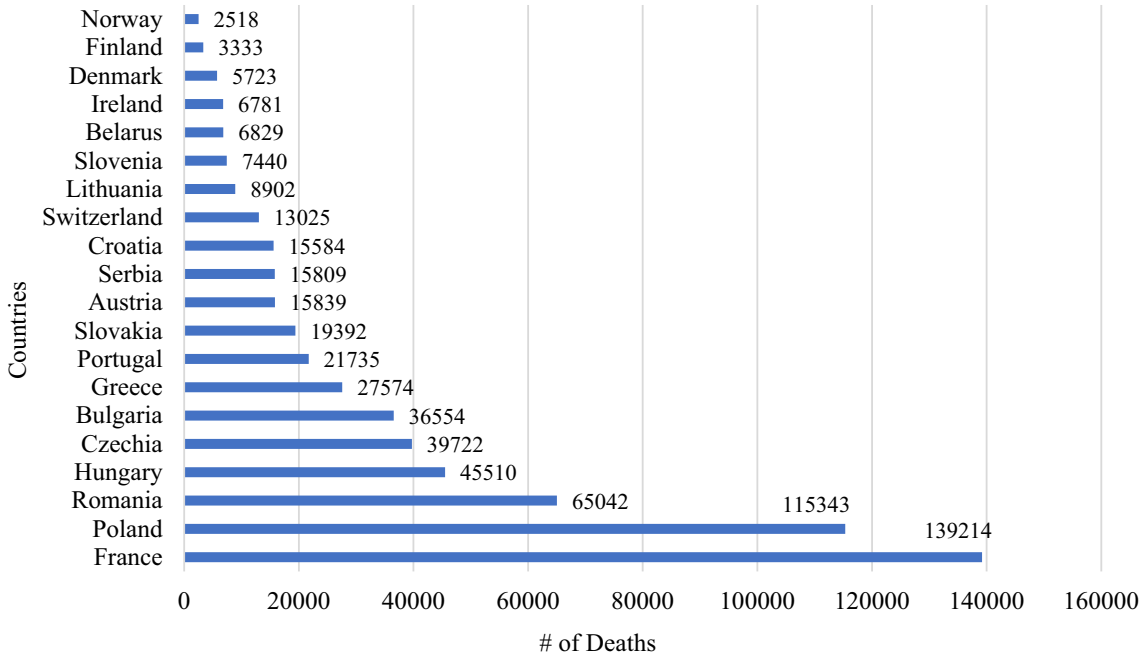

Fig. 5 Total number of deaths for the top 20 countries

here $\hat{y}$ represents the predicted value and $\bar{y}$ represents the mean of y.

3 The experimental results

To comprehensively evaluate the developed model, a comparative analysis is conducted against CNN, k-NN, LR, LSTM, MLP, RF, XGBoost, and SVM models.

3.1 The prediction of COVID-19 spread

This section provides the findings of experimental investigations to predict the incidence of cases and fatalities in European nations. Experimental studies were carried out according to each model's MSE, RMSE, R^2, and MAE.

The experimental results for predicting the case count, assessed through the MSE metric, are presented in Table 1.

Table 1 demonstrates that the CNN-LSTM-based model outperforms other models regarding MSE values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

The experimental results for predicting the case count, assessed through the RMSE metric, are presented in Table 2.

Table 2 demonstrates that the CNN-LSTM-based model outperforms other models regarding RMSE values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

The experimental results for predicting the case count, assessed through the MAE metric, are presented in Table 3.

Table 3 demonstrates that the CNN-LSTM-based model outperforms other models regarding MAE values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

The experimental results for predicting the case count, assessed through the R^2 metric, are presented in Table 4 and Fig. 6.

Table 4 demonstrates that the CNN-LSTM-based model outperforms other models regarding R^2 values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

Figure 6 displays the experimental results for R^2 values.

The experimental results for predicting the death count, assessed through the MSE metric, are presented in Table 5.

Table 5 demonstrates that the CNN-LSTM-based model outperforms other models regarding MSE values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

The experimental outcomes for predicting the death count, assessed through the RMSE metric, are presented in Table 6.

Table 6 demonstrates that the CNN-LSTM-based model outperforms other models regarding RMSE values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

Table 1 Results of the MSE for cases prediction (MSE*10^4)

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	114,498,201	3,226,323	8,574,886	4,154,520	11,323,782	3,204,442	2,493,307	3,147,843	**2,128,603**
PL	35550.71	13132.98	34331.14	13142.01	37899.88	11556.20	21384.07	11043.99	**9450.04**
AU	61288.85	25593.36	44157.52	26883.85	27123.30	24428.13	30821.93	20121.16	**16653.35**
CZ	56591.73	33571.11	85466.83	34206.14	56424.87	27776.04	27947.45	28077.80	**26495.17**
PT	165761.45	20610.02	172023.28	23671.37	155080.04	22237.09	24670.17	23366.30	**10968.90**
CH	93005.03	27261.55	116349.11	34531.28	89029.22	26993.56	27838.67	26903.31	**26358.41**
DK	197352.62	18340.11	193693.05	16356.58	200475.49	12209.21	13893.80	12190.31	**12170.00**
GR	36271.54	14218.14	35188.41	30761.41	34827.96	14556.76	14763.31	14219.14	**11334.98**
RO	41751.15	8829.13	41783.42	9970.95	45597.03	6955.01	11144.57	6545.60	**6955.01**
SK	21569.62	5070.55	22128.52	6931.80	22375.20	3793.65	5930.79	3569.29	**2693.77**
RS	8678.59	882.39	7855.22	930.70	8149.90	1099.57	1566.97	794.81	**699.68**
HU	40754.88	95052.32	49908.01	101896.11	43227.45	22331.55	76561.37	16435.31	**16140.86**
IE	11942.26	4660.59	11754.33	4589.96	13551.41	4370.72	5341.64	4321.43	**4254.24**
NO	46491.58	8660.57	43838.02	25871.59	56457.95	6099.43	9858.28	4933.51	**3919.08**
BG	3498.49	2003.68	4088.15	2203.44	4170.36	1410.49	2186.60	1234.08	**630.13**
HR	5056.55	3158.15	5334.13	3250.93	4641.38	3107.56	2999.92	3016.63	**1742.12**
LT	6647.87	865.26	5288.48	1023.23	6410.78	860.16	1289.03	860.16	**606.82**
SI	10108.85	3587.75	9667.38	3744.48	9875.63	3257.74	2816.93	3196.66	**2358.71**
BY	2137.58	593.69	2215.78	478.98	2132.56	492.62	610.36	460.94	**403.10**
FI	12568.56	1748.60	12453.43	4513.81	12580.46	1457.44	2035.76	1074.49	**768.39**

Bold values indicate the best of the compared results

Table 2 Results of the RMSE for cases prediction

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	107003.83	56800.73	92600.67	64455.56	106413.26	56607.79	49933.02	56105.63	**46136.78**
PL	5962.44	3623.94	5859.27	3625.19	6156.28	3399.44	4624.29	3323.25	**3074.09**
AU	7828.72	5058.98	6645.11	5184.96	5208.00	4942.48	5551.75	4485.66	**4080.85**
CZ	7522.74	5794.05	9244.82	5848.60	7511.64	5270.29	5286.53	5298.84	**5147.34**
PT	12874.83	4539.82	13115.76	4865.32	12453.11	4715.62	4966.90	4833.87	**3311.93**
CH	9643.91	5221.26	10786.52	5876.33	9435.53	5195.53	5276.23	5186.84	**5134.04**
DK	14048.22	4282.53	13917.36	4044.32	14158.93	3494.16	3727.43	3491.46	**3488.55**
GR	6022.58	3770.69	5931.98	5546.29	5901.52	3815.33	3842.30	3770.82	**3366.74**
RO	6461.51	2971.38	6464.00	3157.68	6752.55	2637.23	3338.34	2558.43	**2637.23**
SK	4644.31	2251.78	4704.09	2632.83	4730.24	1947.73	2435.32	1889.25	**1641.27**
RS	2945.94	939.36	2802.71	964.73	2854.80	1048.60	1251.79	891.52	**836.47**
HU	6383.95	9749.47	7064.56	10094.36	6574.75	4725.62	8749.93	4054.04	**4017.57**
IE	3455.75	2158.84	3428.45	2142.41	3681.22	2090.62	2311.19	2078.80	**2062.58**
NO	6818.47	2942.88	6621.02	5086.41	7513.85	2469.70	3139.79	2221.15	**1979.66**
BG	1870.42	1415.51	2021.91	1484.39	2042.14	1187.64	1478.71	1110.89	**793.80**
HR	2248.67	1777.12	2309.57	1803.03	2154.38	1762.82	1732.02	1736.84	**1319.89**
LT	2578.34	930.19	2299.67	1011.55	2531.95	927.45	1135.35	927.45	**778.98**
SI	3179.44	1894.13	3109.24	1935.06	3142.55	1804.92	1678.37	1787.92	**1535.81**
BY	1462.04	770.51	1488.55	692.08	1460.33	701.87	781.25	678.92	**634.90**
FI	3545.21	1322.35	3528.94	2124.57	3546.89	1207.24	1426.80	1036.57	**876.58**

Bold values indicate the best of the compared results

Table 3 Results of the MAE for cases prediction

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	52661.46	28825.11	40072.51	32337.30	51840.43	28830.23	25973.59	29205.45	**21748.66**
PL	3041.13	2285.45	3153.92	2403.89	3234.84	2105.07	2800.44	2060.52	**1947.70**
AU	4591.49	3191.04	3360.87	3260.09	4368.52	2915.52	3219.64	2301.82	**2002.60**
CZ	3537.84	2940.08	5338.05	3018.48	3572.38	2571.78	3079.16	2880.07	**2773.69**
PT	5768.33	2363.84	6030.71	2453.80	5725.82	2901.69	2603.04	3008.28	**1538.50**
CH	4763.22	2859.87	6095.45	2831.61	4668.04	2997.99	3031.54	2830.57	**2812.46**
DK	6649.67	2615.23	6535.63	2087.47	6748.05	1787.09	1939.02	1764.12	**1755.34**
GR	3873.90	2190.98	3795.76	3384.47	3759.50	2194.78	2221.78	2186.69	**1989.55**
RO	3105.88	1538.47	3120.38	1589.23	3278.30	1366.68	1692.91	1133.21	**1366.68**
SK	2362.42	1467.80	2507.06	1718.62	2524.02	1230.82	1609.15	1111.66	**901.21**
RS	1398.87	610.27	1354.13	593.21	1369.25	665.63	785.53	574.34	**499.54**
HU	3414.03	5432.42	3473.61	5652.95	3329.31	2927.90	4956.28	2245.37	**2150.06**
IE	1881.69	1254.20	1907.67	1264.89	2110.93	1192.60	1372.33	1229.16	**1204.81**
NO	3353.45	1313.36	3770.03	2896.04	4277.94	1158.59	1512.61	936.98	**872.04**
BG	1058.19	903.44	1168.36	850.20	1185.60	777.74	1003.49	732.87	**509.99**
HR	1385.70	1134.85	1482.77	1136.80	1354.94	1140.73	1119.08	1144.85	**844.09**
LT	1327.90	609.80	1224.83	642.02	1347.40	609.84	712.50	609.84	**506.05**
SI	1318.94	972.09	1306.61	989.64	1323.86	874.58	869.16	944.44	**667.31**
BY	592.00	447.93	632.87	396.54	605.62	401.29	484.67	392.41	**360.18**
FI	1984.69	722.15	1979.86	1263.83	1992.59	594.41	819.78	460.16	**419.07**

Bold values indicate the best of the compared results

Table 4 Results of the R^2 for cases prediction

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	0.336	0.722	0.442	0.642	0.342	0.727	0.683	0.729	**0.816**
PL	0.789	0.922	0.796	0.922	0.775	0.931	0.656	0.934	**0.944**
AU	0.401	0.877	0.548	0.871	0.650	0.883	0.572	0.903	**0.920**
CZ	0.446	0.671	0.385	0.665	0.448	0.728	0.667	0.725	**0.741**
PT	0.261	0.912	0.253	0.898	0.265	0.921	0.556	0.917	**0.942**
CH	0.475	0.799	0.371	0.746	0.485	0.801	0.747	0.802	**0.806**
DK	0.248	0.919	0.252	0.921	0.244	0.940	0.851	0.940	**0.941**
GR	0.448	0.842	0.458	0.658	0.462	0.838	0.782	0.842	**0.874**
RO	0.351	0.862	0.351	0.845	0.291	0.891	0.719	0.886	**0.891**
SK	0.060	0.825	0.184	0.761	0.175	0.869	0.504	0.869	**0.901**
RS	0.552	0.954	0.594	0.951	0.579	0.956	0.672	0.958	**0.963**
HU	0.415	0.260	0.407	0.250	0.425	0.528	0.285	0.653	**0.659**
IE	0.366	0.752	0.376	0.756	0.281	0.768	0.721	0.771	**0.774**
NO	0.478	0.800	0.428	0.558	0.377	0.859	0.771	0.867	**0.894**
BG	0.453	0.686	0.360	0.655	0.348	0.779	0.511	0.807	**0.918**
HR	0.370	0.606	0.335	0.595	0.422	0.613	0.674	0.624	**0.783**
LT	0.143	0.888	0.318	0.868	0.174	0.889	0.848	0.889	**0.921**
SI	0.110	0.729	0.148	0.717	0.130	0.754	0.789	0.759	**0.792**
BY	0.148	0.799	0.116	0.838	0.150	0.820	0.823	0.827	**0.863**
FI	0.301	0.830	0.302	0.562	0.300	0.846	0.730	0.886	**0.918**

Bold values indicate the best of the compared results

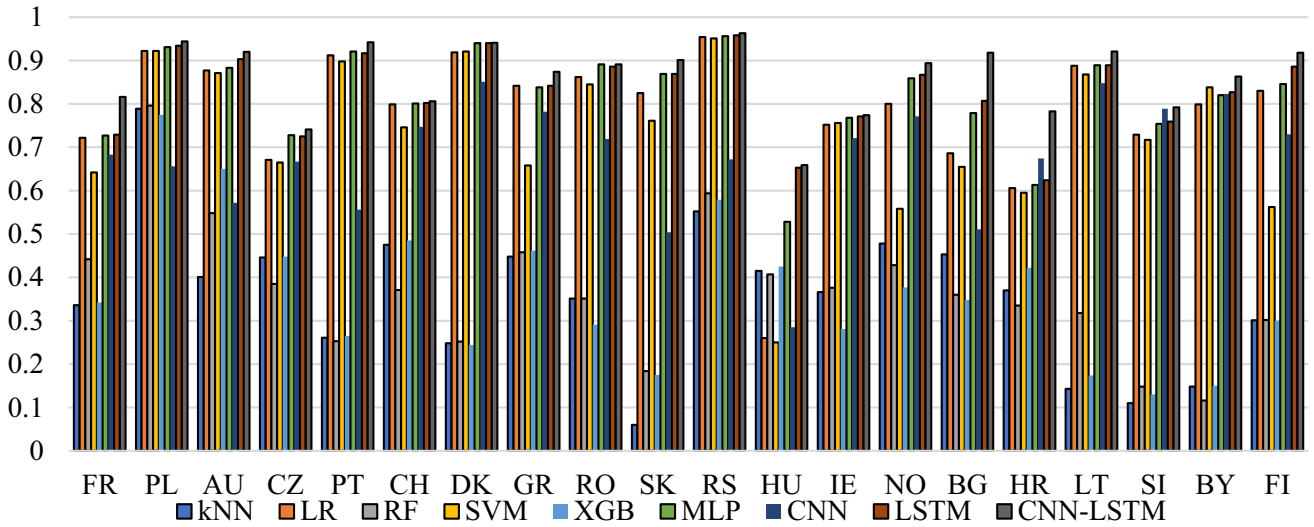

Fig. 6 The experimental results for R^2 values

Table 5 Results of the MSE for deaths prediction

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	5285.85	4735.67	5136.04	4939.63	8722.96	5616.23	4993.00	4773.82	**4672.47**
PL	11403.24	17576.68	14310.23	18768.08	16224.59	11782.82	25184.68	11395.92	**11146.41**
AU	693.05	289.65	426.88	295.46	285.64	273.59	936.07	289.12	**159.94**
CZ	118.15	68.18	71.25	68.35	102.39	70.54	72.96	68.39	**67.59**
PT	41.44	29.67	34.71	28.22	28.15	27.92	34.90	27.83	**25.62**
CH	18.06	15.98	18.14	16.09	21.70	15.90	17.00	15.89	**15.78**
DK	37.50	16.24	36.95	16.66	41.12	16.23	17.23	16.18	**15.84**
GR	193.70	132.27	166.58	134.73	267.41	131.39	146.25	130.99	**129.16**
RO	4294.89	2797.79	4105.01	2931.63	4184.58	2900.35	2950.45	2831.88	**2766.35**
SK	121.77	103.63	111.35	95.29	219.86	110.46	104.53	97.35	**83.43**
RS	21.51	10.45	12.19	10.23	17.46	10.51	12.99	10.11	**8.75**
HU	7901.94	9056.36	9144.96	9067.26	7984.54	6092.08	9084.84	6057.35	**5618.96**
IE	9.75	10.24	10.33	10.01	9.89	8.04	10.80	7.84	**7.82**
NO	48.57	21.21	41.70	24.69	53.89	21.31	36.13	20.88	**19.55**
BG	2613.43	1908.95	2523.54	2803.26	2588.05	2039.00	2486.18	1892.16	**1845.10**
HR	83.37	51.64	57.37	50.44	76.90	52.87	51.10	55.06	**50.07**
LT	32.70	29.79	36.58	29.90	42.38	29.35	30.95	29.38	**29.27**
SI	17.38	15.60	18.56	15.74	22.42	15.56	15.24	15.30	**12.28**
BY	6.28	4.48	5.98	4.31	5.89	4.48	4.38	4.28	**3.92**
FI	61.30	26.77	52.63	31.17	68.01	26.90	45.60	26.35	**24.67**

Bold values indicate the best of the compared results

The experimental outcomes for predicting the death count, assessed through the MAE metric, are presented in Table 7.

Table 7 demonstrates that the CNN-LSTM-based model outperforms other models regarding MAE values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

The experimental results for predicting the death count, assessed through the R^2 metric, are presented in Table 8.

Table 8 and Fig. 7 demonstrate that the CNN-LSTM-based model outperforms other models regarding R^2

Table 6 Results of the RMSE for deaths prediction

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	72.70	68.81	71.66	70.28	93.39	74.94	70.66	69.09	**68.35**
PL	106.78	132.57	119.62	136.99	127.37	108.54	158.69	106.75	**105.57**
AU	26.32	17.01	20.66	17.18	16.90	16.54	30.59	17.00	**12.64**
CZ	10.87	8.25	8.44	8.26	10.11	8.39	8.54	8.27	**8.22**
PT	6.43	5.44	5.89	5.31	5.30	5.28	5.90	5.27	**5.06**
CH	4.25	3.99	4.25	4.01	4.65	3.98	4.12	3.98	**3.97**
DK	6.12	4.03	6.07	4.08	6.41	4.02	4.15	4.02	**3.98**
GR	13.91	11.50	12.90	11.60	16.35	11.46	12.09	11.44	**11.36**
RO	65.5	52.89	64.07	54.14	64.68	53.85	54.31	53.21	**52.59**
SK	11.03	10.18	10.55	9.76	14.82	10.51	10.22	9.86	**9.13**
RS	4.63	3.23	3.49	3.19	4.17	3.24	3.60	3.18	**2.95**
HU	88.89	95.16	95.62	95.22	89.35	78.05	95.31	77.82	**74.96**
IE	3.12	3.20	3.21	3.16	3.14	2.83	3.28	2.80	**2.79**
NO	6.96	4.60	6.45	4.96	7.34	4.61	6.01	4.56	**4.42**
BG	51.12	43.69	50.23	52.94	50.87	45.15	49.86	43.49	**42.95**
HR	9.13	7.18	7.57	7.10	8.76	7.27	7.14	7.42	**7.07**
LT	5.71	5.45	6.04	5.46	6.50	5.41	5.56	5.42	**5.41**
SI	4.16	3.94	4.30	3.96	4.73	3.94	3.90	3.91	**3.50**
BY	2.50	2.11	2.44	2.07	2.42	2.11	2.09	2.07	**1.98**
FI	7.82	5.17	7.25	5.58	8.24	5.18	6.75	5.13	**4.96**

Bold values indicate the best of the compared results

Table 7 Results of the MAE for deaths prediction

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	49.823	50.537	50.864	49.261	65.569	54.141	49.591	48.881	**47.973**
PL	69.610	84.067	77.055	84.698	77.937	65.005	102.404	66.220	**65.395**
AU	14.907	11.017	9.971	11.172	10.800	10.977	19.054	10.986	**6.522**
CZ	7.343	5.497	5.536	5.556	6.841	5.832	5.600	5.439	**5.434**
PT	4.823	4.187	4.353	4.095	3.958	4.003	4.582	4.021	**3.779**
CH	3.235	3.074	3.236	3.110	3.511	3.072	3.154	3.054	**3.049**
DK	3.977	2.753	3.816	2.769	3.963	2.712	2.829	2.736	**2.693**
GR	11.030	9.007	10.091	9.024	12.867	8.991	9.093	9.025	**8.960**
RO	43.367	31.862	41.517	32.796	41.459	36.222	34.978	35.674	**31.189**
SK	7.633	7.107	7.213	6.890	9.515	7.625	7.114	6.669	**6.108**
RS	3.579	2.541	2.736	2.501	3.189	2.547	2.849	2.477	**2.278**
HU	51.150	49.580	49.935	50.109	49.464	47.017	53.695	43.788	**42.415**
IE	2.477	2.541	2.520	2.517	2.468	2.152	2.642	2.146	**2.115**
NO	4.164	2.735	3.810	2.923	4.409	2.731	3.550	2.701	**2.634**
BG	33.354	30.240	33.342	30.965	34.156	30.860	31.108	29.731	**29.065**
HR	6.660	5.268	5.481	5.228	6.457	5.314	5.252	5.483	**5.228**
LT	4.471	4.231	4.707	4.221	4.974	4.191	4.277	4.196	**4.183**
SI	3.035	2.959	3.151	2.952	3.540	2.960	2.892	2.956	**2.488**
BY	1.900	1.597	1.839	1.543	1.780	1.604	1.588	1.594	**1.524**
FI	5.255	3.452	4.809	3.690	5.565	3.448	4.481	3.409	**3.325**

Bold values indicate the best of the compared results

Table 8 Results of the R^2 for deaths prediction

Country	k-NN	LR	RF	SVR	XGB	MLP	CNN	LSTM	CNN-LSTM
FR	0.482	0.536	0.429	0.516	0.397	0.450	0.525	0.533	**0.543**
PL	0.714	0.559	0.641	0.530	0.593	0.704	0.459	0.714	**0.720**
AU	0.296	0.581	0.443	0.572	0.553	0.604	0.232	0.582	**0.678**
CZ	0.913	0.949	0.942	0.949	0.924	0.948	0.921	0.949	**0.950**
PT	0.755	0.824	0.794	0.833	0.833	0.835	0.700	0.835	**0.849**
CH	0.670	0.708	0.669	0.706	0.604	0.710	0.688	0.710	**0.712**
DK	0.786	0.907	0.789	0.905	0.765	0.907	0.866	0.907	**0.910**
GR	0.758	0.835	0.792	0.832	0.666	0.836	0.826	0.837	**0.839**
RO	0.758	0.844	0.769	0.837	0.764	0.836	0.794	0.841	**0.847**
SK	0.809	0.838	0.826	0.851	0.656	0.847	0.759	0.858	**0.885**
RS	0.944	0.972	0.968	0.973	0.954	0.973	0.784	0.973	**0.981**
HU	0.264	0.272	0.270	0.270	0.273	0.261	0.251	0.265	**0.319**
IE	0.449	0.438	0.421	0.442	0.451	0.513	0.421	0.525	**0.527**
NO	0.483	0.735	0.528	0.688	0.456	0.736	0.566	0.745	**0.764**
BG	0.294	0.436	0.318	0.416	0.300	0.449	0.388	0.430	**0.444**
HR	0.813	0.884	0.871	0.886	0.827	0.881	0.885	0.877	**0.888**
LT	0.607	0.642	0.561	0.641	0.491	0.647	0.636	0.648	**0.649**
SI	0.734	0.761	0.715	0.759	0.656	0.761	0.765	0.766	**0.814**
BY	0.244	0.464	0.280	0.484	0.290	0.465	0.474	0.488	**0.510**
FI	0.401	0.610	0.438	0.571	0.378	0.611	0.470	0.610	**0.634**

Bold values indicate the best of the compared results

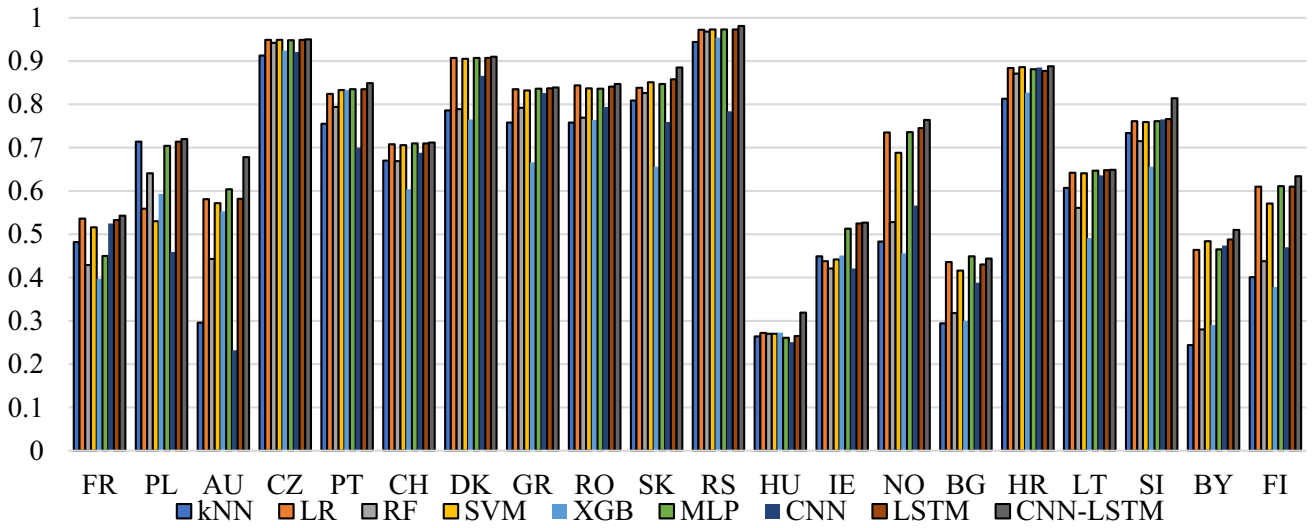

Fig. 7 The experimental results for R^2 values

values. Following the developed model, LSTM, MLP, and CNN exhibit more favorable outcomes than the other models.

Figure 8 illustrates the prediction charts of the proposed model for the top 20 European countries with the highest incidence of cases. The developed model has demonstrated superior performance compared to the other models, effectively capturing the variations in the number of cases across different countries.

3.2 Analysis of transmission dynamics of COVID-19 among European countries

The dates of the peak numbers of confirmed cases and deaths were employed to analyze the transmission dynamics of COVID-19 among European nations. The incubation period of COVID-19 is thought to extend to 14 days by WHO. Because the early onset of symptoms has been reported in 5 days, analysis has been performed

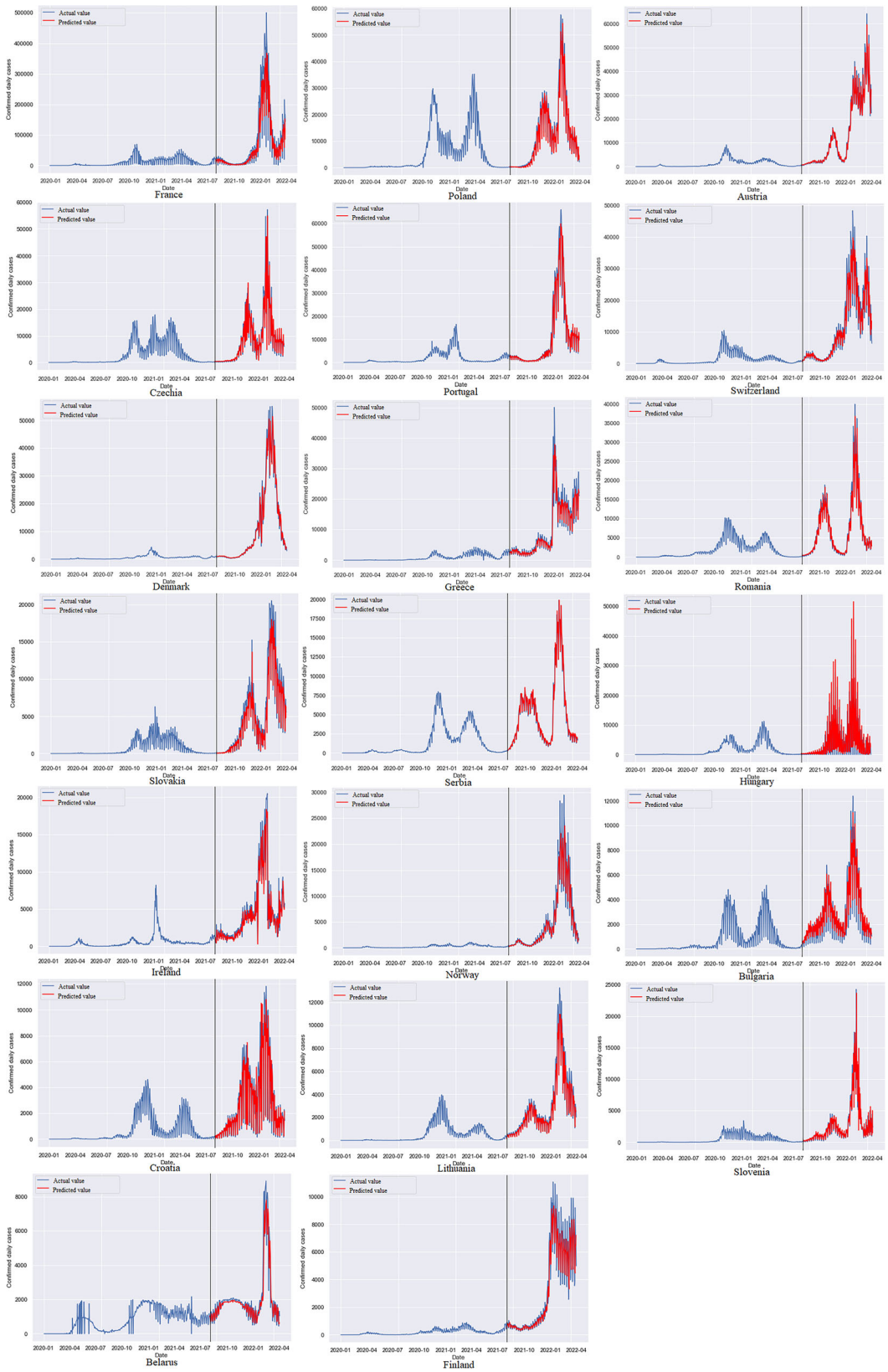

Fig. 8 Prediction charts of the developed model

for 5 days and 14 days. Also, the chord diagrams were drawn to analyze the inter-country transmission patterns of COVID-19 over 5-day and 14-day intervals. Table 9 presents the dates with the maximum case count in each country and the dates that are 5 and 14 days before and after these peak dates.

Figure 9 illustrates the dispersion of COVID-19 cases across European countries considering a 5-day incubation period.

Based on Fig. 9 and Table 9, it is apparent that Bulgaria and France, Croatia and Poland, Czechia and Romania, Hungary and Ireland, Switzerland and Serbia, Bulgaria, and Switzerland, Bulgaria and Poland, Bulgaria and Serbia, Croatia and France, Denmark and Slovakia, France and Switzerland, France and Poland, France and Serbia, Bulgaria and Poland, Poland and Portugal, Romania and Slovenia exhibit a similar spread pattern of COVID-19 cases when considering 5-day incubation period.

Figure 10 illustrates the dispersion of COVID-19 cases across European countries considering a 14-day incubation period.

Based on Fig. 10 and Table 9, it is apparent that Belarus and Denmark, Belarus and Slovakia, Bulgaria and Croatia, Bulgaria and Czechia, Bulgaria and France, Bulgaria and Hungary, Bulgaria and Ireland, Bulgaria and Switzerland, Bulgaria and Poland, Bulgaria and Portugal, Bulgaria and Serbia, Croatia and France, Croatia and Hungary, Croatia and Ireland, Croatia and Switzerland, Croatia and Poland, Croatia and Portugal, Croatia and Serbia, Czechia and Hungary, Czechia and Ireland, Czechia and Romania,

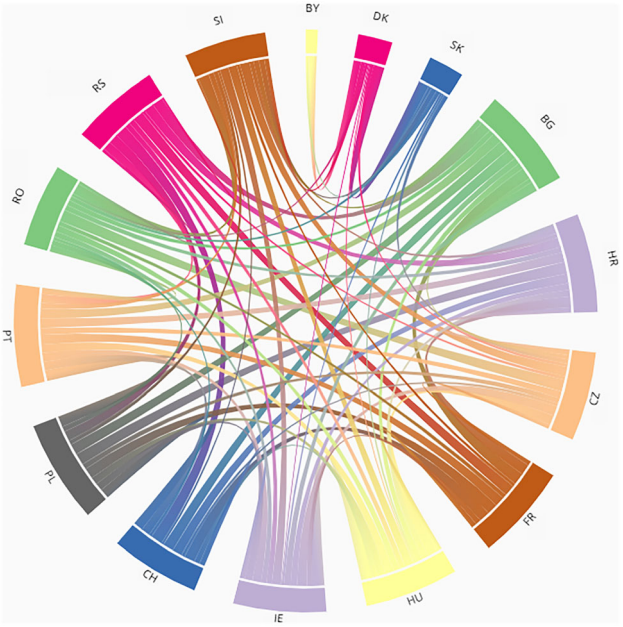

Fig. 9 The dispersion of COVID-19 cases across European countries considering a 5-day incubation period

Table 9 The dates with the maximum case count in each country

Country	14 days before peak date	5 days before peak date	Peak date	5 days after peak date	14 days after peak date
AT	2022/03/03–2022/03/16	2022/03/12–2022/03/16	2022/03/17	2022/03/18–2022/03/22	2022/03/18–2022/04/01
BY	2022/01/31–2022/02/13	2022/02/09–2022/02/13	2022/02/14	2022/02/15–2022/02/19	2022/02/15–2022/02/28
BG	2022/01/12–2022/01/25	2022/01/21–2022/01/25	2022/01/26	2022/01/27–2022/01/31	2022/01/27–2022/02/09
HR	2022/01/13–2022/01/26	2022/01/22–2022/01/26	2022/01/27	2022/01/28–2022/02/01	2022/01/28–2022/02/10
CZ	2022/01/28–2022/02/01	2022/01/18–2022/02/01	2022/02/02	2022/02/03–2022/02/07	2022/02/03–2022/02/16
DK	2022/01/29–2022/02/09	2022/01/05–2022/02/09	2022/02/10	2022/02/11–2022/02/15	2022/02/11–2022/02/24
FI	2019/12/29–2020/01/10	2020/01/06–2020/01/10	2020/01/11	2020/01/12–2020/01/16	2020/01/12–2020/01/25
FR	2022/01/12–2022/01/25	2022/01/21–2022/01/25	2022/01/26	2022/01/27–2022/01/31	2022/01/27–2022/02/09
GR	2021/12/21–2022/01/04	2021/12/31–2022/01/04	2022/01/05	2022/01/06–2022/01/10	2022/01/06–2022/01/19
HU	2022/01/17–2022/01/30	2022/01/26–2022/01/30	2022/01/31	2022/02/01–2022/02/05	2022/02/01–2022/02/14
IE	2022/01/17–2022/01/30	2022/01/26–2022/01/30	2022/01/31	2022/02/01–2022/02/05	2022/02/01–2022/02/14
CH	2022/01/11–2022/01/24	2022/01/20–2022/01/24	2022/01/25	2022/01/26–2022/01/30	2022/01/26–2022/02/08
LT	2020/12/12–2020/12/25	2020/12/21–2020/12/25	2020/12/26	2020/12/27–2020/12/31	2020/12/27–2021/01/09
NO	2020/10/11–2020/10/24	2020/10/20–2020/10/24	2020/10/25	2020/10/26–2020/10/30	2020/10/26–2020/11/09
PL	2022/01/13–2022/01/26	2022/01/22–2022/01/26	2022/01/27	2022/01/28–2022/02/01	2022/01/28–2022/02/10
PT	2022/01/14–2022/01/27	2022/01/23–2022/01/27	2022/01/28	2022/01/29–2022/02/02	2022/01/29–2022/02/11
RO	2022/01/18–2022/02/01	2022/01/28–2022/02/01	2022/02/02	2022/02/03–2022/02/07	2022/02/03–2022/02/16
RS	2022/01/11–2022/01/24	2022/01/20–2022/01/24	2022/01/25	2022/01/26–2022/01/30	2022/01/26–2022/02/08
SK	2022/01/26–2022/02/08	2022/02/04–2022/02/08	2022/02/09	2022/02/10–2022/02/14	2022/02/10–2022/02/23
SI	2022/01/20–2022/02/02	2022/01/29–2022/02/02	2022/02/03	2022/02/04–2022/02/08	2022/02/04–2022/02/17

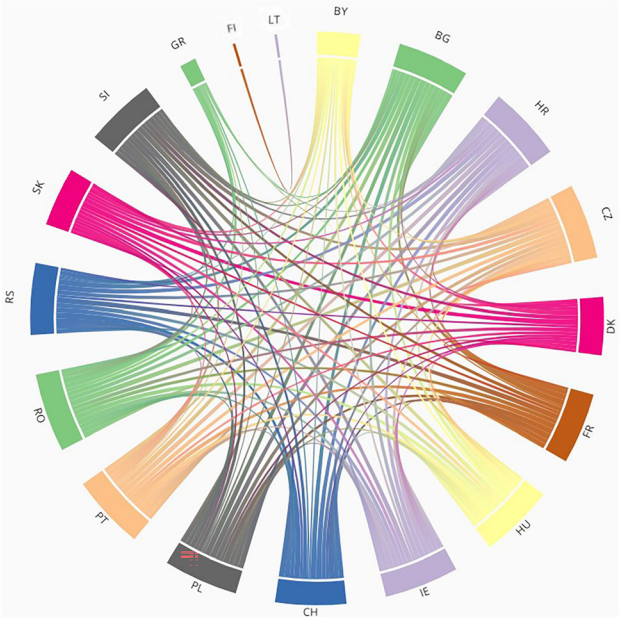

Fig. 10 The dispersion of COVID-19 cases across European countries considering a 14-day incubation period

Czechia and Slovenia, Czechia and Denmark, France and Switzerland, France and Poland, France and Portugal, France and Serbia, Greece and Romania, Hungary and Ireland, Hungary and Poland, Hungary and Portugal, Hungary and Romania, Hungary and Slovenia exhibit a similar spread pattern of COVID-19 cases when considering 14-day incubation period.

Table 10 shows the dates when the countries experienced the highest number of deaths, as well as the dates that are 5 and 14 days prior to and following these peak dates.

Figure 11 illustrates the dispersion of COVID-19 deaths across European countries considering a 5-day incubation period.

Based on Fig. 11 and Table 10, it is apparent that Austria and Serbia, Austria and Slovenia, Serbia and Slovenia, Portugal and Ireland, Belarus and France, Bulgaria and Switzerland, Denmark and Finland exhibit a similar spread pattern of COVID-19 cases when considering 14-day incubation period.

Figure 12 illustrates the dispersion of COVID-19 deaths across European countries considering a 14-day incubation period.

Based on Fig. 12 and Table 10, it is apparent that Austria and Croatia, Austria and Serbia, Austria and Slovenia, Bulgaria and Romania, Denmark and Finland, Ireland and Portugal, Serbia and Slovenia, Lithuania and Croatia, Switzerland and Czechia exhibit a similar spread pattern of COVID-19 cases when considering 14-day incubation period.

Table 10 The dates when the countries experienced the highest number of deaths

Country	14 days before peak date	5 days before peak date	Peak date	5 days after peak date	14 days after peak date
AT	2020/11/21–2020/12/04	2020/11/30–2020/12/04	2020/12/05	2020/12/06–2020/12/10	2020/12/06–2020/12/19
BY	2021/05/21–2021/05/30	2021/05/26–2021/05/30	2021/05/31	2021/06/01–2021/06/05	2021/06/01–2021/06/14
BG	2021/10/26–2021/11/08	2021/11/04–2021/11/08	2021/11/09	2021/11/10–2021/11/14	2021/11/10–2021/11/23
HR	2020/12/03–2020/12/16	2020/12/12–2020/12/16	2020/12/17	2020/12/18–2020/12/22	2020/12/18–2020/12/31
CZ	2020/10/21–2020/11/03	2020/10/30–2020/11/03	2020/11/04	2020/11/05–2020/11/09	2020/11/05–2020/11/18
DK	2022/02/19–2022/03/04	2022/02/28–2022/03/04	2022/03/05	2022/03/06–2022/03/10	2022/03/06–2022/03/19
FI	2022/02/15–2022/02/28	2022/02/24–2022/02/28	2022/03/01	2022/03/02–2022/03/06	2022/03/02–2022/03/15
FR	2020/03/21–2020/04/03	2020/03/30–2020/04/03	2020/04/04	2020/04/05–2020/04/09	2020/04/05–2020/04/18
GR	2021/04/20–2021/05/03	2021/04/29–2021/05/03	2021/05/04	2021/05/05–2021/05/09	2021/05/05–2021/05/18
HU	2021/11/23–2021/12/05	2021/12/01–2021/12/05	2021/12/06	2021/12/07–2021/12/11	2021/12/07–2021/12/20
IE	2021/01/20–2021/02/02	2021/01/29–2021/02/02	2021/02/03	2021/02/04–2021/02/08	2021/02/04–2021/02/17
CH	2020/10/30–2020/11/12	2020/11/08–2020/11/12	2020/11/13	2020/11/14–2020/11/18	2020/11/14–2020/11/27
LT	2020/12/12–2020/12/25	2020/12/21–2020/12/25	2020/12/26	2020/12/27–2020/12/31	2020/12/27–2021/01/09
NO	2022/03/14–2022/03/27	2022/03/23–2022/03/27	2022/03/28	2022/03/29–2022/04/02	2022/03/29–2022/04/11
PL	2021/03/25–2021/04/07	2021/04/03–2021/04/07	2021/04/08	2021/04/09–2021/04/13	2021/04/09–2021/04/22
PT	2021/01/18–2021/01/31	2021/01/27–2021/01/31	2021/02/01	2021/02/02–2021/02/06	2021/02/02–2021/02/15
RO	2021/10/20–2021/11/02	2021/10/29–2021/11/02	2021/11/03	2021/11/04–2021/11/08	2021/11/04–2021/11/17
RS	2020/11/22–2020/12/04	2020/11/31–2020/12/04	2020/12/05	2020/12/06–2020/12/10	2020/12/06–2020/12/19
SK	2020/12/21–2021/01/03	2020/12/30–2021/01/03	2021/01/04	2021/01/04–2021/01/04	2021/01/04–2021/01/18
SI	2020/11/21–2020/12/04	2020/11/31–2020/12/04	2020/12/05	2020/12/06–2020/12/10	2020/12/06–2020/12/19

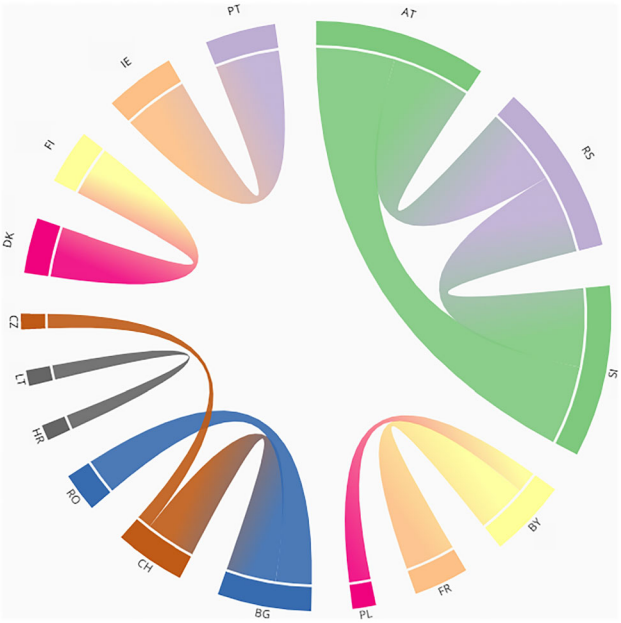

Fig. 11 The dispersion of COVID-19 deaths across European countries considering a 5-day incubation period

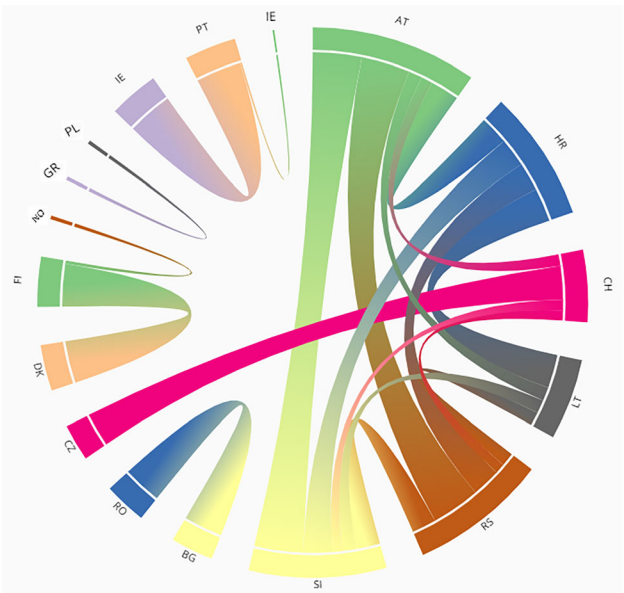

Fig. 12 The dispersion of COVID-19 deaths across European countries considering a 14-day incubation period

4 Conclusions and discussions

This study presents the development of a hybrid prediction model that integrates CNN and LSTM models, enabling the prediction of case and death numbers while facilitating analysis of the inter-country transmission patterns in European nations. Extensive comparisons were conducted between the developed model and CNN, k-NN, LR, LSTM, MLP, RF, SVM, and XGBoost models. The dataset consisted of WHO-confirmed cases and deaths up to April 2, 2022. The experimental studies focused on the top 20 countries in Europe with the highest case count. The models were evaluated using metrics such as MSE, RMSE, R^2, and MAE. Experimental findings revealed that the CNN-LSTM model showed superior prediction performance compared to alternative models.

In order to analyze the inter-country, spread of COVID-19 in European countries, the incubation period of COVID-19 was extended to 14 days, according to WHO's report. Because the early onset of symptoms has been reported in 5 days, analysis has been performed for 5 and 14 days. According to the experimental results, Bulgaria and France, Croatia and Poland, Czechia and Romania, Hungary and Ireland, Switzerland and Serbia, Bulgaria and Switzerland, Bulgaria and Poland, Bulgaria and Serbia, Croatia and France, Denmark and Slovakia, France and Switzerland, France and Poland, France and Serbia, Bulgaria and Poland, Poland and Portugal, Romania and Slovenia exhibit a similar spread pattern of COVID-19 cases when considering a 5-day incubation period.

This study is an essential source of information to guide decision-makers and researchers in areas such as epidemic management, health policy, and resource planning. The research results showed that the developed CNN-LSTM model effectively predicts the spread of COVID-19. The developed model possesses the capability to predict the trajectory of the epidemic by leveraging historical data.

The theoretical innovation of the hybrid model presented in this study is that a model has been developed that determines the spread pattern of COVID-19 more successfully than popular machine learning and deep learning models. By combining CNN and LSTM models, the developed model is capable of better modeling the spread dynamics of COVID-19 in European countries. In the developed hybrid model, CNN is used to extract the features' spatial features, while LSTM enables the extraction of time-dependent relationships. The experimental results showed that the CNN-LSTM model had a better prediction performance than the compared models and was more effective in managing the epidemic. Such theoretical innovations of the developed model enable more successful predictions to overcome future epidemics such as COVID-19.

The CNN-LSTM hybrid model takes advantage of the prominent features of CNN and LSTM models. CNN automatically extracts features in input data, while LSTM effectively captures long-term dependencies between consecutive time series data. In this way, complex patterns in the data are extracted.

The fact that CNN-LSTM is more successful than kNN, LR, RF, SVM, and XGBoost can be explained by the ability of LSTM to model temporal and spatial features.

CNN-LSTM has a feature learning ability that can automatically extract complex data patterns. However, traditional machine learning methods require data engineering processes. In addition, with LSTM in the structure of the model, a more effective learning process is achieved by performing operations such as remembering long-term dependencies, updating the model, and remembering and forgetting.

The fact that CNN-LSTM is more successful than MLP can be explained by processing historical data through feature extraction processes. CNN-LSTM reduces feature engineering by automating the feature extraction process by capturing relationships between different features and components in the data. However, in MLP, features must be defined beforehand. Additionally, LSTM can model changes over time, allowing trends in the data to be captured. The fact that CNN-LSTM is more successful than CNN and LSTM can be interpreted as CNN and LSTM alone are effective only at specific points. CNN is particularly effective at feature extraction and dimensionality reduction on image data and 1D time series data. LSTM, on the other hand, is effective in remembering and learning long-term dependencies and identifying trends in data. While CNN alone is not effective in the learning step, LSTM is not effective in the feature extraction stage. Therefore, the CNN-LSTM model, which combines these two models in a hybrid model, was more successful than the compared models.

In the studies examined in the introduction section, it is seen that the studies in the literature are aimed at predicting the number of COVID-19 cases/deaths and detecting COVID-19 from lung X-ray images. There is no study in the literature to analyze the spread of COVID-19 between countries and to predict the number of cases and deaths in European countries. This research provides a valuable tool for controlling the epidemic and tackling future outbreaks. Using deep learning techniques provides a solid basis for monitoring the epidemic's spread rate, assessing risks, and taking appropriate action. In addition, it is essential not to use estimates alone for outbreak management and policy decisions but to consider other factors and expert opinions.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The authors did not receive support from any organization for the submitted work.

Data availability The data that support the finding of this study are available from https://covid19.who.int/WHO-COVID-19-global-data.csv.

Declarations

Conflict of interest The authors of this manuscript declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Muhammad LJ, Algehyne EA, Usman SS, Mohammed IA, Abdulkadir A, Jibrin MB, Malgwi YM (2022) Deep learning models for predicting COVID-19 using chest X-ray images. In: Trends and advancements of image processing and its applications. Springer, Cham, pp 127–144
2. Mesgarpour M, Abad JMN, Alizadeh R, Wongwises S, Doranehgard MH, Jowkar S, Karimi N (2022) Predicting the effects of environmental parameters on the spatio-temporal distribution of the droplets carrying coronavirus in public transport—a machine learning approach. Chem Eng J 430:132761
3. Ghany KKA, Zawbaa HM, Sabri HM (2022) Gulf area COVID-19 cases prediction using deep learning. In: Digital transformation technology. Springer, Singapore, pp 521–530
4. Bisanzio D, Reithinger R, Alqunaibet A, Almudarra S, Alsukait RF, Dong D, Zhang Y, El-Saharty S, Herbst CH (2022) Estimating the effect of non-pharmaceutical interventions to mitigate COVID-19 spread in Saudi Arabia. BMC Med 20(1):1–14
5. Xiong Y, Ma Y, Ruan L, Li D, Lu C, Huang L (2022) Comparing different machine learning techniques for predicting COVID-19 severity. Infect Dis Poverty 11(1):1–9
6. Shorten C, Khoshgoftaar TM, Furht B (2021) Deep learning applications for COVID-19. J Big Data 8(1):1–54
7. Worldometers (2021) https://www.worldometers.info/coronavirus/. Accessed 15 Dec 2021
8. Shibuya K (2022) Formalizing models on COVID-19 pandemic. In: The rise of artificial intelligence and big data in pandemic society. Springer, Singapore, pp 95–125
9. Menezes B, Franzoi R, Yaqot M, Sawaly M, Sanfilippo A (2022) Advanced analytics for medical supply chain resilience in healthcare systems: an infection disease case. In: International conference of reliable information and communication technology. Springer, Cham, pp 759–768
10. Zeroual A, Harrou F, Dairi A, Sun Y (2020) Deep learning methods for forecasting COVID-19 time-series data: a comparative study. Chaos Solitons Fractals 140:110121
11. Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R (2020) COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. Mathematics 8(6):890
12. Babukarthik RG, Adiga VAK, Sambasivam G, Chandramohan D, Amudhavel JJIA (2020) Prediction of COVID-19 using genetic deep learning convolutional neural network (GDCNN). IEEE Access 8:177647–177666

13. Zoabi Y, Deri-Rozov S, Shomron N (2021) Machine learning-based prediction of COVID-19 diagnosis based on symptoms. npj Digit Med 4(1):3
14. Zhao W, Jiang W, Qiu X (2021) Deep learning for COVID-19 detection based on CT images. Sci Rep 11(1):1–12
15. Ismael AM, Şengür A (2021) Deep learning approaches for COVID-19 detection based on chest X-ray images. Expert Syst Appl 164:114054
16. Alassafi MO, Jarrah M, Alotaibi R (2022) Time series predicting of COVID-19 based on deep learning. Neurocomputing 468:335–344
17. Al-Waisy AS, Al-Fahdawi S, Mohammed MA, Abdulkareem KH, Mostafa SA, Maashi MS, Arif M, Garcia-Zapirain B (2023) COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. Soft Comput 27(5):2657–2672
18. Aslam H, Biswas S (2023) Analysis of COVID-19 death cases using machine learning. SN Comput Sci 4(4):403
19. Tang H (2022) Using machine learning techniques to study economic trends in various US industries in the post-epidemic era. In: International conference on computational modeling, simulation, and data analysis (CMSDA 2021), vol 12160. SPIE, pp 472–479
20. Prihatmono MW, Arni S, Iin JN, Moeis D (2022) Application of the K-NN algorithm for predicting data card sales at PT. XL Axiata Makassar. In: Conference series, vol 4, pp 59–64
21. Nayak S, Bhat M, Reddy NS, Rao BA (2022) Study of distance metrics on k-nearest neighbor algorithm for star categorization. J Phys Conf Ser 2161(1):012004
22. Vaulet T, Al-Memar M, Fourie H, Bobdiwala S, Saso S, Pipi M, Stalder C, Bennett P, Timmerman D, Bourne T, De Moor B (2022) Gradient boosted trees with individual explanations: an alternative to logistic regression for viability prediction in the first trimester of pregnancy. Comput Methods Programs Biomed 213:106520
23. Jui SJJ, Ahmed AM, Bose A, Raj N, Sharma E, Soar J, Chowdhury MWI (2022) Spatiotemporal hybrid random forest model for tea yield prediction using satellite-derived variables. Remote Sens 14(3):805
24. Alnahit AO, Mishra AK, Khan AA (2022) Stream water quality prediction using boosted regression tree and random forest models. Stoch Environ Res Risk Assess 36:1–20
25. Zouhri W, Homri L, Dantan JY (2022) Identification of the key manufacturing parameters impacting the prediction accuracy of support vector machine (SVM) model for quality assessment. Int J Interact Des Manuf 16:1–20
26. Harimoorthy K, Thangavelu M (2021) Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. J Ambient Intell Humaniz Comput 12(3):3715–3723
27. Alim M, Ye GH, Guan P, Huang DS, Zhou BS, Wu W (2020) Comparison of ARIMA model and XGBoost model for prediction of human brucellosis in mainland China: a time-series study. BMJ Open 10(12):e039676
28. Wang Y, Guo Y (2020) Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. China Commun 17(3):205–221
29. Li RYM, Tang B, Chau KW (2019) Sustainable construction safety knowledge sharing: a partial least square-structural equation modeling and a feedforward neural network approach. Sustainability 11(20):5831
30. Demir I, Karaboga HA (2021) Modeling mathematics achievement with deep learning methods. Sigma J Eng Nat Sci 39(5):33–40
31. Cheon S, Lee H, Kim CO, Lee SH (2019) Convolutional neural network for wafer surface defect classification and the detection of unknown defect class. IEEE Trans Semicond Manuf 32(2):163–170
32. Rahman T, Chowdhury ME, Khandakar A, Islam KR, Islam KF, Mahbub ZB, Kashem S (2020) Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Appl Sci 10(9):3233
33. Liu T, Bao J, Wang J, Zhang Y (2018) A hybrid CNN–LSTM algorithm for online defect recognition of CO_2 welding. Sensors 18(12):4369
34. Hartawan DR, Purboyo TW, Setianingsih C (2019) Disaster victims detection system using convolutional neural network (CNN) method. In: 2019 IEEE international conference on Industry 4.0, artificial intelligence, and communications technology (IAICT). IEEE, pp 105–111
35. Shewalkar A (2019) Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. J Artif Intell Soft Comput Res 9(4):235–245
36. Liu Y, Gong C, Yang L, Chen Y (2020) DSTP-RNN: a dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. Expert Syst Appl 143:113082
37. Dalgkitsis A, Louta M, Karetsos GT (2018) Traffic forecasting in cellular networks using the LSTM RNN. In: Proceedings of the 22nd Pan-Hellenic conference on informatics, pp 28–33
38. Sunny MAI, Maswood MMS, Alharbi AG (2020) Deep learning-based stock price prediction using LSTM and bi-directional LSTM model. In: 2020 2nd novel intelligent and leading emerging sciences conference (NILES). IEEE, pp 87–92
39. Poornima S, Pushpalatha M (2019) Prediction of rainfall using intensified LSTM based recurrent neural network with weighted linear units. Atmosphere 10(11):668
40 Brownlee J (2018) Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery, Vermont
41. Brownlee, J. (2022). Data preparation for machine learning.
42. Wan R, Mei S, Wang J, Liu M, Yang F (2019) Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting. Electronics 8(8):876

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