Journal of Information Systems and Management Research
Bilişim Sistemleri ve Yönetim Araştırmaları Dergisi

http://dergipark.gov.tr/jismar

Araştırma Makalesi / Research Article

COVID-19'un Yayılım Modelini Tahmin Etmek İçin Hibrit GA-ConvLSTM Modeli: Şanghay İşbirliği Örgütü İçin Bir Vaka Çalışması

Anıl UTKU[*,a]

[a,*] *Munzur Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, TUNCELİ, 62000, TÜRKİYE*

MAKALE BİLGİSİ

Alınma: 20.12.2024
Kabul: 29.01.2025

Anahtar Kelimeler:
Derin öğrenme,
Sangay İşbirliği Örgütü,
CNN,
LSTM,
Genetik algoritma,
COVID-19

Sorumlu Yazar
e-posta:
anilutku@munzur.edu.tr

ÖZET

COVID-19 dünyanın hemen her yerine çok hızlı bir şekilde yayılarak birçok insanın ciddi semptomlar yaşamasına ve hayatını kaybetmesine neden olmuştur. Bu çalışmada, sağlık sistemleri üzerindeki yükü hafifletmek ve salgının dağılımını tahmin etmek için planlar yapılabilmesi amacıyla derin öğrenme yöntemleriyle COVID-19'un yayılım örüntüsünü belirlemek amaçlandı. Bu amaçla CNN ve LSTM modelleri kullanılarak geliştirilen hibrit derin öğrenme modelinin hiper parametreleri genetik algoritma ile optimize edilerek daha başarılı bir tahmin performansı elde edilmiştir. GA-ConvLSTM modeli, SCO üyesi ülkelerde salgının yayılımını belirlemek için XGBoost, SVM, CNN, MLP, LSTM ve ConvLSTM ile test edilmiştir. Çalışmada, WHO tarafından sunulan 2020/01/03 ile 2022/05/31 tarihleri arasındaki günlük COVID-19 vaka ve ölüm verileri kullanılmıştır. Deneyler, GA-ConvLSTM'nin tüm ülkeler için vaka tahmininde 0,9'un üzerinde R^2 değerine sahip olduğunu göstermiştir. Deneyler, GA-ConvLSTM'nin ölüm tahmininde ülkelerin çoğunluğu için 0,9'un üzerinde R^2'ye sahip olduğunu göstermiştir. Ayrıca, COVID-19'un SCO ülkeleri arasındaki yayılım örüntüsü, 5 ve 14 günlük kuluçka dönemleri kullanılarak oluşturulan akor diyagramlarıyla belirlenmiştir.

DOI: 10.59940/jismar.1604942

Hybrid GA-ConvLSTM Model for Predicting the Transmission Pattern of COVID-19: A Case Study for Shanghai Cooperation Organisation

ARTICLE INFO

Received: 20.12.2024
Accepted: 29.01.2025

Keywords:
Deep learning,
Shanghai Cooperation Organization,
CNN,
LSTM,
Genetic algorithm,
COVID-19

Corresponding Authors
e-mail:
anilutku@munzur.edu.tr

ABSTRACT

COVID-19 has spread very quickly to almost every part of the world, causing many people to experience severe symptoms and lose their lives. In this study, it is aimed to determine the transmission pattern of COVID-19 with deep learning methods so that plans can be made to alleviate the burden on healthcare systems and predict the distribution of the epidemic. For this purpose, the hyper-parameters of the hybrid deep learning model developed using CNN and LSTM models were optimized with genetic algorithm, and a more successful prediction performance was achieved. The GA-ConvLSTM model was tested with XGBoost, SVM, CNN, MLP, LSTM, and ConvLSTM to determine the spread of the epidemic in the member countries of SCO. The study used daily COVID-19 case and death data between 2020/01/03 and 2022/05/31 presented by WHO. Experiments showed that GA-ConvLSTM has over 0.9 R^2 in case prediction for all countries. Experiments showed that GA-ConvLSTM has above 0.9 R^2 for the majority of countries when it comes to death prediction. In addition, the transmission pattern of COVID-19 among the SCO countries was determined with the chord diagrams created using 5 and 14 days' incubation periods.

DOI: 10.59940/jismar.1604942

1. *INTRODUCTION* *(GİRİŞ)*

Coronavirus Disease 2019 (COVID-19) has high fever and respiratory symptoms that can feel like a cold, flu, or pneumonia. [1]. Symptoms of COVID-19 are high fever, shortness of breath, muscle, joint, and body aches, weakness, and severe cough [2]. The incubation period of coronavirus disease varies between 5-14 days [3]. These symptoms may be very mild or severe. The disease can be severe in those who pose a risk for COVID-19 and are most likely to be affected, especially the elderly, those with cancer or immune-suppressing diseases, and people with lung diseases [4]. It can be said that almost all of those who lost their lives due to COVID-19 had different underlying diseases.

COVID-19 has caused many people to be admitted to intensive care units or even die [5]. During this period, the health systems of most countries remained inadequate. Health personnel were insufficient in number, intensive care units were at capacity, and morgues were full [6]. For this reason, artificial intelligence emerges to make future predictions in epidemic management. Using artificial intelligence methods, it is essential to analyze epidemics such as COVID-19, predict epidemics' spread, and develop strategies to combat the epidemic. The inferences obtained can be used to ensure that countries' health systems can control the epidemic and prevent its spread.

Shanghai Cooperation Organization (SCO), whose main area of cooperation is security, is a regional cooperation organization [7]. SCO is a security-based cooperation organization founded in 1996 with the cooperation of China, Kazakhstan, Russia, Tajikistan and Kyrgyzstan [8]. The SCO, which Uzbekistan joined in 2001, India and Pakistan in 2017, and Iran in 2023, aims to ensure neighborly relations and trust among its member countries and to establish security and peace on a regional basis [7, 8].

In this study, we aimed to develop preventive strategies by determining the spread of pandemics. In this way, countries' health systems are optimized, the workload of healthcare professionals is lightened, and strategies are developed to prevent the spread of pandemics. The hyper-parameters of the developed ConvLSTM model were optimized with the GA to create the GA-ConvLSTM hybrid model. GA-ConvLSTM was tested with Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), ConvLSTM, Long-Short Term Memory (LSTM) and Support Vector Machine (SVM) using a dataset of daily COVID-19 deaths and cases provided by World Health Organization (WHO). Additionally, experimental studies were conducted to determine the transmission pattern of COVID-19 among SCO member countries and chord graphs were created according to 5 and 14-day incubation times.

The innovations of this study are as follows:

• There is no study in the literature determining the distribution of COVID-19 in SCO member countries.
• GA-ConvLSTM was developed by optimizing the hyper-parameters of the ConvLSTM hybrid model using the GA.
• GA-ConvLSTM was compared in detail with SVM, CNN, XGBoost, SVM, MLP, and LSTM and the ConvLSTM hybrid model.
• To identify the transmission pattern of COVID-19 among SCO member countries, detailed analyzes were conducted for incubation periods.

2. *RELATED WORKS* *(İLİŞKİLİ ÇALIŞMALAR)*

Artificial intelligence is effectively used to predict the spread pattern of pandemics and diagnose diseases. Artificial intelligence techniques analyze big data and identify complex relationships and patterns between data, enabling planning and developing strategies for pandemics. This section examines studies in the literature on COVID-19 and deep learning.

Shahid et al. presented an evaluation of Autoregressive Integrated Moving Average (ARIMA), Bidirectional LSTM (Bi-LSTM), SVM, and LSTM for COVID-19 case prediction [10]. COVID-19 data from 10 countries were used as the dataset for approximately five months. Experiments showed that the Bi-LSTM model is more effective than the other models.

Abbasimehr and Paki presented a comparative analysis in which the parameters of CNN, multi-head attention, and LSTM were optimized with Bayes for predicting COVID-19 cases [11]. The applied models were compared with fuzzy fractals. Two different datasets consisting of 10 countries were used: long-term and short-term. Experiments showed that LSTM is more effective than the compared models in most countries.

Rauf et al. presented an evaluation of the Gated Recurrent Unit (GRU), Recurrent Neural Network (RNN), and LSTM to predict the spread of coronavirus [12]. Experiments showed that LSTM outperformed other models for the four compared countries.

Zhou et al. developed an application of Bi-LSTM, GRU, LSTM, and Dense-LSTM to determine COVID-19 deaths and cases [13]. Data from 12 countries until June 2022 were used as the dataset. Experiments showed that Dense-LSTM outperformed the other models.

Ukwuoma et al. presented an analysis of VGG-16, DenseNet, and InceptionV3 for coronavirus disease prediction from chest images [14]. A dataset consisting of approximately 21000 images was used in the study. Experiments showed that the DenseNet201 model outperformed the benchmark models with 96% and 98% accuracy in different test scenarios.

Al-Rashedi and Al-Hagery presented an application of LSTM, ARIMA, and CNN to determine the transmission pattern of COVID-19 [15]. Three different time intervals were used for predicting cases, recoveries, and deaths: long, medium, and short. Experiments showed that LSTM, in particular, and then CNN were quite successful.

Solayman et al. developed an application of ensemble models, k-Nearest Neighbor (kNN), CNN, Decision Tree (DT), LSTM, Logistic Regression (LR), SVM, and Random Forest (RF) for COVID-19 detection [16]. The dataset was preprocessed using synthetic oversampling. Experiments showed that the hybrid CNN-LSTM outperformed the benchmark models with over 96% accuracy.

3. *MATERIAL AND METHOD* *(MATERYAL VE METOT)*

SCO is an important regional and political cooperation organization. The model presented in this study may be effective for SCO member countries to combat epidemics that burden healthcare systems, such as COVID-19. It can contribute to developing strategies for epidemic management and developing cooperation between member countries.

3.1. Dataset *(Veriseti)*

This study aimed to develop a perspective for SCO member countries using the daily COVID-19 information presented by WHO. The dataset used includes the period between 2020/01/03 and 2022/05/31 when COVID-19 peaked. The dataset consists of the attributes date, region code determined by WHO, country name, country code, number of daily cases, total number of cases, total number of deaths, and daily deaths. For SCO member countries, the attributes of date and daily case and death numbers were selected according to country codes. Fig. 1 shows the daily COVID-19 cases for SCO member countries.

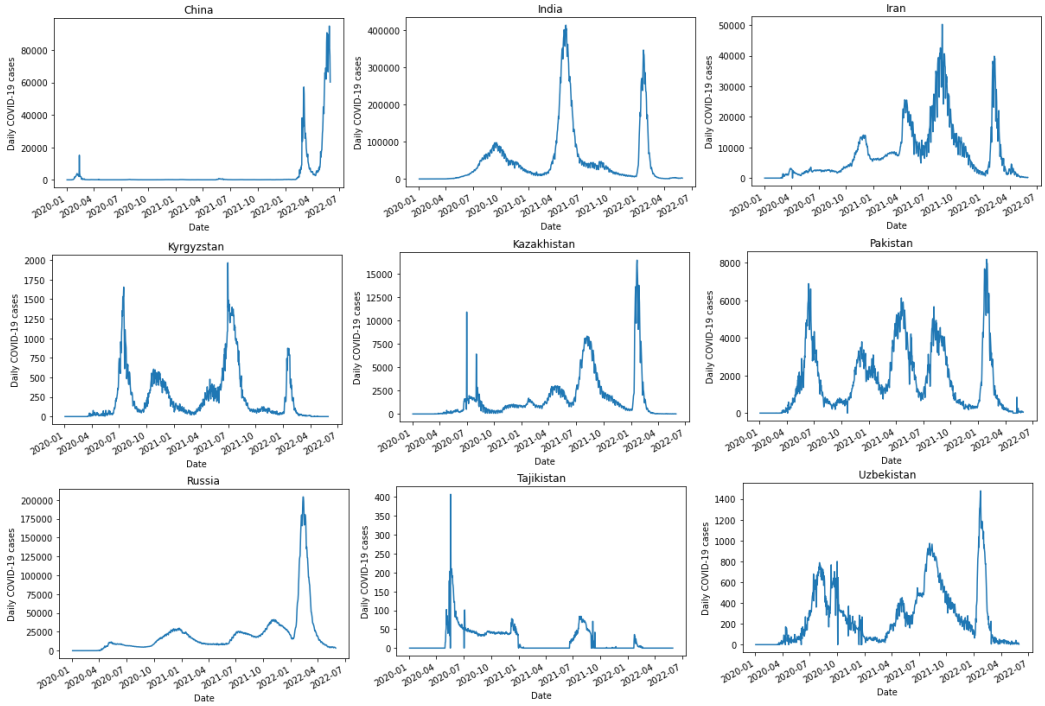

Figure 1. The daily COVID-19 cases for SCO member countries *(SCO üyesi ülkelerdeki günlük COVID-19 vaka sayıları)*

Fig. 2 shows the daily COVID-19 deaths for SCO member countries.

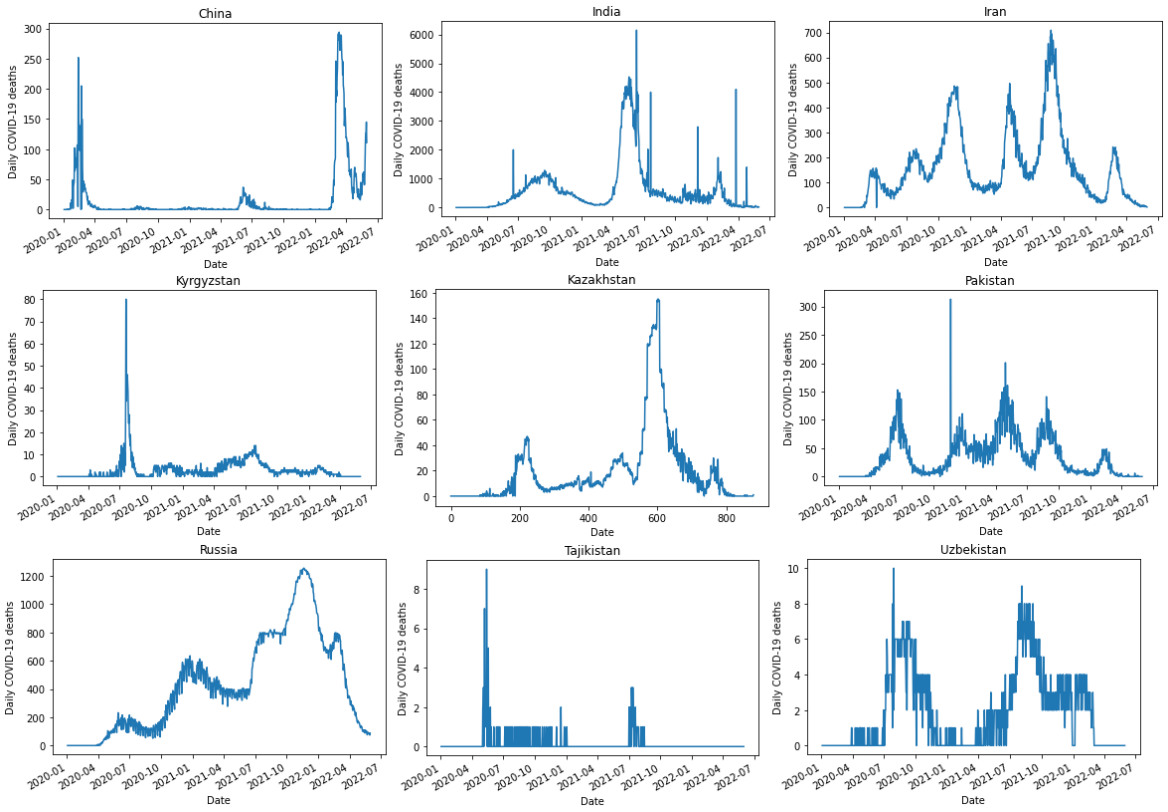

Figure 2. The daily COVID-19 deaths for SCO member countries *(SCO üyesi ülkelerdeki günlük COVID-19 ölüm sayıları)*

Fig. 3 indicates the total cases in the SCO member countries.

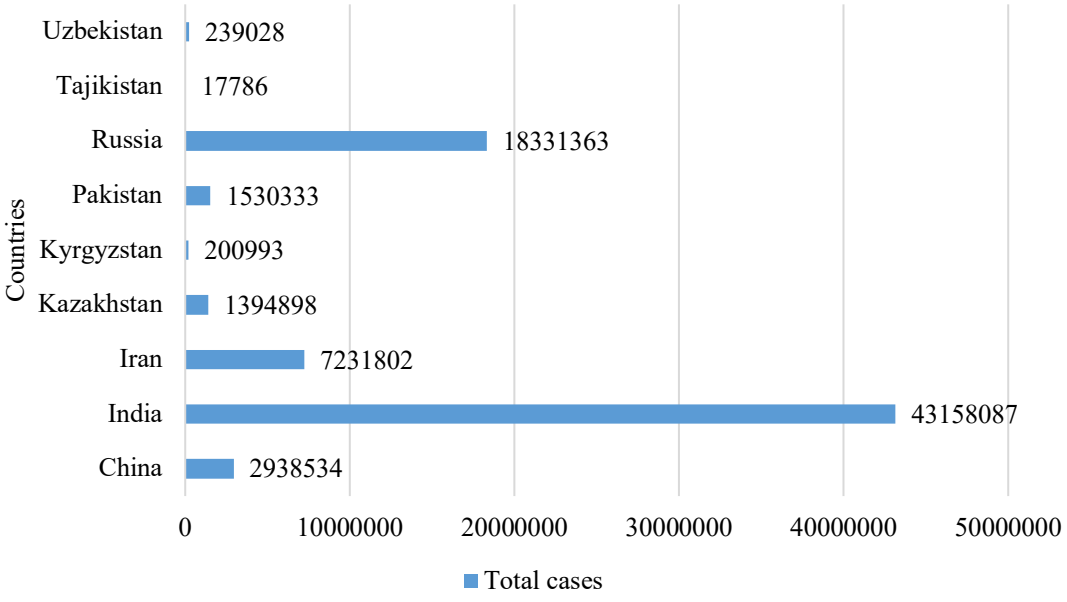

Figure 3. The total cases in the SCO member countries *(SCO üyesi ülkelerdeki toplam COVID-19 vaka sayıları)*

The total cases in the India is 43,158,087. After India, Russia, China and Iran have higher total cases. Fig. 4 shows the total deaths in the SCO member countries.

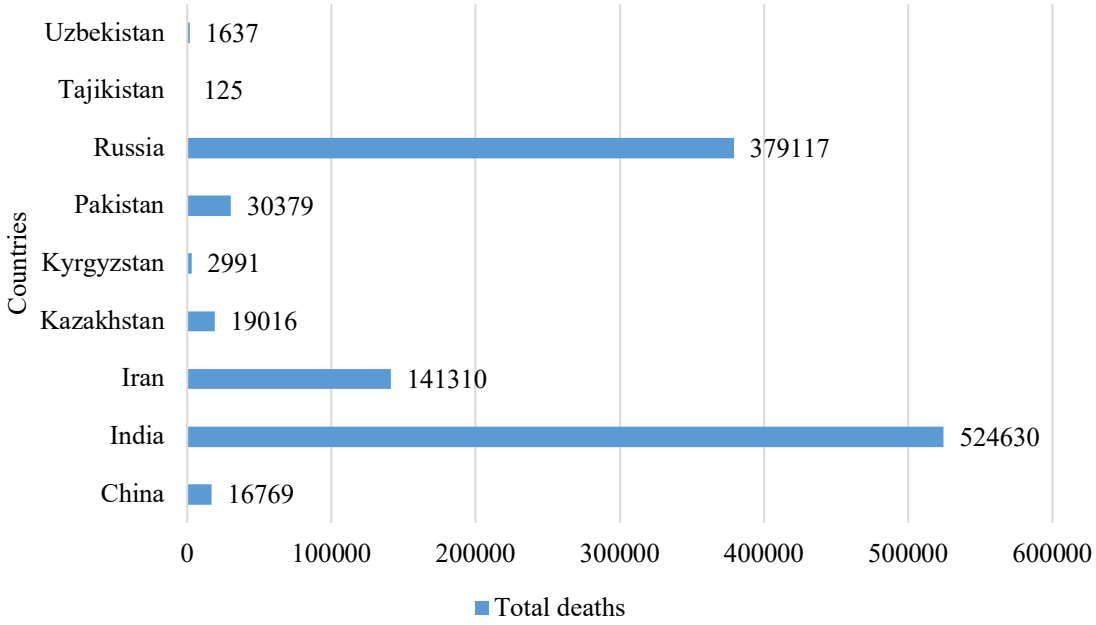

Figure 4. The cumulative number of deaths in the SCO member countries *(SCO üyesi ülkelerdeki toplam COVID-19 ölüm sayıları)*

The total deaths in India is 524,630. After India, Iran and Russia have maximum death counts.

Table 1 indicates the information of the deaths and cases in the SCO member countries.

Table 1. The information of the deaths and cases in the SCO member countries *(SCO üyesi ülkelerdeki ölüm ve vaka bilgileri)*

Country	Maximum case count	Maximum death count	Total cases	Total deaths
China	94753	294	2938534	16769
India	414188	6148	43158087	524630
Iran	50228	709	7231802	141310
Kazakhstan	16442	155	1394898	19016
Kyrgyzstan	1965	80	200993	2991
Pakistan	8183	313	1530333	30379
Russia	203949	1254	18331363	379117
Tajikistan	407	9	17786	125
Uzbekistan	1478	10	239028	1637

As seen in Table 1, India and Russia have the highest number of cases and deaths.

3.2. Data Pre-processing *(Veri Ön-işleme)*

The dataset used consists of daily COVID-19 time series data. For this reason, it is necessary to transform the dataset into a regression problem. The sliding window method was used for this purpose. This method ensures that the given input is set to the

specified window size, and the data point in the next time step is set as output [17]. For instance, if the window size is 3, t_1, t_2 and, t_3 will be the input, and t_4 will be the output, as shown in Fig. 5.

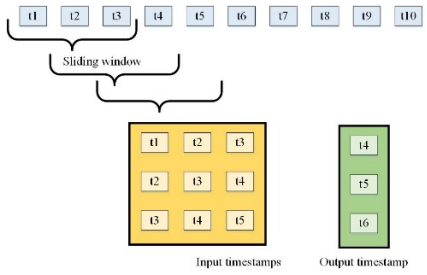

Figure 5. Transform the dataset into regression problem *(Verisetinin regresyon problemine dönüştürülmesi)*

Experimental studies showed that the lowest error rates were obtained when the sliding window size was 5. After transforming the data into the regression problem structure, it was scaled using MinMaxScaler. 33% of the dataset was used for testing and 67% for training. 10% of the train data was used for hyper-parameter optimization. For each compared model to achieve the most successful result, the hyper-parameters of the models were adjusted using grid search.

3.3. Prediction Models *(Tahmin Modelleri)*

XGBoost combines the prediction results of multiple decision tree predictors [18]. It creates new models by correcting errors in the models it creates until the training data is trained correctly [19]. Thanks to its ability to handle missing values, XGBoost can run without requiring data pre-processing. Additionally, it can work quickly on large data sets thanks to its parallel processing ability.

SVM determines a hyperplane that separates different classes in multidimensional space so that the margin between them is maximum [20]. Using the resulting hyperplane, data samples are classified according to their locations. SVM uses support vectors to maximize the margin between classes. Kernel functions are used to determine the decision boundary for the hyperplane [21].

MLP is a model inspired by the human brain's information-processing ability [22]. It has hidden layers consisting of interconnected neurons, except for input and output layers [23]. Hidden layers enable complex relationships to be learned in the dataset. MLP updates the parameters used in the model by calculating the errors between the outputs obtained using backpropagation and the actual values [24].

CNN, generally used in image processing problems, consists of input and output layers, a pooling layer, a convolution layer, and a flatten layer [25]. The input layer receives the data and passes it to the convolution layer. The convolution layer extracts features from the input data using filters [26]. CNN determines patterns in the data by applying convolution along the temporal dimension in time series problems [27]. The flattened layer converts multidimensional feature maps into a 1D vector. Fully connected layers make predictions by learning representations of features [28]. The output layer also presents the prediction.

LSTM is a model that allows long-term dependencies in sequential data to be remembered [29]. LSTM uses gates and memory cells to control the flow of information, enabling information to be selectively forgotten or remembered. LSTM's forget gate decides which information to discard from the memory cell [30]. Thanks to memory cells, LSTM stores information from previous time steps. The outputs of LSTM cells are transferred to the next cell, allowing consecutive data to be processed [31].

3.4. Proposed Model *(Önerilen Model)*

The developed GA-ConvLSTM model enables the hyper-parameters of the ConvLSTM model to be optimized using GA. Proposed GA-ConvLSTM model architecture is seen in Fig. 6.

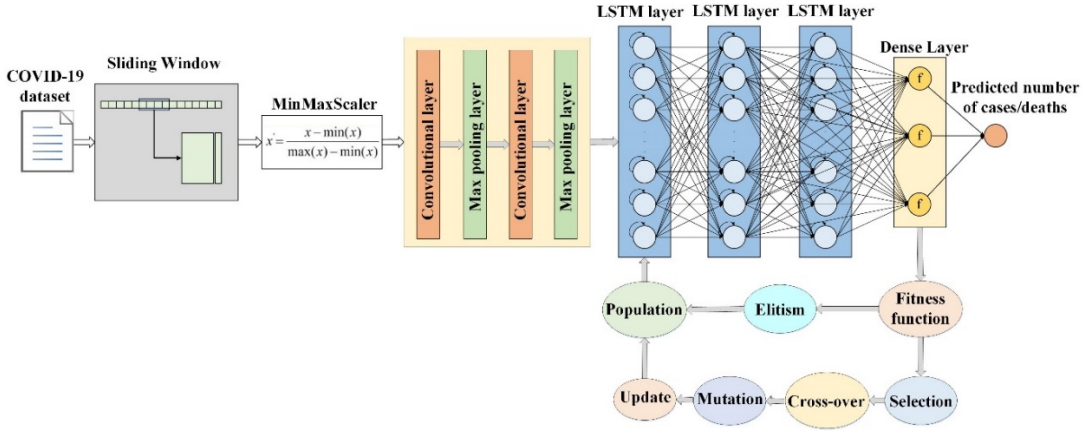

Figure 6. Proposed GA-ConvLSTM model architecture *(Önerilen GA-ConvLSTM model mimarisi)*

CNN is responsible for extracting features in time series data and learning the relationships and patterns between the data. LSTM enables increasing prediction performance by learning long- and short-term dependencies between data. GA uses different hyper-parameters to find combinations with the lowest error rate. GA evolves the model to achieve lower error rates with each iteration using population-based search. GA speeds up the optimization process by

making fewer attempts than Grid Search. GA enables the evaluation of parameter combinations simultaneously by processing different hyper-parameters in parallel. Additionally, GA can determine parameter ranges in more detail than Grid Search. Grid Search works with specified fixed value ranges.

GA-ConvLSTM takes daily COVID-19 cases and the death numbers of SCO member countries as input. After the data is transformed into a regression problem structure using a sliding window, it is scaled using the MinMax scaler. The CNN component uses convolution layers to discover patterns in the data and perform feature extraction. The LSTM component models complex relationships in the data by processing feature maps sent from the CNN. In this way, LSTM learns the data's long and short-term dependencies. GA is used to optimize the hyper-parameters of CNN-LSTM. The hyper-parameter combinations are evaluated as an individual, and selection, crossover, and mutation are performed between individuals. The fitness function enables the determination of hyper-parameters with the lowest error value. When it comes to the GA-ConvLSTM model, the hyper-parameters optimized with GA play a crucial role. For the CNN component, the number of convolutional layers is set at 2, with an activation function of ReLU, a pooling size of 2, and a number of filters at 64. The LSTM component, on the other hand, has 128 neurons, uses the Adam optimizer, has a dropout rate of 0.2, a batch size of 64, and runs for 50 epochs. As for GA itself, the crossover probability is set at 0.8, the population size at 50, the mutation probability at 0.1, and the number of generations at 100.

4. EXPERIMENTAL RESULTS *(DENEYSEL SONUÇLAR)*

In this study, the transmission pattern of COVID-19 in SCO member countries was predicted, and chord diagrams were created for 5- and 14-day periods to determine the transmission of the epidemic among SCO member countries.

4.1. Prediction of the Spread Pattern of COVID-19 in SCO Member Countries (*SCO Üye Ülkelerinde COVID-19'un Yayılma Örüntüsünün Tahmini)*

GA-ConvLSTM was comprehensively checked with CNN, XGBoost, SVM, LSTM, MLP, and CNN-LSTM according to Root Mean Squared Error (RMSE), R-Squared (R^2), and Mean Absolute Error (MAE). Table 2 indicates the case prediction outcomes of RMSE in SCO member countries.

Table 2. The case prediction outcomes according to RMSE in SCO member countries *(SCO üye ülkelerinde RMSE'ye göre vaka tahmin sonuçları)*

Country	XGBoost	SVM	MLP	CNN	LSTM	ConvLSTM	GA-ConvLSTM
China	6477.20	3948.72	3542.48	3751.11	3433.66	3284.30	3018.18
India	8470.77	8286.26	7673.31	8170.56	7515.76	7169.42	5086.96
Iran	2847.86	2821.38	2180.75	2301.22	2113.95	1937.19	1814.97
Kazakhstan	1084.52	751.46	573.25	576.60	568.15	506.54	290.32
Kyrgyzstan	43.44	43.34	41.86	44.09	40.99	39.41	27.13
Pakistan	387.46	383.17	379.19	382.98	365.04	342.89	240.18
Russia	10245.13	4238.41	3654.05	3691.76	3602.80	3290.55	3074.55
Tajikistan	7.14	7.09	5.03	5.98	4.94	4.12	3.93
Uzbekistan	65.44	65.07	49.62	65.10	49.25	48.07	44.66

As seen in Table 2, GA-ConvLSTM is more successful than the other models according to the RMSE. After GA-ConvLSTM, ConvLSTM, LSTM, MLP, CNN, SVM, and XGBoost were successful, respectively. As seen in Table 2, RMSEs are low for Kyrgyzstan, Tajikistan, and Uzbekistan, where the number of cases is low. However, RMSEs are also high for China, India, and Russia, where the number of cases is high. Table 3 indicates the case prediction outcomes of MAE in SCO member countries.

Table 3. The case prediction outcomes in terms of MAE in SCO member countries *(SCO üye ülkelerinde MAE'ye göre vaka tahmin sonuçları)*

Country	XGBoost	SVM	MLP	CNN	LSTM	ConvLSTM	GA-ConvLSTM
China	2538.34	1705.76	1598.44	1547.50	1234.45	1146.30	1113.96
India	4349.47	3771.15	3349.16	3607.85	3039.46	2984.93	2079.43
Iran	1535.99	1561.28	1197.70	1229.09	1251.79	1088.78	1021.40
Kazakhstan	447.46	338.98	252.38	258.74	234.55	214.33	145.16
Kyrgyzstan	20.53	19.47	18.17	22.57	18.30	17.97	12.33
Pakistan	221.48	218.10	217.99	225.87	204.98	194.49	139.23

Russia	5130.96	2136.44	1751.11	2269.28	1858.14	1627.02	1501.16
Tajikistan	4.59	3.07	1.58	2.05	1.65	1.43	1.07
Uzbekistan	41.55	41.52	32.98	45.83	32.04	31.12	28.35

As seen in Table 3, GA-ConvLSTM is more successful than the other models, according to the MAE. After GA-ConvLSTM, ConvLSTM, LSTM, MLP, CNN, SVM, and XGBoost were successful, respectively. As seen in Table 3, MAE values are low for Kyrgyzstan, Tajikistan, and Uzbekistan, where the number of cases is low. However, MAE values are also high for China, India, Iran, and Russia, where the number of cases is high. Fig. 7 and Table 4 show the case prediction results according to R^2 in SCO member countries.

Table 4. The case prediction outcomes according to R^2 in SCO member countries *(SCO üye ülkelerinde R^2'ye göre vaka tahmin sonuçları)*

Country	XGBoost	SVM	MLP	CNN	LSTM	ConvLSTM	GA-ConvLSTM
China	0.902	0.964	0.971	0.967	0.972	0.974	0.976
India	0.986	0.986	0.988	0.986	0.988	0.989	0.994
Iran	0.926	0.927	0.956	0.951	0.959	0.965	0.967
Kazakhstan	0.889	0.946	0.969	0.968	0.969	0.975	0.995
Kyrgyzstan	0.926	0.926	0.931	0.924	0.934	0.939	0.956
Pakistan	0.957	0.958	0.959	0.958	0.962	0.966	0.976
Russia	0.948	0.991	0.993	0.993	0.993	0.994	0.995
Tajikistan	0.605	0.611	0.804	0.724	0.810	0.869	0.965
Uzbekistan	0.961	0.961	0.977	0.962	0.978	0.979	0.982

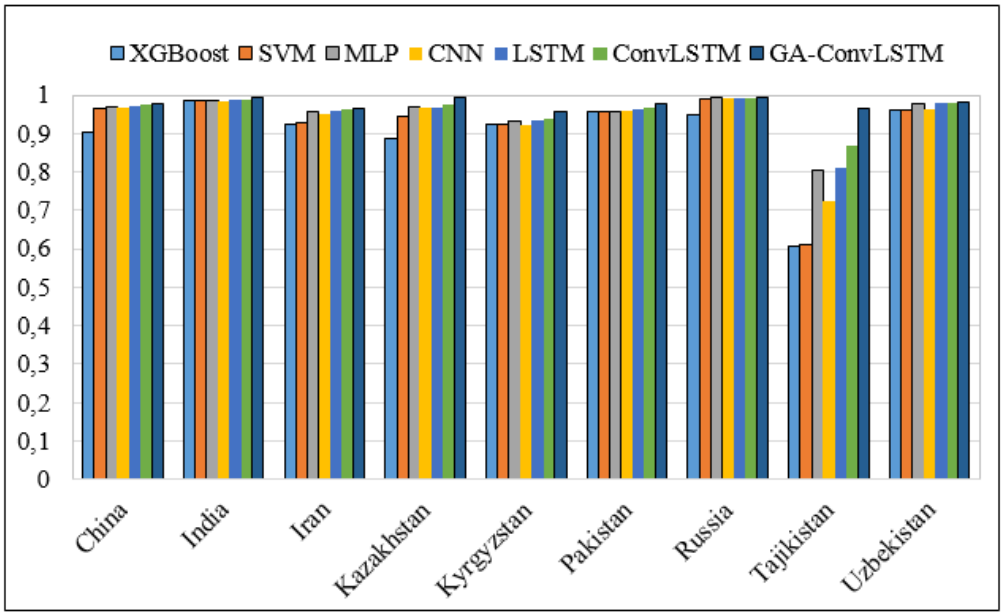

Figure 7. The case prediction results according to R^2 in SCO member countries *(SCO üye ülkelerinde R^2'ye göre vaka tahmin sonuçları)*

Table 4 and Fig. 7 show that all compared models have R^2 above 0.9, except Tajikistan. For Tajikistan, GA-ConvLSTM has 0.965 R^2, and ConvLSTM has 0.869 R^2. Fig. 8 shows the prediction graphs of GA-ConvLSTM for predicting daily COVID-19 cases.

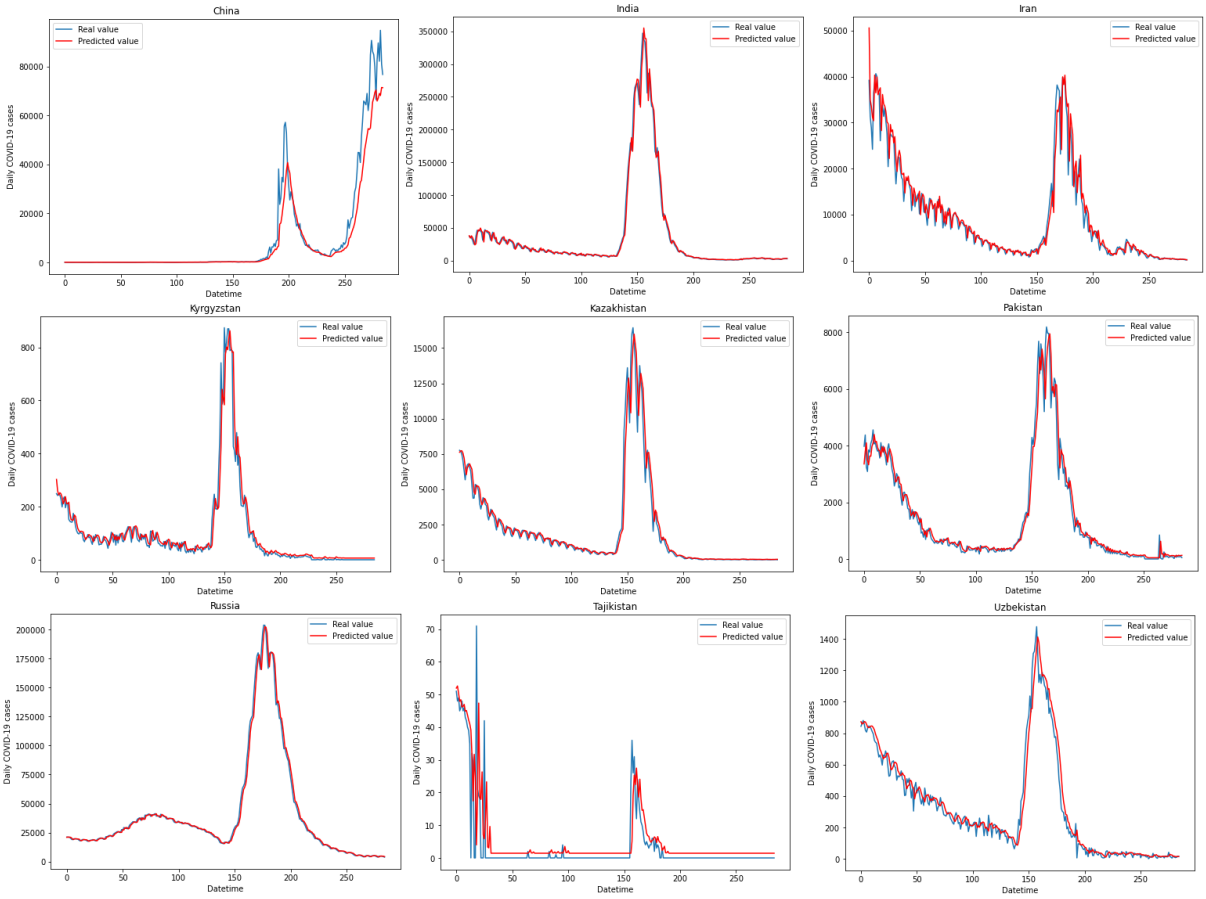

Figure 8. Prediction graphs of GA-ConvLSTM for predicting daily COVID-19 cases *(Günlük COVID-19 vakalarını tahmin etmek için GA-ConvLSTM tahmin grafikleri)*

As seen in Fig. 8, GA-ConvLSTM successfully modelled the changes in daily COVID-19 cases. Table 5 indicates the death prediction outcomes of RMSE in SCO member countries.

Table 5. The death prediction outcomes according to RMSE in SCO member countries *(SCO üye ülkelerinde RMSE'ye göre ölüm tahmini sonuçları)*

Country	XGBoost	SVM	MLP	CNN	LSTM	ConvLSTM	GA-ConvLSTM
China	21.30	19.39	12.79	16.68	12.57	12.40	10.67
India	350.54	313.29	312.02	314.36	311.91	311.14	242.15
Iran	24.75	24.73	24.18	24.50	23.82	23.66	16.80
Kazakhstan	6.66	6.66	6.39	6.60	6.20	5.76	4.86
Kyrgyzstan	0.77	0.76	0.73	0.76	0.72	0.56	0.46
Pakistan	9.10	9.01	8.91	8.97	8.85	8.79	6.52
Russia	25.89	26.00	24.29	26.30	23.98	22.44	15.08
Tajikistan	0.06	0.03	0.02	0.04	0.03	0.02	0.01
Uzbekistan	0.82	0.81	0.78	0.78	0.77	0.75	0.70

As seen in Table 5, GA-ConvLSTM is more successful than the other models according to the RMSE. After GA-ConvLSTM, ConvLSTM, LSTM, MLP, CNN, SVM, and XGBoost were successful, respectively. As seen in Table 5, RMSEs are low for Kyrgyzstan, Kazakhstan, Pakistan, Tajikistan, and Uzbekistan, where the number of cases is low. However, RMSEs are also high for China, India, Iran, and Russia, where the number of cases is high. Table 6 indicates The death prediction outcomes of MAE in SCO member countries.

Table 6. The deaths prediction outcomes according to MAE in SCO member countries *(SCO üye ülkelerinde MAE'ye göre ölüm tahmini sonuçları)*

Country	XGBoost	SVM	MLP	CNN	LSTM	ConvLSTM	GA-ConvLSTM
China	9.17	9.00	5.46	7.04	5.15	5.09	4.46
India	136.40	110.46	109.56	111.08	108.22	108.84	85.30
Iran	15.94	15.89	14.62	14.70	14.96	14.68	10.28
Kazakhstan	3.85	3.84	3.65	3.78	3.69	3.11	2.41
Kyrgyzstan	0.62	0.56	0.54	0.56	0.56	0.56	0.53
Pakistan	5.83	5.35	5.12	5.23	5.12	5.05	3.70
Russia	19.52	19.76	17.80	19.48	17.64	15.50	12.81
Tajikistan	0.05	0.03	0.03	0.03	0.02	0.01	0.01
Uzbekistan	0.56	0.57	0.55	0.51	0.54	0.52	0.47

According to the MAE, GA-ConvLSTM is more successful than the other models. After GA-ConvLSTM, ConvLSTM, LSTM, MLP, CNN, SVM, and XGBoost were successful, respectively. As seen in Table 6, MAE values are low for Kyrgyzstan, Tajikistan, and Uzbekistan, where the number of cases is low. However, MAE values are also high for India. Fig. 9 and Table 7 show the death prediction outcomes according to R^2 in SCO member countries.

Table 7. The death prediction outcomes according to R^2 in SCO member countries *(SCO üye ülkelerinde R^2'ye göre ölüm tahmini sonuçları)*

Country	XGBoost	SVM	MLP	CNN	LSTM	ConvLSTM	GA-ConvLSTM
China	0.917	0.931	0.970	0.949	0.971	0.972	0.982
India	0.332	0.507	0.512	0.502	0.513	0.516	0.623
Iran	0.977	0.978	0.978	0.978	0.979	0.979	0.994
Kazakhstan	0.963	0.963	0.966	0.964	0.969	0.972	0.975
Kyrgyzstan	0.719	0.730	0.747	0.730	0.755	0.760	0.765
Pakistan	0.880	0.883	0.885	0.883	0.887	0.888	0.980
Russia	0.994	0.994	0.995	0.994	0.995	0.995	0.997
Tajikistan	0.421	0.423	0.425	0.422	0.428	0.429	0.432
Uzbekistan	0.838	0.841	0.854	0.854	0.858	0.865	0.882

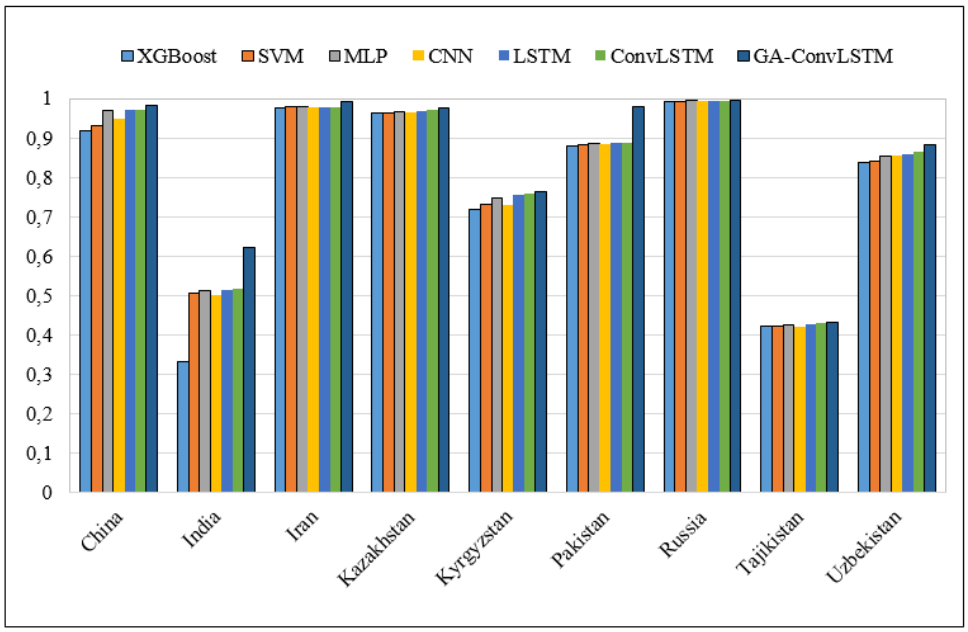

Figure 9. The death prediction results according to R^2 in SCO member countries *(SCO üye ülkelerinde R^2'ye göre ölüm tahmini sonuçları)*

Table 7 and Fig. 9 show that the R^2 value is above 0.9 in all compared models for China, Iran, Kazakhstan, and Russia. However, the R^2 value of all models compared, especially for Tajikistan, is low. Because

the daily death numbers reported for Tajikistan in the dataset are close to 0. Fig. 10 shows the prediction graphs of GA-ConvLSTM for predicting daily COVID-19 deaths.

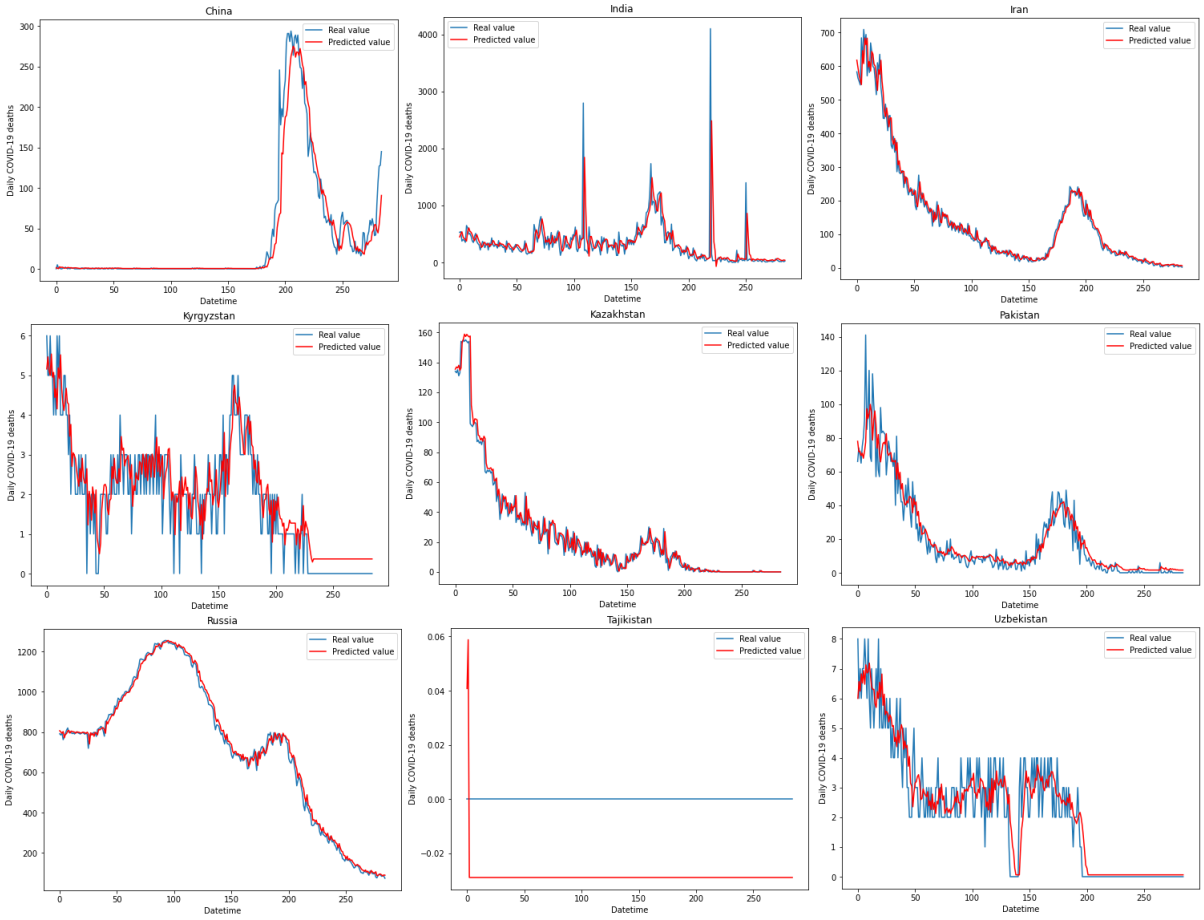

Figure 10. Prediction graphs of GA-ConvLSTM for predicting daily COVID-19 deaths *(Günlük COVID-19 ölümlerini tahmin etmek için GA-ConvLSTM tahmin grafikleri)*

As seen in Fig. 10, in predicting the daily COVID-19 deaths, GA-ConvLSTM has effectively modeled the changes in the daily deaths in countries other than Tajikistan and Uzbekistan. Daily death numbers for Tajikistan and Uzbekistan are reported as very low perhaps countries' failure to declare their death numbers prevented successful modeling of GA-ConvLSTM.

In this study, the GA-ConvLSTM model was created by optimizing the hyper-parameters of the ConvLSTM model with GA. GA-ConvLSTM model was compared with base ConvLSTM, LSTM, CNN, MLP, SVM, and XGBoost, where hyper-parameters were optimized with Grid Search. Experiments showed that the GA-ConvLSTM model was more successful than the other models.

ConvLSTM combines CNN and LSTM to analyze time series data, such as daily COVID-19 data. In the ConvLSTM hybrid model, CNN identifies local patterns in the data, while LSTM discovers time dependencies and long-term relationships in time series data. Grid Search determined the hyper-parameters of base ConvLSTM. However, since Grid Search only tested the values in the specified parameter ranges, successful prediction performance needed to be improved. However, GA enabled the model to have a more successful prediction performance by testing a more extensive hyper-parameter range.

The effectiveness of ConvLSTM over LSTM can be interpreted by CNN's success in detecting local patterns. In this way, ConvLSTM extracts important features in time series data. The effectiveness of ConvLSTM over MLP and CNN can be interpreted by the time dependencies of MLP and CNN in the data and their limitations in capturing complex relationships in time series data. ConvLSTM is more effective than XGBoost and SVM because classical machine learning algorithms cannot model the dynamic structures and dependencies in time-dependent data.

4.2. Determining the Intercountry Spread Pattern of COVID-19 in SCO Member Countries *(SCO Üye Ülkelerinde COVID-19'un Ülkelerarası Yayılma Örüntüsünün Belirlenmesi)*

It has been determined that COVID-19 symptoms emerge in most patients on the fifth day of the illness. It was observed that in some patients, the emergence of symptoms could take up to the 14th day of the disease. In this study, 5 and 14-day incubation times were applied to determine the distribution of COVID-19 in SCO member countries. Moreover, the dates when the number of deaths and cases in each SCO member country peaked were determined, and the date ranges 5 and 14 days before and 5 and 14 days after these dates were determined. Chord diagrams were created according to these dates. The dates when the case counts in each SCO member country were highest, 5 and 14 days before and 5 and 14 days after these dates, are seen in the following table.

Table 8. Dates when each SCO member country has the maximum case counts and incubation periods *(Her SCO üye ülkesinin maksimum vaka sayısına ve kuluçka süresine sahip olduğu tarihler)*

Country	14 days before	5 days before	Peak date	5 days after	14 days after
China	2022/05/14-2022/05/27	2022/05/23-2022/05/27	2022/05/28	2022/05/29-2022/06/02	2022/05/29-2022/06/11
India	2021/04/24-2021/05/06	2021/05/02-2021/05/06	2021/05/07	2021/05/08-2021/05/12	2021/05/08-2021/05/21
Iran	2021/08/04-2021/08/17	2021/08/13-2021/08/17	2021/08/18	2021/08/19-2021/08/23	2021/08/19-2021/09/02
Kazakhstan	2022/01/07-2022/01/20	2022/01/16-2022/01/20	2022/01/21	2022/01/22-2022/01/26	2022/01/22-2022/02/04
Kyrgyzstan	2021/06/16-2021/06/29	2021/06/25-2021/06/29	2021/06/30	2021/07/01-2021/07/05	2021/07/01-2021/07/14
Pakistan	2022/01/15-2022/01/28	2022/01/24-2022/01/28	2022/01/29	2022/01/30-2022/02/03	2022/01/30-2022/02/12
Russia	2022/01/28-2022/02/10	2022/02/06-2022/02/10	2022/02/11	2022/02/12-2022/02/16	2022/02/12-2022/02/25
Tajikistan	2020/05/05-2020/05/18	2020/05/14-2020/05/18	2020/05/19	2020/05/20-2020/05/24	2020/05/20-2020/06/02
Uzbekistan	2022/01/09-2022/01/22	2022/01/18-2022/01/22	2022/01/23	2022/01/24-2022/01/28	2022/01/24-2022/02/06

According to the dates presented in Table 8, the distributions of the transmission pattern of cases among SCO member countries were determined by drawing chord diagrams for the 5-day incubation time in Fig. 11 and the 14-day incubation time in Fig. 12.

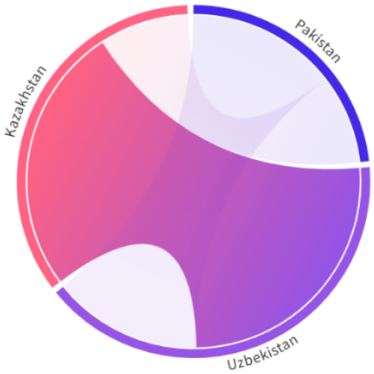

Figure 11. Spread pattern of COVID-19 cases among SCO member countries for 5-day incubation time *(SCO üye ülkeleri arasında COVID-19 vakalarının 5 günlük kuluçka süresi boyunca yayılma örüntüsü)*

Fig. 11 shows that the transmission pattern of cases in Pakistan, Kazakhstan and Uzbekistan showed a similar pattern. As seen in Fig. 11, Pakistan and Uzbekistan have 5 days in common, Pakistan and Kazakhstan have 3 days, and Kazakhstan and Uzbekistan have 9 days in common.

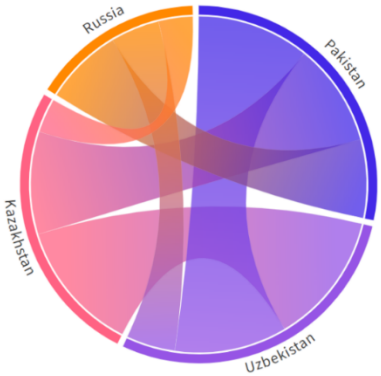

Figure 12. Spread pattern of COVID-19 cases among SCO member countries for 14-day incubation time *(SCO üye ülkeleri arasında COVID-19 vakalarının 14 günlük kuluçka süresi boyunca yayılma örüntüsü)*

Fig. 12 shows that the transmission pattern of cases in Pakistan, Kazakhstan, Russia, and Uzbekistan showed a similar pattern. As seen in Fig. 11, Pakistan and Uzbekistan have 23 days in common, Pakistan and Kazakhstan have 21 days in common, Kazakhstan and Uzbekistan have 27 days in common, Pakistan and Russia have 16 days in common, Russia and Uzbekistan have 10 days in common, and Kazakhstan and Russia have 7 days in common.

Table 9 shows the dates when the number of deaths in each SCO member country was highest, as well as the 5 and 14 days before and 5 and 14 days after these dates.

Table 9. Dates with the highest number of deaths and incubation periods of each SCO member country *(Her SCO üyesi ülkenin en fazla ölüm ve kuluçka döneminin olduğu tarihler)*

Country	14 days before	5 days before	Peak date	5 days after	14 days after
China	2020/04/04-2020/04/17	2020/04/13-2020/04/17	2020/04/18	2020/04/19-2020/04/23	2020/04/19-2020/05/01
India	2021/05/27-2021/06/09	2021/05/05-2021/06/09	2021/06/10	2021/06/11-2021/06/15	2021/06/11-2021/06/24
Iran	2021/08/11-2021/08/24	2021/08/20-2021/08/24	2021/08/25	2021/08/26-2021/08/30	2021/08/26-2021/09/08
Kazakhstan	2021/08/12-2021/08/25	2021/08/21-2021/08/25	2021/08/26	2021/08/27-2021/08/31	2021/08/27-2021/09/09
Kyrgyzstan	2020/07/05-2020/07/18	2020/07/14-2020/07/18	2020/07/19	2020/07/20-2020/07/24	2020/07/20-2020/08/02
Pakistan	2020/11/07-2020/11/20	2020/11/16-2020/11/20	2020/11/21	2020/11/22-2020/11/26	2020/11/22-2020/12/05
Russia	2021/11/05-2021/11/18	2021/11/14-2021/11/18	2021/11/19	2021/11/20-2021/11/24	2021/11/20-2021/12/03
Tajikistan	2020/04/30-2020/05/13	2020/05/09-2020/05/13	2020/05/14	2020/05/15-2020/05/19	2020/05/15-2020/05/28
Uzbekistan	2020/07/15-2020/07/28	2020/07/24-2020/07/28	2020/07/29	2020/07/30-2020/08/03	2020/07/30-2020/08/12

According to the dates presented in Table 9, the distributions of the transmission pattern of deaths among SCO member countries were determined by drawing chord diagrams for the 14-day incubation period in Fig. 13 and the 14-day incubation time in Fig. 14.

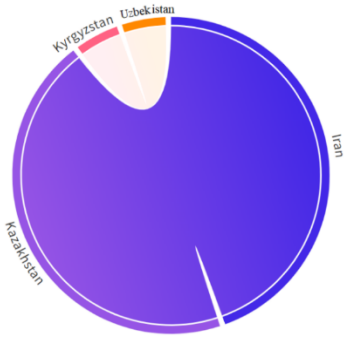

Figure 13. Spread pattern of COVID-19 deaths among SCO member countries for 5-day incubation time *(SCO üyesi ülkelerde COVID-19 ölümlerinin 5 günlük kuluçka süresi boyunca yayılma örüntüsü)*

Fig. 13 shows that for the 5-day incubation time, the transmission pattern of deaths in Iran and Kazakhstan, Kyrgyzstan and Uzbekistan showed a similar pattern. As seen in Fig. 13, Iran and Kazakhstan have 10 days in common and Kyrgyzstan and Uzbekistan have 1 days in common.

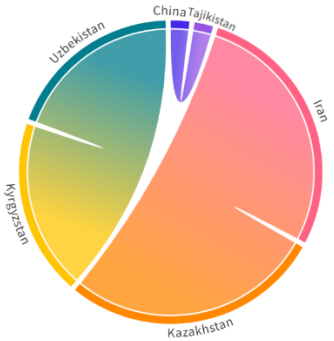

Figure 14. Spread pattern of COVID-19 deaths among SCO member countries for 14-day incubation time *(SCO üyesi ülkelerde COVID-19 ölümlerinin 14 günlük kuluçka süresi boyunca yayılma örüntüsü)*

Fig. 14 shows that for the 14-day incubation time, the transmission pattern of deaths in China, Iran, Kazakhstan, Kyrgyzstan, Tajikistan and Uzbekistan showed a similar pattern. As shown in Fig. 14, Tajikistan and China have 2 days in common, Iran and Kazakhstan have 28 days in common, Kyrgyzstan and Uzbekistan have 19 days in common.

5. CONCLUSIONS *(SONUÇLAR)*

COVID-19 has spread very quickly to almost every part of the world, causing many people to experience severe symptoms and lose their lives. Thanks to the intense and powerful scientific studies conducted in the past four years, many unknown issues regarding the disease have been clarified. With the introduction of vaccines that are effective in protecting against COVID-19 and antivirals that are effective in treatment and the increase in the immunity level of people who have the disease, the morbidity and mortality rates caused by the disease have decreased significantly. However, as COVID-19 spread worldwide, countries' health systems remained inadequate. Due to the complications caused by COVID-19, people were admitted to intensive care units and even lost their lives. Hospital intensive care units were complete, and healthcare personnel and medical equipment were insufficient.

In this study, the hybrid GA-ConvLSTM model was created to predict the daily deaths and cases due to COVID-19 in SCO member countries and to determine the distribution of the spread pattern of COVID-19 among SCO member countries. It aimed to obtain a more successful prediction performance by optimizing the hyper-parameters of the ConvLSTM model using GA. GA-ConvLSTM, XGBoost, SVM, MLP, CNN, LSTM, and ConvLSTM have been extensively tested using RMSE, R^2, and MAE. The study used daily COVID-19 case and death data provided by WHO from 2020/01/03 to 2022/05/31. Comparing models were evaluated. Experiments have shown that GA-ConvLSTM outperforms the compared models based on RMSE, R^2, and MAE.

For daily COVID-19 cases, GA-ConvLSTM had 0.9 R^2 for all SCO member countries. For daily COVID-19 deaths, GA-ConvLSTM had an R^2 value above 0.9 for China, Iran, Kazakhstan, Pakistan, and Russia. However, it had a value of 0.623 R^2 for India, 0.765 R^2 for Kyrgyzstan, 0.432 R^2 for Tajikistan, and 0.882 R^2 for Uzbekistan. The relatively low R^2 values obtained for India and Tajikistan can be expressed as the countries not reporting their actual mortality values. Specifically, the total number of deaths presented for Tajikistan was reported as 9. Unrealistic death numbers prevented the compared models from modeling the dataset successfully. However, the experimental results showed that GA-ConvLSTM can successfully model undulations in the daily deaths and cases of epidemic diseases such as COVID-19.

5 and 14-day incubation periods were used because WHO reported that symptoms caused by COVID-19 appear within the first 5 days but can extend up to the first 14 days in some patients. To determine the distribution of the spread of COVID-19 among SCO member countries, the dates with the highest number of COVID-19 deaths and cases were identified for SCO member countries. Moreover, chord graphs were created by determining 5 and 14 days before and 5 and 14 days after these peak dates.

The spread of cases in Pakistan, Kazakhstan, and Uzbekistan for the 5-day incubation time showed a similar pattern. Pakistan and Uzbekistan have 5 days in common, Pakistan and Kazakhstan have 3 days, and Kazakhstan and Uzbekistan have 9 days in common. For the 14-day incubation time, the transmission pattern of cases in Pakistan, Kazakhstan, Russia, and Uzbekistan showed a similar pattern. Pakistan and Uzbekistan have 23 days in common, Pakistan and Kazakhstan have 21 days, Kazakhstan and Uzbekistan have 27 days, Russia and Pakistan have 16 days, Russia and Uzbekistan have 10 days, and Kazakhstan and Russia have 7 days in common.

For the 5-day incubation time, the spread pattern of deaths in Iran and Kazakhstan, Kyrgyzstan, and Uzbekistan showed a similar pattern. Iran and Kazakhstan have 10 days in common, and Kyrgyzstan and Uzbekistan have 1 day in common. During the 14-day incubation period, the spread pattern of COVID-19 deaths in China, Iran, Kazakhstan, Kyrgyzstan, Tajikistan and Uzbekistan showed a similar pattern. China and Tajikistan have 2 days in common, Iran and Kazakhstan have 28 days in common, and Kyrgyzstan and Uzbekistan have 19 days in common.

REFERENCES *(KAYNAKLAR)*

[1] E. C. Abebe, T. A., Dejenie, M. Y., Shiferaw, and T. Malik, "The newly emerged COVID-19 disease: a systemic review" *Virology journal*, vol. 17, no. 1, 2020.

[2] I. H. Elrobaa, and K. J. New, "COVID-19: pulmonary and extra pulmonary manifestations" *Frontiers in public health*, vol. 9, 2021.

[3] C. Lai, R. Yu, M. Wang,W. Xian, X. Zhao, Q. Tang, and F. Wang, "Shorter incubation period is associated with severe disease progression in patients with COVID-19" *Virulence*, vol. 11, no. 1, pp. 1443-1452, 2020.

[4] A. H. Al-Ani, R. E. Prentice, C. A. Rentsch, D. Johnson, Z. Ardalan, N. Heerasing, and B. Christensen, "Prevention, diagnosis and management of COVID-19 in the IBD patient" *Alimentary Pharmacology & Therapeutics*, vol. 52, no. 1, pp. 54-72, 2020.

[5] A. Ayaz, A. Arshad, H. Malik, H. Ali, E. Hussain, and B. Jamil, "Risk factors for intensive care unit admission and mortality in hospitalized COVID-19 patients" *Acute and critical care*, vol. 35, no. 4, 2020.

[6] S. Singh, G. C. Ambooken, R. Setlur, S. K. Paul, M. Kanitkar, S. S. Bhatia, and R. S. Kanwar, "Challenges faced in establishing a dedicated 250 bed COVID-19 intensive care unit in a temporary structure" *Trends in Anaesthesia and Critical Care*, vol. 36, pp. 9-16, 2021.

[7] Y. Xue, and B. M. Makengo, "Twenty years of the Shanghai Cooperation Organization: Achievements, challenges and prospects" *Open Journal of Social Sciences*, vol. 9, no. 10, pp. 184-200, 2021.

[8] I. Ahmad, "Shanghai Cooperation Organization: China, Russia, and Regionalism in Central Asia" *Initiatives of Regional Integration in Asia in Comparative Perspective: Concepts, Contents and Prospects*, pp. 119-135, 2018.

[9] S. Azizi, "China's Belt and Road Initiative (BRI): The Role of the Shanghai Cooperation Organization (SCO) in Geopolitical Security and Economic Cooperation" *Open Journal of Political Science*, vol. 14, no. 1, pp. 111-129, 2024.

[10] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning

models of LSTM, GRU and Bi-LSTM" *Chaos, Solitons & Fractals*, vol. 140, 2020.

[11] H. Abbasimehr, and R. Paki, "Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization" *Chaos, Solitons & Fractals*, vol. 142, 2021.

[12] H. T. Rauf, M. I. U. Lali, M. A. Khan, S. Kadry, H. Alolaiyan, A. Razaq, and R. Irfan, "Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks" *Personal and Ubiquitous Computing*, pp. 1-18, 2023.

[13] L. Zhou, C. Zhao, N. Liu, X. Yao, and Z. Cheng, "Improved LSTM-based deep learning model for COVID-19 prediction using optimized approach" *Engineering applications of artificial intelligence*, vol. 122, 2023.

[14] C. C. Ukwuoma, D. Cai, M. B. B. Heyat, O. Bamisile, H. Adun, Z. Al-Huda, and M. A. Al-Antari, "Deep learning framework for rapid and accurate respiratory COVID-19 prediction using chest X-ray images" *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 7, 2023.

[15] A. Al-Rashedi, and M. A. Al-Hagery, "Deep learning algorithms for forecasting COVID-19 cases in Saudi Arabia" *Applied Sciences*, vol. 13, no. 3, 2023.

[16] S. Solayman, S. A. Aumi, C. S. Mery, M. Mubassir, and R. Khan, "Automatic COVID-19 prediction using explainable machine learning techniques" *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 36-46, 2023.

[17] M. Kim, and H. Kim, "A Dynamic Analysis Data Preprocessing Technique for Malicious Code Detection with TF-IDF and Sliding Windows" *Electronics*, vol. 13, no. 5, 2024.

[18] M. A. Muslim, and Y. Dasril, "Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning" *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 5549-5557, 2021.

[19] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm" *Interactive Learning Environments*, vol. 31, no. 6, pp. 3360-3379, 2023.

[20] A. Rizwan, N. Iqbal, R. Ahmad, and D. H. Kim, "WR-SVM model based on the margin radius approach for solving the minimum enclosing ball problem in support vector machine classification" *Applied Sciences*, vol. 11, no. 10, 2021.

[21] M. Aslani, S. Seipel, "Efficient and decision boundary aware instance selection for support vector machines" *Information Sciences*, vol. 577, pp. 579-598, 2021.

[22] R. Sharma, M. Kim, and A. Gupta, "Motor imagery classification in brain-machine interface with machine learning algorithms: Classical approach to multi-layer perceptron model" *Biomedical Signal Processing and Control*, 71, 2022.

[23] D. D. Oliveira, M. Rampinelli, G. Z. Tozatto, R. V. Andreão, and S. M. Müller, "Forecasting vehicular traffic flow using MLP and LSTM" *Neural Computing and applications*, vol. 33, pp. 17245-17256, 2021.

[24] N. Calik, M. A. Belen, and P. Mahouti, "Deep learning base modified MLP model for precise scattering parameter prediction of capacitive feed antenna". *International journal of numerical modelling: electronic networks, devices and fields*, vol. 33, no. 2, 2020.

[25] R. Shyam, "Convolutional neural network and its architectures" *Journal of Computer Technology & Applications*, vol. 12, no. 2, pp. 6-14, 2021.

[26] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, A. Mehmood, and M. T. Sadiq, "Document-level text classification using single-layer multisize filters convolutional neural network" *IEEE Access*, vol. 8, pp. 42689-42707, 2020.

[27] H. V. Dudukcu, M. Taskiran, Z. G. C. Taskiran, and T. Yildirim, "Temporal Convolutional Networks with RNN approach for chaotic time series prediction" *Applied soft computing*, vol. 133, 2023.

[28] K. Hermann, and A. Lampinen, "What shapes feature representations? exploring datasets, architectures, and training" *Advances in Neural Information Processing Systems*, vol. 33, pp. 9995-10006, 2020.

[29] S. Patil, V. M. Mudaliar, P. Kamat, and S. Gite, "LSTM based Ensemble Network to enhance the learning of long-term dependencies in

chatbot" *International Journal for Simulation and Multidisciplinary Design Optimization*, vol. 11, no. 25, 2020.

[30] J. Duan, P. F. Zhang, R. Qiu, and Z. Huang, "Long short-term enhanced memory for sequential recommendation" *World Wide Web*, vol. 26, no. 2, pp. 561-583, 2023.

[31] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells" *Chaos, Solitons & Fractals*, vol. 146, 2021.