electronics

Article

Machine Learning-Driven Approach for a COVID-19 Warning System

Mushtaq Hussain [1], Akhtarul Islam [2], Jamshid Ali Turi [3], Said Nabi [1,*], Monia Hamdi [4], Habib Hamam [5,6,7,8], Muhammad Ibrahim [9,10,*], Mehmet Akif Cifci [11,12] and Tayyaba Sehar [1]

[1] Department of Computer Science and Information Technology, Virtual University of Pakistan, Lahore 54000, Pakistan
[2] Statistics Discipline, Science, Engineering and Technology (SET) School, Khulna University, Khulna 9208, Bangladesh
[3] Department of Management Information System, University of Tabuk, Tabuk 47512, Saudi Arabia
[4] Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[5] Faculty of Engineering, Moncton University, Moncton, NB E1A3E9, Canada
[6] International Institute of Technology and Management, Commune d'Akanda, BP, Libreville 1989, Gabon
[7] Department of Electrical and Electronic Engineering Science, School of Electrical Engineering, University of Johannesburg, Johannesburg 2006, South Africa
[8] Spectrum of Knowledge Production & Skills Development, Sfax 3027, Tunisia
[9] Department of Computer Engineering, Jeju National University, Jeju 63014, Republic of Korea
[10] Department of Information Technology, University of Haripur, Haripur 22620, Pakistan
[11] Department of Computer Engineering, Bandirma Onyedi Eylul University, Balikesir 10200, Turkey
[12] Informatics, Klaipeda State University of Applied Sciences, LT-91274 Klaipeda, Lithuania
* Correspondence: said.nabi@vu.edu.pk (S.N.); ibrahimmayar@uoh.edu.pk (M.I.)

Citation: Hussain, M.; Islam, A.; Turi, J.A.; Nabi, S.; Hamdi, M.; Hamam, H.; Ibrahim, M.; Cifci, M.A.; Sehar, T. Machine Learning-Driven Approach for a COVID-19 Warning System. *Electronics* **2022**, *11*, 3875. https://doi.org/10.3390/electronics11233875

Academic Editors: Taiyong Li, Wu Deng, Jiang Wu and Juan M. Corchado

Received: 19 August 2022
Accepted: 14 November 2022
Published: 23 November 2022

Abstract: The emergency of the pandemic and the absence of treatment have motivated researchers in all the fields to deal with the pandemic situation. In the field of computer science, major contributions include the development of methods for the diagnosis, detection, and prediction of COVID-19 cases. Since the emergence of information technology, data science and machine learning have become the most widely used techniques to detect, diagnose, and predict the positive cases of COVID-19. This paper presents the prediction of confirmed cases of COVID-19 and its mortality rate and then a COVID-19 warning system is proposed based on the machine learning time series model. We have used the date and country-wise confirmed, detected, recovered, and death cases features for training of the model based on the COVID-19 dataset. Finally, we compared the performance of time series models on the current study dataset, and we observed that PROPHET and Auto-Regressive (AR) models predicted the COVID-19 positive cases with a low error rate**.** Moreover, death cases are positively correlated with the confirmed detected cases, mainly based on different regions' populations. The proposed forecasting system, driven by machine learning approaches, will help the health departments of underdeveloped countries to monitor the deaths and confirm detected cases of COVID-19. It will also help make futuristic decisions on testing and developing more health facilities, mostly to avoid spreading diseases.

Keywords: time series; forecasting; COVID-19; machine learning; warning system; PROPHET; health

1. Introduction

The introduction should briefly place the study in a broad context and highlight why it is important. Coronaviruses (termed as CoVs) are a group of viruses that infect birds and mammals. They also cause widespread diseases, such as Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), Middle East Respiratory Syndrome, Coronavirus (MERS-CoV), and the 2019 Novel Coronavirus (2019-nCoV, also known as COVID-19) [1].

The COVID-19 outbreak started in Wuhan Province, China, in late December 2019, and patients died due to organ dysfunction syndrome [2,3]. The Chinese government reported that the causative pathogen was a coronavirus identified by genomic sequencing and electron microscopy. The virus originated in bats and was eventually transmitted to humans via an intermediate host (probably the raccoon dog) [4].

In many instances, the major symptoms of COVID-19, were fever, cough, and shortness of breath, resembling those of seasonal influenza [5]. Since it was first recognized, COVID-19 has spread exponentially across the world. According to world meters, as of the 5 October 2022, 11:13 GMT, the COVID-19 pandemic has affected 228 countries and territories worldwide and two international conveyances with 624,430,759 confirmed 6,553,537 deaths, 604,462,445 recovered cases, and 13,414,777 active cases. Even after the substantial efforts made by scientists and scholars worldwide, COVID-19 has no standard cure method through vaccines [6]. Nonetheless, some of the patients of the COVID-19 pandemic are recovering with the aid and the proper administration of antibiotic medications. Right now, the world needs a speedy solution to tackle the further spread of COVID-19. The emergence of COVID-19 infection has forced researchers from various disciplines to explore this novel virus. Machine learning is a branch of AI that essentially focuses on the production of systems that can learn from trained examples and improve without being explicitly programmed [7]. Machine Learning has played a significant part in many fields, e.g., medical care [8], medical informatics [9], and agriculture [10]. Moreover, different ML models have optimization problems and mathematical techniques [11] that can be used to solve these problems. Similarly, ML algorithms have been used to understand and detect COVID-19, which has alleviated the enormous strain on healthcare systems while offering the most effective diagnostic and prognostic tools for COVID-19 pandemic patients.

The COVID-19 pandemic has seriously affected population health across the globe. The forecasting of COVID-19 research efforts has become critical and, with the advancement of computers and software technology, AI has played a vital role in the healthcare system in the detection and clinical diagnosis of diseases. Much research has focused on the treatment, prediction, as well as the formulation of COVID-19 [12].

A variety of ML techniques have been used to predict the mortality risk of COVID-19 patients. Pourhomayoun et al. [13] have used a support vector machine (SVM), artificial neural network (ANN), random forest (RF), decision tree, logistic regression, and K-nearest neighbor to detect the mortality risk of patients due to COVID-19 infection.

Researchers have also focused on modeling, predicting, and forecasting the spread of COVID-19 based on the time-series recorded data of COVID-19. Sarkar et al. [14] proposed the SARII mathematical model to forecast the dynamic transmission of COVID-19, which was an extended version of the SEIR model. The proposed model is based on six dynamics behaviors, i.e., susceptible, asymptomatic, recovered, infected, and quarantined. An alternative version of the SEIR model was proposed by Abbasi et al. [15], named SQEIAR, which considered the two parameters, quarantined individuals and asymptomatic individuals, to describe COVID-19. Similarly, Ribeiro et al. [16] used applied regression models ARIMA, cubist regression, random forest, SVR, rigid regression, and stacking-ensemble learning for the forecasting of COVID-19 cases in Brazil. According to the obtained results, researchers observed that SVM regression and stacking-ensemble are better in forecasting. Apart from linearity, many researchers have used nonlinearity structures to predict COVID-19 cases. Peng et.al [17], used the SVR with a Gaussian kernel and claimed the better prediction of COVID-19 cases.

Various ML algorithms and deep learning techniques have been utilized in the literature to compute COVID-19. Different methodologies, including long short-term memory (LSTM), ARIMA, and JNARNN, were built using ML and deep learning [18,19]. However, this research did not analyze the performance model's link between positive cases and input features. This study explores the performance time series of the ML model on the COVID-19 dataset and identifies the characteristics most closely associated with positive COVID-19 cases. The prognosis of death and verified detection cases (of COVID-19) is a weekly concern for numerous nations. The current dataset displayed daily confirmations and death cases in various nations; however, such a dataset was not ordered weekly, and not all the observations of the existing dataset were available (many attributes were missing), which must be fixed.

For this research, we utilized the COVID-19 virus dataset which is available online for research purposes. In this dataset, the COVID-19 observations, such as confirmed cases, death cases, and recovery cases, are organized by date for many U.S. states. In addition, the dataset comprises data from 10 March 2020–29 March 2020.

In the subsequent section, the relevant literature will be presented. Materials and methods will be discussed in the following section. Section 4 will next exhibit the experiment described in this paper. The final section will present issues, difficulties, and conclusions.

Consequently, a new COVID-19 warning system can be constructed using an ML technique, for instance, by comparing the performance of the "Time Series ML Algorithm" to the "Statistical Time Series Model." This would aid healthcare professionals and physicians in diagnosing COVID-19 pandemic patients and recommending recent anti-bodies medication (for recovery). Additionally, implementing the time series ML algorithm (to avoid the pandemic) would limit the spread of the COVID-19 pandemic in situations where human-to-human interaction is inevitable.

The main objective of this study is twofold. First, to estimate the weekly-confirmed instances of COVID-19 and potential deaths using patient history data in different nations; second, to create a warning system that can evaluate the performance of various ML time series models with statistical time series models. Since the present pandemic data (of COVID-19) is only available in abundance, the investigation of the following research issues constitutes a significant contribution to the body of research.

1. What are the appropriate time series models for predicting patients infected with the COVID-19 virus?
2. What will the number of death cases and possible confirmed detected cases in the coming weeks be, based on various given features in the form of data at input, such as date, country, detected cases, and deaths?

2. Related Work

False positives are often observed in research when the literature of papers and methodologies are considered. As a result, it is essential to develop methods that make becoming faster and gaining more accurate results easier while simultaneously reducing human-induced errors. This section of the study examines the procedures and methodologies of the literature.

To help policymakers manage the disease and related emerging situations, the authors [6] devised a COVID-19 pandemic prediction tool. This tool was based on data from patients from India to keep track of infected cases. They assumed that control strategies, such as quarantines and lockdowns, would prevail. Their results suggested that India could experience the end of the pandemic by March 2021. The model was developed on the basis of least-square fitting of the novel coronavirus behavior and is based on real-world data for a particular time, but the least-square technique was unable to address the overfitting issue.

Ganiny et al. [19], based on the Indian perspective, employed an autoregressive integrated moving average (ARIMA) model that utilizes the past trajectory and forecasts the future evolution of COVID-19. Their model predicted the number of infected cases, active cases, recoveries, and deaths due to the pandemic. They suggest some robust control strategies to mitigate the spread of COVID-19.

Wadhwa et al. [20] predicted recovery, death, and active cases of COVID-19 patients by applying a linear regression technique from Indian records. Their model predicted the extension of lockdown based on empirical results. They applied graphical tools to showcase the predicted results more comprehensively.

Saima et al. [21] studied the trends of COVID-19 in the eastern Mediterranean regions using a statistical method. Their analysis revealed that Iran was the worst affected country, followed by Saudi Arabia and Pakistan. The United Arab Emirates and Saudi Arabia had the lowest fatality rates, while Pakistan and Lebanon had moderate fatalities. They suggest following strict recommendations, based on epidemiological principles, to reduce COVID-19 cases.

Yadav et al. [22] utilized ML tools to analyze the transmission and growth rates of COVID-19 patients across various countries. They further correlated the weather conditions and the COVID-19 cases and predicted the pandemic's end time frame. They exploited support vector machine algorithms (SVM) for these tasks.

The model demonstrated a high accuracy of 98% and proved its efficacy compared to recent forecasting models.

Ricardo, M. A. V., et al. [23] applied reduced-space Gaussian process regression, related to chaotic dynamical systems, to forecast COVID-19-related deaths from 82 days' data. Empirical results asserted that Gaussian mean-field models were able to be employed to gather information regarding the pandemic's spread, recovery, and fatality rates. They also devised a reduced-space Gaussian process regression model to estimate when saturation would be achieved in the USA (regarding the pandemic).

Hamzah et al. [24] also introduced a predictive model based on the Corona Tracker (an online platform for reliable analysis, and statistics, of COVID-19) to forecast COVID-19-related cases, recoveries, and deaths. They exploited susceptible exposed infectious recovered (SEIR) modeling to keep track and predict COVID-19 outbreaks.

Moreover, they classified and analyzed the queried news into positive and negative categories based on the people's sentiments. Furthermore, they tried to understand the economic and political impacts of COVID-19. Overall, they observed that more negative articles exist in the given domain than positive ones.

Mahajan et al. [25] utilized a compartmental epidemic model (SIPHERD) to predict COVID-19 active, confirmed, and death cases in India. Their results show that social-distancing measures, increasing daily tests, and strict lockdown significantly impacted the reduction of COVID-19.

Moreover, the authors [26] employed the SEIR model to extract the epidemic curve from the epidemiological data of COVID-19. They also applied an AI framework to forecast the disease. Their model was trained using 2003 SARS data. They predicted that the epidemic peak would gradually rise and then fall in China. Their dynamic model demonstrated its efficacy in forecasting COVID-19 epidemic sizes and peaks.

Shahid et al. [27] also presented a COVID-19 time series prediction model by employing LSTM, bidirectional long short-term memory (Bi-LSTM), support vector regression (SVR), and autoregressive integrated moving average model (ARIMA) techniques. They evaluated their model using the R square score, root mean square error (RMSE), and mean absolute error indices (MAEI). Their results suggest that the Bi-LSTM model is the best-suited model for such pandemic predictions, especially for better management and planning**.**

According to Xue et al. [28], in 2020, COVID-19 still needed to be completely understood. The authors believe that scientists and doctors were struggling to find COVID-19 instances. COVID-19 tests include viral tests to determine whether the patients are infected and antibody tests to determine if the patients have been infected before. The paper aims to reduce the false positive rate.

Various ML algorithms and deep learning techniques have been utilized in the literature to compute COVID-19. Different methodologies, including LSTM, ARIMA, and JNARNN, were built using ML and deep learning [29,30]. However, this research did not analyze the performance model's link between positive cases and input features. This study explores the performance time series of the ML model on the COVID-19 dataset and identifies the characteristics most closely associated with positive COVID-19 cases. The prognosis of death and the verified detection cases (of COVID-19) is a weekly concern for numerous nations.

Mansour et al. [17] provide a unique unsupervised DL-based variational autoencoder model for COVID-19 identification and classification. They utilized the Adagrad approach to modify the Inception v4 model hyperparameters to improve the classification performance.

Accordingly, an intelligent COVID-19 positive cases detection system was developed in this study using ML time series algorithms. The main task of the proposed architecture is to provide an efficient method for predicting COVID-19 patient-positive cases. Based on the performance of the proposed system, the health department can find daily positive cases in different areas of the country.

3. Materials and Methods

We studied COVID-19 data from many countries across the globe, which are freely accessible online for research purposes. The forecasting system of the current study, driven by machine learning approaches, will help the health departments of underdeveloped countries to monitor the death and confirmed cases of COVID-19. It will also help make futuristic decisions on testing and developing more health facilities, mostly to avoid spreading diseases. Using an ML technique, for instance, by comparing the performance of the "Time Series ML Algorithm" to the "Statistical Time Series Model", would aid healthcare professionals and physicians in diagnosing COVID-19 pandemic patients and recommending recent anti-bodies medication (for recovery).

The dataset contains information regarding the COVID-19 virus in various countries and cities. In addition, the dataset contains daily record data for many countries/cities. The dataset includes records for other countries and cities beginning on 11 March 2020 and ending on 29 March 2020. Using the time series models depicted in Figure 1, we followed the steps below to explore COVID-19 predictions for the subsequent week.

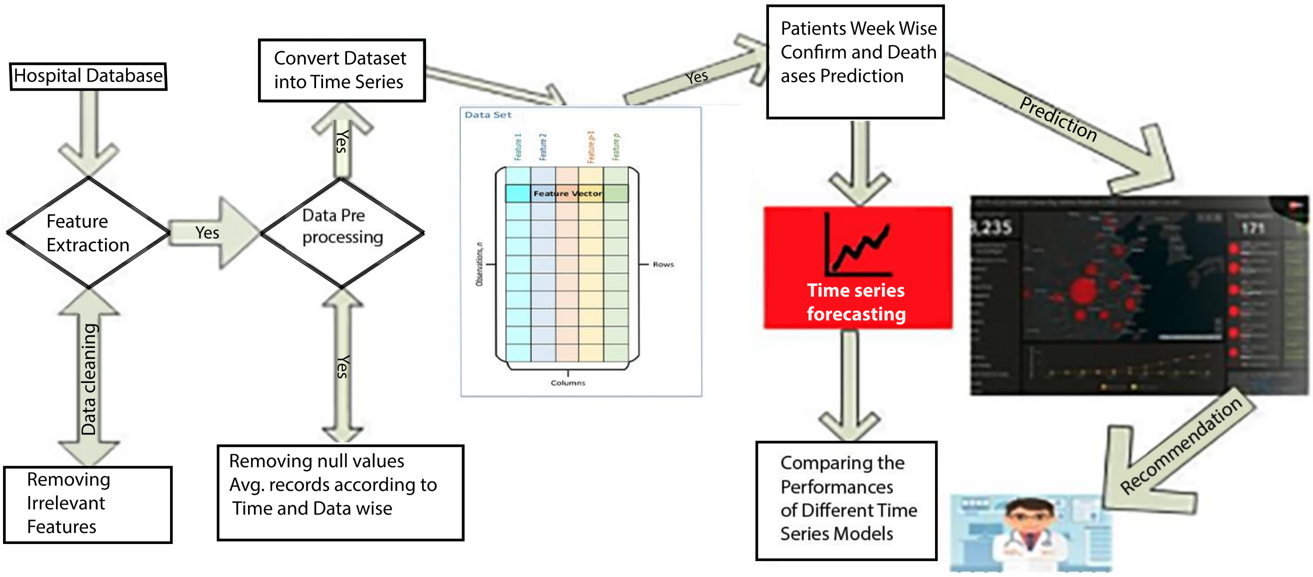

Hospital Database
Feature Extraction
Yes
Data Pre processing
Yes
Convert Dataset into Time Series
Yes
Data cleaning
Removing Irrelevant Features
Yes
Removing null values Avg. records according to Time and Data wise
Data Set
Feature 1
Feature 2
Feature p-1
Feature p
Feature Vector
Observations, n
Rows
Columns
Yes
Patients Week Wise Confirm and Death ases Prediction
Prediction
Time series forecasting
Comparing the Performances of Different Time Series Models
Recommendation

Figure 1. Proposed framework.

Figure 1 depicts the data collection process, followed by the feature extraction procedure. If there are irrelevant characteristics, they are eliminated. Following this, we transfer the data into the preprocessing procedures, eliminating null values and transforming the data into a time series. The whole dataset is disseminated to the Week

Wise section, where the death and survival instances are verified. Either it is moved to time series forecasting to be compared with several models or the prediction is moved to an expert.

Data description: Table 1 provides specifics on the features extracted from the dataset. It describes how to exclude the pertinent features/attributes of the dataset to construct time-series prediction models for the COVID-19 cases of different counties, including laboratory-confirmed cases, recovered cases, and deaths in the following week or hours. Extracted features include date, state, country, confirmed cases, recovered cases, deaths, and population.

Table 1. Attribute description of the COVID-19 dataset.

Attribute	Description
Date	The date on which the data was recorded
State	State from where the COVID-19 patients belong
County	Count y from where COVID-19 patients belong
Confirmed	Number of confirmed COVID-19 patients
Recovered	Number of recovered COVID-19 patients
Deaths	Number of deceased COVID-19 patients
Population	The total population in the state

Data preprocessing: Before applying time series forecasting models, we removed missing values by applying the median imputation method. We checked the dataset's stationarity property, which is a relevant feature of time series and that confirms the data suitability, specifically for the time series-related issues. Moreover, many time series models only work on stationary data. Stationarity data has a constant up and down movement, and it also has a constant mean and variance. Since the data used in this study is not stationary because the status of the time series of COVID-19 dataset statistics and the properties that changed over time, i.e., MSE and RMSE, we applied Python differencing techniques (to convert the data into the stationary format). For that reason, we subtracted the current value from the next value. Then we used a partial autocorrelation function (PACF) plot to check for stationary properties in the dataset.

Machine learning techniques: In order to apply time series forecasting models to predict next week and hours of confirmed detected and death cases (based on ML), there are multiplication classification techniques for the datasets (such as PROPHET, auto regression, ARIMA, and LSTM). The technique that we use (in this study) consists of different ways to extract and classify features that help predict futuristic issues. The details of the current study ML model are below.

3.1. PROPHET

Without a high level of expertise, the prediction is difficult for ML researchers because it often needs more skills than they possess in terms of programming language. PROPHET is the Facebook data ML technique, which is open source and available in Python and R languages. Researchers can use this tool without any programming skills. It is an algorithm that is used to build a forecasting model for time series data based on an additive approach. The algorithm was first introduced in 2017, and unlike the traditional time series technique, PROPHET tries to fit additive regression (called curve fitting) [31]. PROPHET is very robust within missing data, handles outliers very well, and is best with time series, strong seasonal effects, and several seasons of historical data [32].

3.2. Autoregressive Model (AR)

The automotive regressive (AR) model predicts the next timestamp value by applying regression and previous values. The analysis of nature, economics, stock markets, and other time series-based systems frequently employs the AR model. AR models provide a number of advantages over other time series models, such as their ability to operate on continuous variables. The AR model predicts the next timestamp value by regressing and using previous values. The AR model is commonly used in analyzing nature, economics, stock markets, and other time series-based processes. AR models have some advantages over other time series models; for example, they work on continuous values.

3.3. Auto Regressive Integrated Moving Average (ARIMA)

We used the auto regressive integrated moving average (ARIMA) model as our dataset was non-stationary, whereas the integration part was "Stationized" the time series.

3.4. Long Short-Term Memory

LSTM is used to solve the learning models for recurrent neural networks to provide promising results on many tasks, such as constructing prediction and language models [33]. It solves challenging tasks (large time-lags) that recurrent network algorithms [34] have never solved. LSTM is used to solve the learning models for recurrent neural networks to produce promising results on a variety of tasks, including building prediction and language models [33]. It solves complex tasks (long time lags) that have never been solved by recurrent network algorithms [34].

4. Results and Discussion

In this study, we used time-series ML models rather than other ML models with no time dimension. The time series forecasting model is based on previously observed values [35].

We discovered a positive but weak correlation (r = 0.032) between "Confirmed" and "Recovered" cases in Table 2. Every day, COVID-19 confirmed and recovered patients move in the same positive direction, while the increase in confirmed patients is very high compared to recovered patients of COVID-19. The maximum number of COVID-19 patients is 21,873, whereas the maximum number of recovered patients with COVID-19 is 10. Between deaths and confirmed cases, we found a strong positive correlation (r = 0.796). As COVID-19 is confirmed and detected, the death rate of COVID-19 patients also increased, with the maximum reported death of 281. The variables of "Population" and "Confirmed" also showed a positive but weak correlation (r = 0.154), which means that as the population increases, the confirmed patients also increase.

Table 2. Correlation analysis and descriptive statistics for different variables (descriptive statistics).

	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**	**Correlation with "Confirmed" (r)**	***p*-Value**
Confirmed	16,585	0.00	21,873.00	22.79	323.776	1	<0.001 ***
Recovered	16,585	0.00	10	0.008	0.1625	0.032	<0.001 ***
Deaths	16,585	0.00	281	0.311	4.0454	0.796	<0.001 ***
Population	16,452	88	39,512,223.0	387,142.67	1,997,001.27	0.154	<0.001 ***

Note: *** $p < 0.001$.

Table 3 demonstrates that when the number of confirmed cases grows, the death rate will increase by 0.010 times. We determined that a one-unit (100,000) increase in population contributes 0.016 times to the death factor for the "Population" variable. Here, R^2 equals 0.640, which shows the amount of a dependent variable's variance explained by the independent variables in a regression model. The model's inputs can explain approximately 64 percent of the observed variation. We also benefited from the same-day confirmed cases to predict death, though, for one day of COVID-19 cases, it can be concluded that 2.2% died while 75.9% recovered and 21.9% were still in isolation or being treated at the last follow-up.

Table 3. Multiple linear regression estimation considering "Deaths" as a dependent variable.

	Unstandardized Coefficients		***p*-Value**	**95.0% Confidence Interval for B**	
	B	**Std. Error**		**Lower Bound**	**Upper Bound**
(Constant)	0.032	0.019	0.094	−0.006	0.07
Confirmed	0.01	0.000	0.000	0.01	0.01
Population	0.016	0.001	0.000	0.014	0.018

Note: B (Biases) is a training parameter that needs to be optimized during the training process. *p*-value is the probability value corresponding to the likelihood of gaining a data value.

Non-stationary data represent that the mean and the standard deviation are not constant for given data during the time curve described by [35]. With the help of data visualization, we can understand the pattern, trend, and correlation between the variables for COVID-19 predictions, based on the time series-based ML approach.

Figure 2 shows how the confirmed COVID-19 patients and the dead COVID-19 patients, from 10 March 2020–28 March 2020. The figure also shows a sharp increase in both confirmed and unconfirmed deaths of COVID-19 patients during this time. As the number of confirmed COVID-19 patients rises, so does the death rate among these patients. The disturbing fact is that the death toll on March 26 exceeded one thousand.

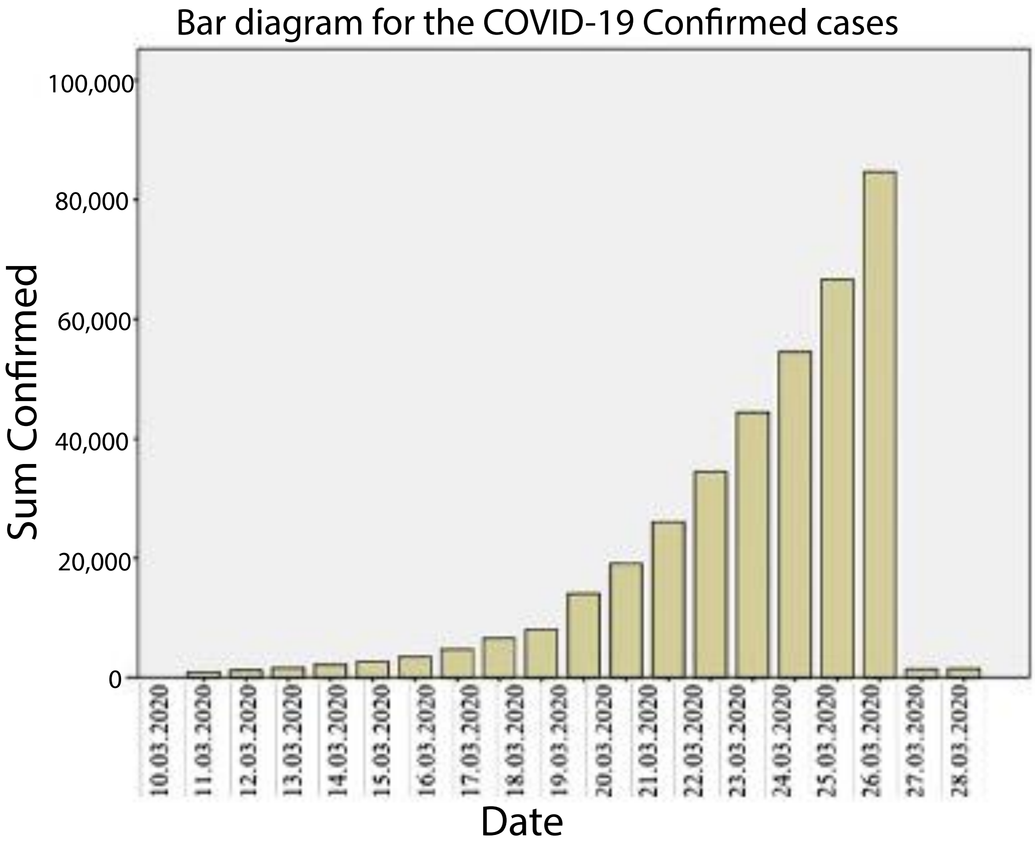


Figure 2. Bar diagram shows the daily confirmed cases of COVID-19 in current study dataset.

Figure 3 depicts the confirmed COVID-19 patients and deceased COVID-19 patients between March 10, 2020, and March 28, 2020. The figure also shows a sharp increase in confirmed and unconfirmed COVID-19 patient deaths over this period. The COVID-19 epidemic affects all sectors of the population but disproportionately negatively impacts the most disadvantaged social groups.

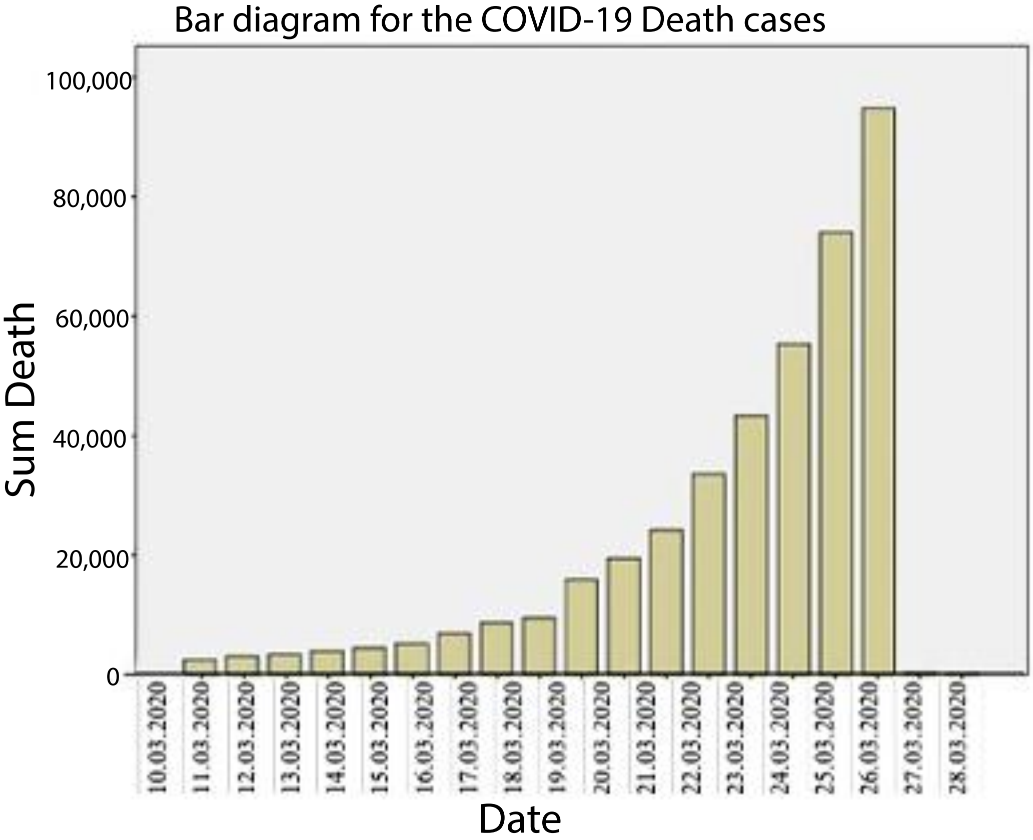


Figure 3. Bar diagram shows the daily death cases due to COVID-19 in current study dataset.

Figure 4 illustrates the daily confirmed cases during COVID-19. In addition, it implies that the rate of confirmed patients jumped dramatically after 21 March 2020. From the beginning of 11 March 2020–21 March 2020, as confirmed in Figure 4, the number of COVID-19-confirmed cases increased gradually. From 21 March 2020, there was an alarming increase in the number of confirmed COVID-19 cases. After 23 March 2020, the number of confirmed COVID-19 cases surged. Globally, there was an increase in the world of confirmed COVID-19 cases. Figures 3 and 4 demonstrate a quick decline in the amount of confirmed and fatal cases due to the lack of data in some date-specific datasets.

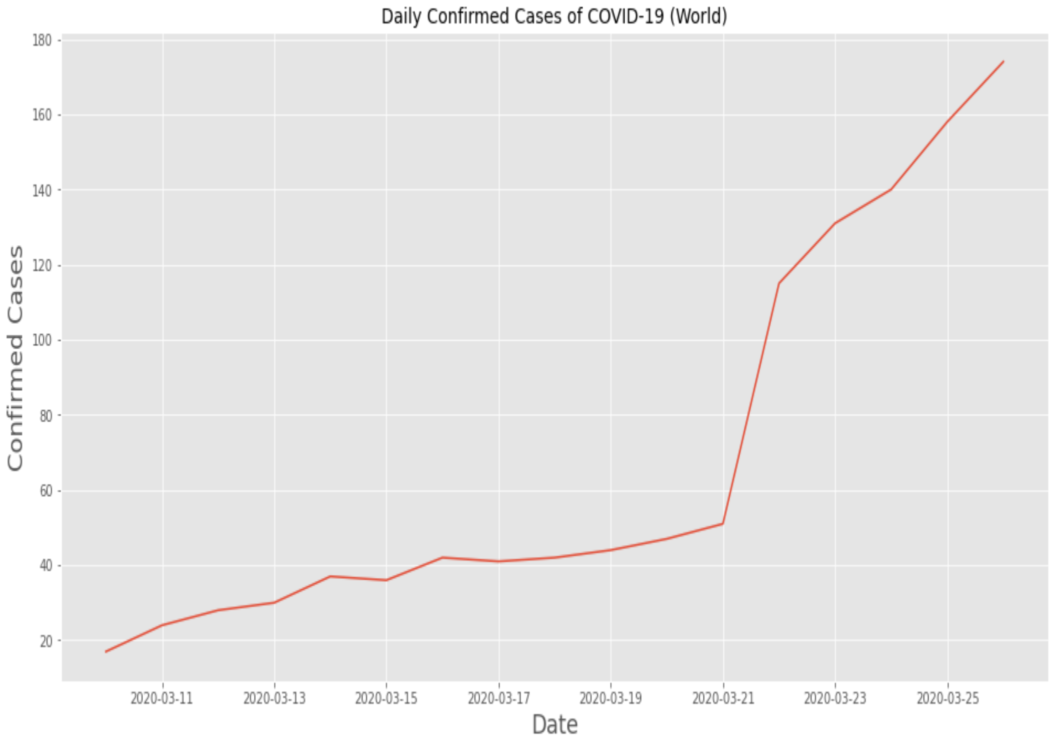


Figure 4. Visualization of daily confirmed cases during COVID-19 disease.

4.1. Design of the Predictive Models and Experimental Setup

In this study, for the COVID-19 predictions, we used the available data for research purposes online [36–41]. From different countries' data from the date 11 March 2020–29 March 2020, we used different Python modules to visualize and describe the data and then trained ML time series models with 80% of data and tested on 20% of data. The PROPHET is a simple time-series algorithm that gives a quick result during the initial stage of modeling. Therefore, we used a Python module to implement the PROPHET algorithm.

Nevertheless, to implement the PROPHET algorithm in Python, the dataset must have NAN-values (or missing values) in the features column; therefore, we leave some NaN values in the dataset. Next, we changed the date column into a date index. We trim the current study dataset to keep only those rows that fall within the period from 10 March 2020–31 March 2020. Before running the model, we rename the dataset column into two columns that are ds (Date) and y (confirmed cases). For LSTM, we also converted the data into three dimensions because LSTM only works on three-dimensional data.

The LSTM model has been enhanced with four layers: the first two layers have 40 neurons each, the third layer has 25 neurons, and the final layer has one neuron. In addition, the model is employed as an Adam optimizer, which utilizes square errors as loss functions. Before applying the AR model, we checked the stationary properties of the dataset because the AR model only works on stationary data. Since the current study dataset was non-stationary (in nature), we took severe differences and finally obtained the stationary data. Nevertheless, the AR and ARIMA models were trained on default values.

4.2. Performance Measurements

To analyze the performance of time series ML models, we employed root mean square error (RMSE) performance metrics. RMSE is the square root of the mean squared error (MSE), which is converted to RMSE by taking its square root. MSE is measured in square units of the target variable, whereas RMSE is measured in the same units. MSE penalizes greater errors more harshly than the squared loss function from which it is derived, and it penalizes greater errors more harshly due to its structure. It measures the deviation between the value predicted by the ML model and the actual value. We predicted the confirmed identified cases and deaths of individuals needing medical care. Consequently, if the number of confirmed deaths is considerable (according to the provided country's statistics), the health agency can take the necessary instances to reduce COVID-19. This is a time series ML-based problem, and the dataset is freely available for research purposes [36].

We investigated the performance of the AR, PROPHET, ARIMA, and LSTM time series classifiers to identify the optimal ML time series models for forecasting the daily confirmed detected cases in different countries during COVID-19. The input variables were daily confirmed cases and death cases of the patient in different countries. The output variables were next week's confirmed and death cases of the patients.

The PROPHET time series model was trained using 80% COVID-19 training data. Then we tested trained PROPHET classifiers on 20% test data. The model received an RMSE value of 29.07, and the results are shown in Table 4. Whereas Figure 5 shows a comparison between the actual confirmed cases and predicted confirmed cases of PROPHET algorithms.

Table 4. ML time series models' performance on COVID-19 test data.

Models	RMSE (Root Mean Square Error)
PROPHET	29.07
LSTM	130.11
AR model	10.49
ARIMA Model	34.75

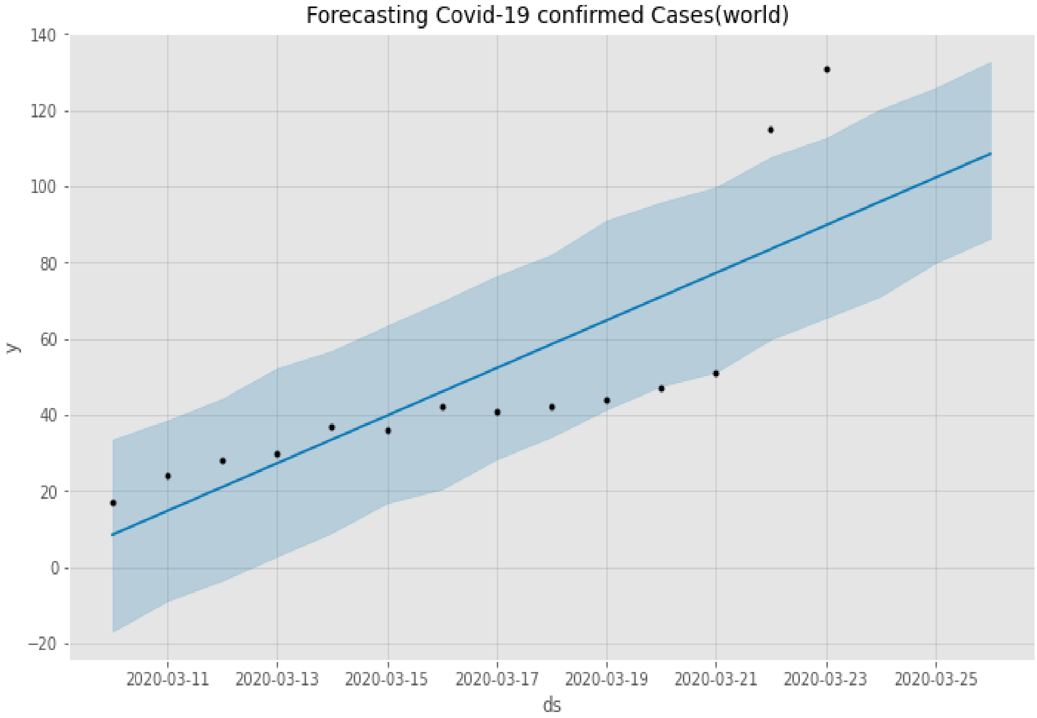


Figure 5. Daily predicted cases of COVID-19 using the PROPHET ML algorithm.

In Figure 6, y-hat or y-hat represent the estimated or predicted values in predictive models. In a regression or other predictive model, the estimated or anticipated values are referred to as y-hat values.

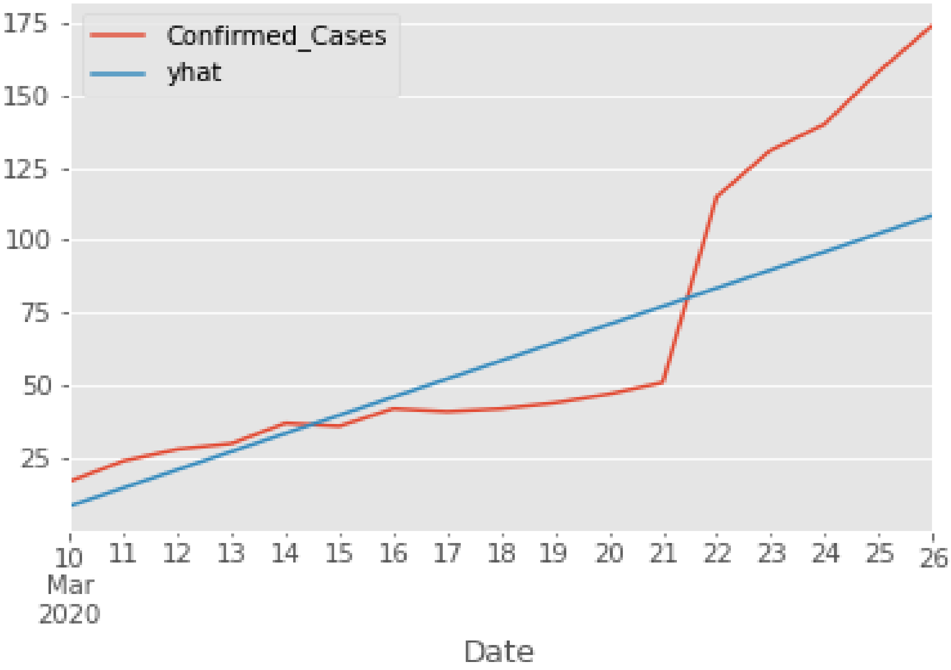
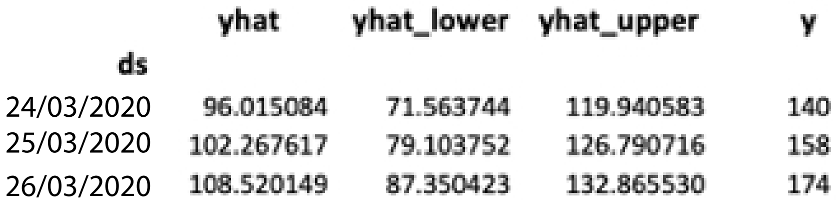


Figure 6. Comparison of actual and predicted cases of COVID-19 using Prophet. The X-axis shows the number of COVID-19 confirmed cases.

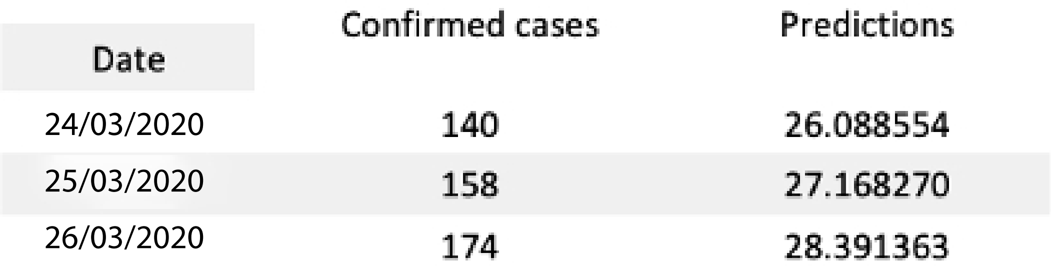
Figure 7 depicts PROPHET's forecasts based on test results. The date is represented by ds, and the confirmed instances for the provided dates are represented by y.



	yhat	yhat_lower	yhat_upper	y
ds				
24/03/2020	96.015084	71.563744	119.940583	140
25/03/2020	102.267617	79.103752	126.790716	158
26/03/2020	108.520149	87.350423	132.865530	174

Figure 7. PROPHET forecasting on test data. The ds represents the date and y represents the confirmed cases for the given dates.

Figure 8 shows the AR predicting outcomes. The AR is shown in the middle of the graph, and residuals at various time steps are displayed beside each observation. Both the proven and predicted instances are evident.



Date	Confirmed cases	Predictions
24/03/2020	140	26.088554
25/03/2020	158	27.168270
26/03/2020	174	28.391363

Figure 8. Comparison of the actual confirmed case and the predicted case of the PROPHET algorithms.

Figure 9 represents the predicted value of the test data, whereas the blue line represents the actual value.

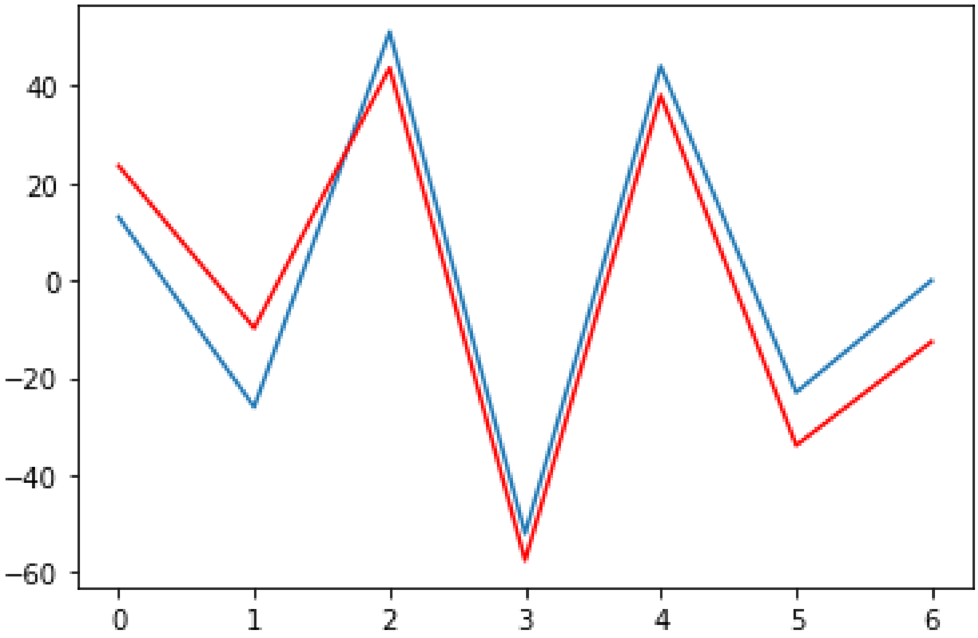


Figure 9. The AR prediction on test data. The red line: predicted value; the blue line: the actual value.

Next, we built the AR and ARIMA models using Python. The Figure 9 shows prediction results of AR model. The AR and ARIMA models (RMSE 10.49 and 34.75, respectively) are shown in Table 4. Additionally, We tuned the AR and ARIMA model parameters using Python because they affect the performance of AR models. Finally, we constructed the time series model LSTM of deep learning using the Python KERAS framework. The performance of LSTM could be better because LSTM requires a considerable amount of data.

Finally, we compared the performance of time series ML models on COVID-19 datasets. The experiment results indicate that developers can integrate the AR and PROPHET time series model into the COVID-19 death warning system and predict confirmed and fatal patient cases in the country (with high performance). Popular algorithms, e.g., the LSTM and ARIMA models, perform well with various real-world issues. Nonetheless, the outcome shows that the performance of these models is inadequate. We found that the size of the confirmed COVID-19 is positively correlated with the level of death caused by COVID-19. Furthermore, from the beginning, the confirmed cases and deaths increased. Regression analysis shows a positive association between the "size of the total population" and the "size of the infected population", along with the number of deaths from COVID-19.

5. Strengths and Limitations

The strength of the study includes, but is not limited to, the performance of PROPHET and AR Time Series algorithms, which is high and can easily integrate with health systems because the PROPHET and AR algorithms are simple and can be easily implemented without programming skills. Moreover, this study also, in the same way, carries certain limitations, which need to be addressed in future research studies. The dataset was limited only to country, state, confirmed cases, recovered cases, deaths, dates, and population as an input related to the COVID-19 disease. At the same time, COVID-19 also depends on other factors, such as age, weather, and even gender, whose inclusion may increase ML models' performance. Similarly, in the current study, the number of patient records was limited and the ML model performance may increase if we increase the number of documents in the dataset.

6. Conclusions

Researchers have encountered various challenges when attempting to construct a warning system that can predict the rapid development and spread of COVID-19. Some issues are hardware resources, DL network architecture repair, and data availability. A massive dataset is required to implement DL methods, such as LSTM, for prediction. The absence of such datasets may result in inaccurate and improper conclusions—consequently, the performance of deep learning architectures declines concerning these warning systems. In addition, there is uncertainty associated with medical datasets. Another problem with the datasets is the lack of phenotypic data, such as gender and age. Moreover, for the prognosis of the disease using computer-assisted early warning systems, several elements (such as infection of neighbor/friend/family member, climatic circumstances, policies to prevent the spread of the disease by countries, and the average age of the community) come into play. The nature of COVID-19 is still largely unclear, so the probability of mutation is a formidable obstacle.

This study examined the performance of time series ML models for predicting patients' confirmed, detected, and death cases over the following week (using a given dataset for research purposes). After training the LSTM, AR, PROPHET, and ARIMA models, we calculated the predictions of confirmed and detected death cases for the next week. The findings predict that PROPHET and AR models have the lowest RMSE error for making predictions concerning the confirmed, detected, and death cases. Furthermore, the present research suggests that we can include PROPHET and AR models in the COVID-19 hospital dashboard. Based on the time series ML technique, it can also predict the medical personnel and government institutions' ability to predict, detect, and confirm COVID-19 death cases in the nation over the next week.

Governments across the globe have adopted various measures to contain the COVID-19 epidemic. Among these measures are the closure of public education and leisure places, such as schools, colleges, universities, movie theaters, retail malls, and parks, and the restriction of face-to-face meetings via obligatory "social distancing". The majority of the global population must adhere to these extraordinary measures. As the number of medical facilities restricted in many developing countries, the exponential development of COVID-19 cases places a tremendous strain on health professionals and services; it causes a shortage of intensive care facilities in hospitals. The early prediction of this pandemic may assist governments, planning officials, and physicians in addressing the health issue more effectively. Thus, a COVID-19 warning system equipped with AI and ML may provide a great source of assistance.

Author Contributions: M.H. (Mushtaq Hussain): conceptualization, methodology, writing, visualization, and supervision; M.A.C. and A.I.: investigation, data curation, conceptualization, methodology, and writing; J.A.T.: formal analysis, writing, review, and editing; S.N.: formal analysis, visualization, writing, review, editing, and hand journal correspondence; M.H. (Monia Hamdi): formal analysis, review, editing, funding; H.H.: validation, analysis, review, and editing; M.I.: review and editing; and T.S.: writing, revising, and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R125), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data Availability Statement: The current study data are publicly available online. No participants' personal information (e.g., name or address) was included in this study. The dataset used in the current study is publicly available at: https://doi.org/10.7910/DVN/URHUOV (accessed on 25 September 2022), https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/URHUOV (accessed on 25 September 2022).

Conflicts of Interest: The authors declare no conflict of interest

References

1. McIntosh, K.; Perlman, S. Coronaviruses, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). In *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*; Bennett, J.E., Dolin, R., Blaser, M.J., Eds.; Elsevier Health Sciences: Amsterdam, The Nederlands, 2015; pp. 1928–1936.e2.
2. Morens, D.M.; Daszak, P.; Taubenberger, J.K. Escaping Pandora's box—Another novel coronavirus. *N. Engl. J. Med.* **2020**, *382*, 1293–1295.
3. Tu, W.-J.; Cao, J.; Yu, L.; Hu, X.; Liu, Q. Clinicolaboratory study of 25 fatal cases of COVID-19 in Wuhan. *Intensiv. Care Med.* **2020**, *46*, 1117–1120. https://doi.org/10.1007/s00134-020-06023-4.
4. Bennett, J.E.; Dolin, R.; Blaser, M.J. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 9th ed.; Elsevier Health Sciences: Amsterdam, The Netherlands, 2014.
5. Gralinski, L.E.; Menachery, V.D. Return of the Coronavirus: 2019-nCoV. *Viruses* **2020**, *12*, 135. https://doi.org/10.3390/v12020135.
6. Sahoo, B.K.; Sapra, B.K. A data driven epidemic model to analyse the lockdown effect and predict the course of COVID-19 progress in India. *Chaos Solitons Fractals* **2020**, *139*, 110034.
7. Shishvan, O.R.; Zois, D.; Soyata, T. Machine intelligence in healthcare and medical cyber physical systems: A survey. *IEEE Access* **2018**, *6*, 46419–46494.
8. Chen, C. Ascent of machine learning in medicine. *Nat. Mater.* **2019**, *18*, 407.
9. Swapnarekha, H.; Behera, H.S.; Nayak, J.; Naik, B. Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. *Chaos Solitons Fractals* **2020**, *138*, 109947. https://doi.org/10.1016/j.chaos.2020.109947.
10. Deng, W.; Ni, H.; Liu, Y.; Chen, H.; Zhao, H. An adaptive differential evolution algorithm based on belief space and generalized opposition-based learning for resource allocation. *Appl. Soft Comput.* **2022**, *127*, 109419. https://doi.org/10.1016/j.asoc.2022.109419.
11. Yao, R.; Guo, C.; Deng, W.; Zhao, H. A novel mathematical morphology spectrum entropy based on scale-adaptive techniques. *ISA Trans.* **2022**, *126*, 691–702. https://doi.org/10.1016/j.isatra.2021.07.017.
12. Lalmuanawma, S.; Hussain, J.; Chhakchhuak, L. Applications of machine learning and artificial intelligence for COVID-19 (SARS-CoV-2) pandemic: A review. *Chaos Solitons Fractals* **2020**, *139*, 110059.
13. Wu, D.; Wu, C. Research on the Time-Dependent Split Delivery Green Vehicle Routing Problem for Fresh Agricultural Products with Multiple Time Windows. *Agriculture* **2022**, *12*, 793. https://doi.org/10.3390/agriculture12060793.
14. Sarkar, K.; Khajanchi, S.; Nieto, J.J. Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fractals* **2020**, *139*, 110049.
15. Ribeiro, M.H.D.M.; da Silva, R.G.; Mariani, V.C.; dos Santos Coelho, L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos Solitons Fractals* **2020**, *135*, 109853.
16. Xue, Y.; Onzo, B.M.; Mansour, R.F.; Su, S.B. Deep Convolutional Neural Network Approach for COVID-19 Detection. *Comput. Syst. Sci. Eng.* **2022**, *42*, 201–211.
17. Mansour, R.F.; Escorcia-Gutierrez, J.; Gamarra, M.; Gupta, D.; Castillo, O.; Kumar, S. Unsupervised Deep Learning based Variational Autoencoder Model for COVID-19 Diagnosis and Classification. *Pattern Recognit. Lett.* **2021**, *151*, 267–274. https://doi.org/10.1016/j.patrec.2021.08.018.
18. Yan, Z.; Wang, Y.; Yang, M.; Li, Z.; Gong, X.; Wu, D.; Zhang, W.; Wang, Y. Predictive and analysis of COVID-19 cases cumulative total: ARIMA model based on machine learning. *medRxiv* 2022. https://doi.org/10.1101/2022.01.24.2226979.
19. Ganiny, S.; Nisar, O. Mathematical modeling and a month ahead forecast of the coronavirus disease 2019 (COVID-19) pandemic: An Indian scenario. *Model. Earth Syst. Environ.* **2021**, *7*, 29–40. https://doi.org/10.1007/s40808-020-01080-6.
20. Wadhwa, P.; Aishwarya; Tripathi, A.; Singh, P.; Diwakar, M.; Kumar, N. Predicting the time period of extension of lockdown due to increase in rate of COVID-19 cases in India using machine learning. *Mater. Today Proc.* **2020**, *37*, 2617–2622. https://doi.org/10.1016/j.matpr.2020.08.509.
21. Dil, S.; Dil, N.; Maken, Z.H. COVID-19 trends and forecast in the Eastern Mediterranean Region with a Particular Focus on Pakistan. *Cureus* **2020**, *12*, e8582.
22. Yadav, M.; Perumal, M.; Srinivas, M. Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos Solitons Fractals* **2020**, *139*, 110050–110050. https://doi.org/10.1016/j.chaos.2020.110050.
23. Velásquez, R.M.A.; Lara, J.V.M. Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression. *Chaos Solitons Fractals* **2020**, *136*, 109924. https://doi.org/10.1016/j.chaos.2020.109924.
24. Hamzah, F.B.; Lau, C.; Nazri, H.; Ligot, D.V.; Lee, G.; Tan, C.L.; Bin Mohd Shaib, M.K.; Binti Zaidon, U.H.; Binti Abdullah, A.; Chung, M.H.; et al. CoronaTracker: Worldwide COVID-19 outbreak data analysis and prediction. *Bull. World Health Organ.* **2020**, *1*, 1–32.
25. Mahajan, A.; A Sivadas, N.; Solanki, R. An epidemic model SIPHERD and its application for prediction of the spread of COVID-19 infection in India. *Chaos Solitons Fractals* **2020**, *140*, 110156–110156. https://doi.org/10.1016/j.chaos.2020.110156.

26. Yang, Z.; Zeng, Z.; Wang, K.; Wong, S.-S.; Liang, W.; Zanin, M.; Liu, P.; Cao, X.; Gao, Z.; Mai, Z.; et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* **2020**, *12*, 165–174.
27. Shahid, F.; Zameer, A.; Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* **2020**, *140*, 110212. https://doi.org/10.1016/j.chaos.2020.110212.
28. Cheikhrouhou, O.; Mahmud, R.; Zouari, R.; Ibrahim, M.; Zaguia, A.; Gia, T.N. One-Dimensional CNN Approach for ECG Arrhythmia Analysis in Fog-Cloud Environments. *IEEE Access* **2021**, *9*, 103513–103523. https://doi.org/10.1109/access.2021.3097751.
29. Kırbaş, İ.; Sözen, A.; Tuncer, A.D.; Kazancıoğlu, F.Ş. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos Solitons Fractals* **2020**, *138*, 110015.
30. Zeroual, A.; Harrou, F.; Dairi, A.; Sun, Y. Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos Solitons Fractals* **2020**, *140*, 110121.
31. Jockers, M.L.; Thalken, R. Introduction to dplyr. In *Text Analysis with R*; Springer: Cham, Switzerland, 2020; pp. 121–132.
32. Liu, S.; Sweeney, C.; Srisarajivakul-Klein, N.; Klinger, A.; Dimitrova, I.; Schaye, V. Evolving oxygenation management reasoning in COVID-19. *Diagnosis* **2020**, *7*, 381–383. https://doi.org/10.1515/dx-2020-0099.
33. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF models for sequence tagging. *arXiv* **2015**, arXiv:150801991. https://doi.org/10.48550/arXiv.1508.01991.
34. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.
35. Nguyen, D.H.D.; Tran, L.P.; Nguyen, V. Predicting Stock Prices Using Dynamic LSTM Models. In *Applied Informatics*; Florez, H., Leon, M., Diaz-Nafria, J., Belli, S., Eds.; Springer: Cham, Switzerland, 2019.
36. Center for Systems Science and Engineering (CSSE). *Coronavirus COVID-19 Global Cases*; Johns Hopkins University (JHU): Baltimore, MD, USA, 2020.
37. Shoeibi, A.; Khodatars, M.; Alizadehsani, R.; Ghassemi, N.; Jafari, M.; Moridian, P.; Khadem, A.; Sadeghi, D.; Hussain, S.; Zare, A.; et al. Automated detection and forecasting of COVID-19 using deep learning techniques: A review. *arXiv* **2020**, arXiv:200710785. https://doi.org/10.48550/arXiv.2007.10785.
38. Cifci, M.A. SegChaNet: A Novel Model for Lung Cancer Segmentation in CT scans. *Appl. Bionics Biomech.* **2022**, *2022*, 1139587.
39. Alizadehsani, R.; Roshanzamir, M.; Hussain, S.; Khosravi, A.; Koohestani, A.; Zangooei, M.H.; Abdar, M.; Beykikhoshk, A.; Shoeibi, A.; Zare, A.; et al. Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991–2020). *Ann. Oper. Res.* **2021**, 1–42. https://doi.org/10.1007/s10479-021-04006-2.
40. Alizadehsani, R.; Sani, Z.A.; Behjati, M.; Roshanzamir, Z.; Hussain, S.; Abedini, N.; Hasanzadeh, F.; Khosravi, A.; Shoeibi, A.; Roshanzamir, M.; et al. Risk Factors Prediction, Clinical Outcomes and Mortality of COVID-19 Patients. *J. Med. Virol.* **2020**, *93*, 2307–2320.
41. Cifci, M.A. Derin Öğrenme Metodu ve Ayrık Dalgacık Dönüşümü Kullanarak BT Görüntülerinden Akciğer Kanseri Teşhisi. *Mühendislik Bilimleri Ve Araştırmaları Derg.* **2022**, *4*, 141–154.