Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM

Farah Shahid, Aneela Zameer*, Muhammad Muneeb

Department of Computer & Information Sciences, Pakistan Institute of Engineering & Applied Sciences (PIEAS), Nilore, Islamabad 45650, Pakistan

ARTICLE INFO

ABSTRACT

COVID-19, responsible of infecting billions of people and economy across the globe, requires detailed study of the trend it follows to develop adequate short-term prediction models for forecasting the number of future cases. In this perspective, it is possible to develop strategic planning in the public health system to avoid deaths as well as managing patients. In this paper, proposed forecast models comprising autoregressive integrated moving average (ARIMA), support vector regression (SVR), long shot term memory (LSTM), bidirectional long short term memory (Bi-LSTM) are assessed for time series prediction of confirmed cases, deaths and recoveries in ten major countries affected due to COVID-19. The performance of models is measured by mean absolute error, root mean square error and r2_score indices. In the majority of cases, Bi-LSTM model outperforms in terms of endorsed indices. Models ranking from good performance to the lowest in entire scenarios is Bi-LSTM, LSTM, GRU, SVR and ARIMA. Bi-LSTM generates lowest MAE and RMSE values of 0.0070 and 0.0077, respectively, for deaths in China. The best r2_score value is 0.9997 for recovered cases in China. On the basis of demonstrated robustness and enhanced prediction accuracy, Bi-LSTM can be exploited for pandemic prediction for better planning and management.

1. Introduction

Corona virus 2019 (COVID-19) epidemic has spread from Wuhan, China to 213 countries across the globe. According to the WHO (World Health Organization) on February 17, 2020, that 80% of coronavirus patients have mild fever and recover, while 2% death rate is reported as compared to other corona diseases, named as SARS (2003) and MERS (2012-2019), that had death rate enclosing of 774 deaths from 8089 confirmed cases as 10% and 858 deaths from 2494 confirmed cases as 34%, respectively [1]. On July 09, 2020, WHO proclaimed COVID-19 outbreak a pandemic including globally infected 559,694 deaths and 10,509,505 confirmed cases. Region-wise this distribution depicts total deaths in Africa (7,559), Americas (272,606), Eastern Mediterranean (29,127), Eu-

rope (201,853), Asia (26,808) and Western Pacific as (7,515); while confirmed cases are (410,744), (6,125,802), (1,222,070), (2,847,887), (1,032,167), and (234,815) [1].

To be precise, COVID-19 has followed specific patterns and these patterns are based on dynamic transmission of the epidemic. When it occurs, superseding measures of different methods are used to find and evaluate such infective diseases. Any epidemic in a state or country has arisen with different aspect of magnitude with respect to time, particularly weather period changes and spread of virus over the time period, and exhibited as nonlinear in nature. To capture these non-linear compelling changes, researchers have gained the attention and designed such nonlinear systems to describe the abruptness of infective diseases [2]. Therefore, mathematical models such as SIR (susceptible-infective-removed) for analyzing the epidemics has been introduced [3]. A transmission model with incubation time for malaria [4] and a deterministic model to analyze the interaction between HIV and tuberculosis is successfully developed to solve the nonlinear behavior of parameters [5]. Similar models of discrete time equations are used to control the infected population [6].

Amid of physical and statistical methods, the difference is to learn the temporal behavior of data such as coronavirus and use of non-linear functions to predict the dynamics [7,8]. Usually statistical approaches are based on autoregressive integrated moving

---

average (ARIMA) model that is employed to predict the spread of epidemic trend COVID-2019 [9] and seasonal autoregressive integrated moving average (SARIMA) model which estimates the fatality rate by use of time series analysis on influenza epidemic [10]. These models have also been used to monitor and predict the dengue hemorrhagic fever (DHF) cases in southern Thailand [11] and hemorrhagic fever with renal syndrome (HFRS) cases in China to control diseases more effectively [12]. Another popular statistical model in the field of health care system is known as artificial intelligence (AI) based which is used to learn and train the COVID-19 dataset of Hubei Province in China to predict the epidemic peaks and trend size [13]. In numerous cases, these methods are not capable to fit actual data utterly and predicted accuracy is very low, while predicting the rise of COVID-19 spread.

In order to get better performance of statistical methods, machine learning (ML) models which cover several fields such as power and energy engineering [14], technology [15], psychology [16], is used for early prediction and real-time spread of data. Recently, one of ML approach namely, infection size aware random forest (iSARF), observed by classification group has been proposed, which highlights the infection size and lung fields [17]. Other models are multilayer perceptron (MLP) and adaptive network-based fuzzy inference system, (ANFIS) utilized for evaluating the complex variation behavior and predicting the COVID-19 transmission [18]. Hybrid approach of support vector regression (SVR) and ARIMA has been suggested to take the confirmed cases and give predictions related to the number of contaminated persons countrywide [19]. Furthermore, Parbat et al. employed SVR model with radial basis function (RBF) kernel method to forecast daily cases, recovered cases and death cases [20]. Hao has constructed the ensemble predictor of SVR and random forest (RF) to predict seven day ahead number of hospitalized patients [21].

Deep learning (DL) algorithms show a vital role in the analysis and prediction of huge outbreak data patterns and help in early exploitation to stop the spread rate of coronavirus [22]. COVID-19 is a time series data and vastly endorsed the use of sequential models to deal with its dynamic nature. Bandyopadhyay et al. has proposed the gated recurrent neural network and long short term memory (LSTM) to evaluate the predictions with confirmed, negative released, and death cases of COVID-19 [23]. Huang et al. have employed DL based convolutional neural network (CNN) model to estimate COVID-19 cumulative confirmed cases [24].

The novelty of the reported work lies in creating the three categories of confirmed cases, death cases and recovered cases from dataset and intelligently developing a COVID-19 predictor to predict and analyze future trends of these three categories. This experiment is based on the data set of confirmed COVID-19 cases available until June 27, 2020. Additionally, owing to the dynamic nature of coronavirus, ML and DL models have been implemented for early predictions. The prominent features of the methodology are summarized in terms of highlights as follows:

- Statistical models as ARIMA, ML technique of SVR with polynomial and RBF kernels, and DL mechanisms of LSTM, GRU and Bi-LSTM are proposed to predict the COVID-19 three categories, confirmed cases, deaths and recovered cases for ten countries.
- Accuracy of models is measured in terms of three performance measures, MAE, RMSE and r2_score.
- Bi-LSTM time series model enhances the learning ability and memorizing the long sequence. Dl techniques in general and Bi-LSTM in specific are proposed for smallest prediction error and higher accuracy.

Rest of the article is organized as follows: Section II describes the proposed methodologies, dataset and performance metrics; Section III includes detailed results of the designed scheme. While the conclusion are provided in the last section.

2. Design methodology for COVID-19 prediction

In this work, two kinds of methodologies, statistical model and machine learning models including simple and deep learning techniques are established for COVID-19 predictions. In the first phase, design of ARIMA and SVR as simple ML algorithm are discussed, whereas in the next phase, description of various DL models are presented. The statistical performance in terms of three error measures, MAE, RMSE and r2_score are also specified in this section for performance evaluation. The graphical overview of the proposed scheme is illustrated in Fig. 1, in which three categories (confirmed case, deaths cases and recovered cases) of data is collected and after preprocessing, data is passed to respective models separately and performance of models are measured through error measures. Furthermore, detail description of proposed models is provided below.

2.1. Autoregressive integerated moving average

ARIMA model comprises three processes named as auto regression, integration and moving average which is data independent and employed for model architecture and parameter estimation that is linear function for past observations and arbitrary error [25,26]. Time series form of underlying process is:

$$y^t = \theta_0 + \phi_1 y^{t-1} + \phi_2 y^{t-2} + ... + \phi_p y^{t-p}$$
$$+ \varepsilon^t - \theta_1 \varepsilon^{t-1} - \theta_2 \varepsilon^{t-2} - ... - \theta_q \varepsilon^{t-q} \qquad (1)$$

In Eq. (1), y^t and ε^t represent the original value and arbitrary error at time step t. $\phi_a (a = 1, 2, ..., p)$ and $\phi_b (b = 0, 1, 2, ..., q)$ are parameters of the model. Arbitrary error symbolizes by ε^t is considered with zero mean and σ^2 of standard variance. Eq. (1) presents ARIMA model mathematically and is used to solve several problems in various applications. Taking the value $q = 0$ in Eq. (1) works as AR model with order p and for $p = 0$ it becomes the MA model with q order. Hence, (p, q) are both important factors to determine ARIMA model.

2.2. Support vector regression

Another effective time series implementation of support machine (SVM) anticipated by Vapnik is known as support vector regression [9]. Both the SVM and SVR are used to minimize the error of margin and employ kernel functions for non-separable classes. The results can be improved by optimizing its parameters; in this regard grid and heuristic search are used to get best parameters [27]. SVR for the multidimensional data is mathematically formulated as:

$$y = f(X) = \sum_{i=1}^{M} W_i X_i + b \qquad (2)$$

In Eq. 2, X_i represent input feature values, W_i are input weights, b is bias and y is used for actual values, whereas M is the total number of data samples. Following equation shows the objective function of SVR and $\|W\|$ is employed for magnitude of the vector.

$$\min_{W} \frac{1}{2} \|W\|^2 \qquad (3)$$

Implementing SVM with soft margin approach comprises two slack variables known as ξ and ξ^* that is used to protect against outliers and $\frac{1}{2}\|w\|^2$ is employed for function smoothness. Both of these parameters depend upon the C parameter. Then, Eq. (3) is formulated into Eq. (4) as:

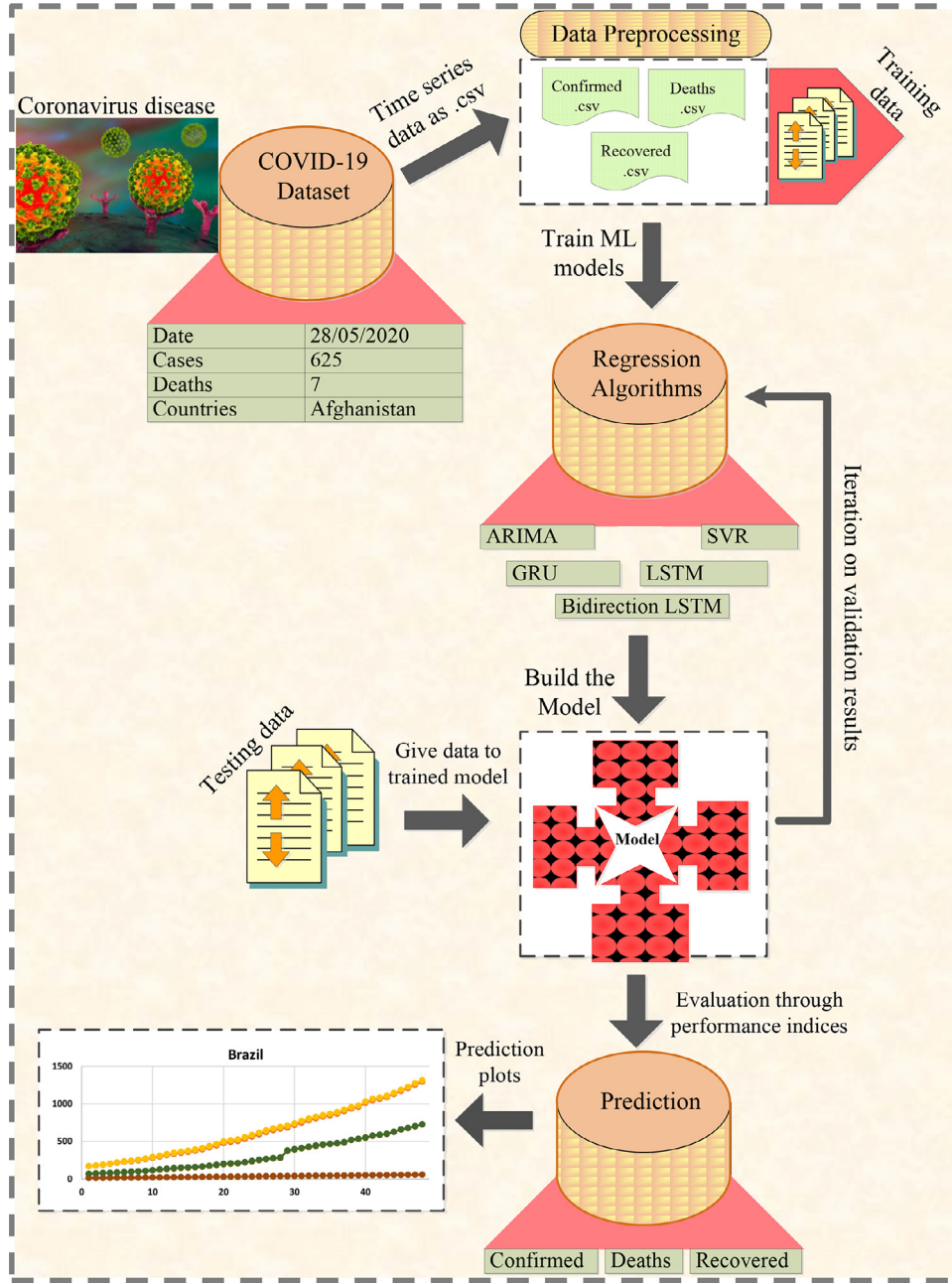$$\min_{W} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^{M} \left(\xi_i + \xi_i^* \right) \qquad (4)$$

Fig. 1. Graphical abstract of the proposed scheme.

With the constraints,

$$\begin{cases} y_i - W^T X_i \le \varepsilon + \xi_i^*, i = 1, 2, ..., M \\ W^T X_i - y_i \le \varepsilon + \xi i, i = 1, 2, ..., M \\ \xi i, \xi_i^* \ge 0 \end{cases} \qquad (5)$$

By solving the Eq. (5) with constraints and getting Lagrangian multipliers that are nonnegative real numbers such as $\alpha i - \alpha_i^*$. This is useful to deal with nonlinear functions in which data is mapped into high dimension space known as kernel space for high accuracy results. Finally, SVR function is mathematically obtained as Eq. (6):

$$f(X) = \sum_{i=1}^{M} \left(\alpha_i^* - \alpha i \right) k(X_i, X) + b \qquad (6)$$

$$k(X_i, X) = \varphi(X_i), \varphi(X) \qquad (7)$$

Here, primal formula of kernel function is $k(X_i, X)$ and $\varphi(X)$ in Eq. 7 represents the features in kernel space. Various kernel functions such as RBF and polynomial kernels are used, and their mathematical formulae is given as:

$$\sigma : k(X_i, X) = \exp(-\|X_i - X\|^2 / 2\sigma^2) \qquad (8)$$

$$k(X, X') = \left((X, X') + 1 \right)^d \qquad (9)$$

In Eqs. (8,9) σ and d is the parameter of kernel that is tuned.

2.3. LSTM and Bi-LSTM

RNN [28] has been employed for sequential time series applications with temporal dependencies. An unfolded RNN has the capability to process current data by use of previous data. Meanwhile, RNN has the problem to train the long term dependencies

data, which is solved by one of the variants of RNN. LSTM anticipated by Hochreiter and Schmidhuber [29], has been used as advance version of RNN network and has overcome the limitation of RNN by use of hidden layer unit known as memory cells. Memory cells have the self-connections that stored the network temporal state and controlled through three gates named as: input gate, output gate and forget gate [30]. The work of input gate and output gate is used to control the flow of memory cell input and outputs into the rest of network. In addition, forget gate has been added to the memory cell, which pass the output information with high weights from previous to next neuron. The information reside in memory depend upon the high activation results; if the input unit has high activation, the information is stored in memory cell. In addition, if the output unit has high activation then it will pass the information to next neuron. Otherwise, input information with high weights resides in memory cell.

LSTM network is compute mapping between input sequence and output sequence, *i.e.* $X = (X_1, X_2, ..., X_n)$ and $y = (y_1, y_2, ..., y_n)$. Calculating by the following equations:

$$f_{orget}g_{ate} = sigmoid\left(W_{fg}X_t + W_{hfg}h_{t-1} + b_{fg}\right) \quad (10)$$

$$i_{nput}g_{ate} = sigmoid\left(W_{ig}X_t + W_{hig}h_{t-1} + b_{ig}\right) \quad (11)$$

$$o_{utput}g_{ate} = sigmoid\left(W_{og}X_t + W_{hog}h_{t-1} + b_{og}\right) \quad (12)$$

$$(C)_t = (C)_{t-1} \otimes (f_{orget}g_{ate})_t + (i_{nput}g_{ate})_t \\ \otimes (tanh (W_C X_t + W_{hC}h_{t-1} + b_C)) \quad (13)$$

$$h_t = o_{utput}g_{ate} \otimes \tanh\left((C)_{t-1}\right) \quad (14)$$

In Eqs. 10, 11, 12, 13 and 14, W_{ig}, W_{og}, W_{hC}, W_{fg} and b_{fg}, b_{ig}, b_{og}, b_C represent the weights and bias variables respectively of three gates and a memory cell. Here, h_{t-1} symbolizes the prior hidden layers units that element-wise adding with weights of three gates. After the processing of Eq. 13, $(C)_t$ turns into current memory cell unit. Eq. 14 shows the element wise multiplication of prior hidden unit outputs and previous memory cell unit. Add the non-linearity on top of the three gates in the form of tanh and sigmoid activation functions, which is shown in Eqs. (10–14). Here, $t - 1$ and t are previous and current time steps.

To overcome the limitations of LSTM cell which is able to work on previous content but cannot use the future one. Schuster and Paliwal [31] proposed bidirectional recurrent neural networks (BRNN) that is comprised of two distinct LSTM hidden layers with similar output in opposite directions. With this architecture, previous and future information is exploited in output layer. An input sequence $X = (X_1, X_2, ..., X_n)$ in Bi-LSTM is calculated in forward direction as $\overrightarrow{h_t} = (\overrightarrow{h_1}, \overrightarrow{h_2}, ..., \overrightarrow{h_n})$ and backward directions as $\overleftarrow{h_t} = (\overleftarrow{h_1}, \overleftarrow{h_2}, ..., \overleftarrow{h_n})$. The final out of this cell y_t is formed by both $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$, the final sequence of out looks like $y = (y_1, y_2, ...y_t..., y_n)$. Fig. 2 displays the single cell of LSTM and Bi-LSTM.

2.4. Gated recurrent unit (GRU)

GRU is the simple variant of LSTM that has two gates, one is "update gate" which comprises of input, forget gates and "reset gate" [32,33]. GRU has no additional memory cell to keep information, therefore, it can only control information inside the unit.

$$u_{pdate}g_{ate} = sigmoid(W_{ug}X_t + W_{ug}h_{t-1}) \quad (15)$$

$$r_{est}g_{ate} = sigmoid(W_{rg}X_t + W_{rg}h_{t-1}) \quad (16)$$

$$\tilde{h}_t = \tanh (W (r_{est}g_{ate})_t \otimes +W h_{t-1}, X_t) \quad (17)$$

$$h_t = \left(1 - \left(u_{pdate}g_{ate}\right)_t\right) \otimes h_{t-1} + \left(u_{pdate}g_{ate}\right)_t \otimes \tilde{h}_t \quad (18)$$

Here, $u_{pdate}g_{ate}$ in Eq. 15 decides for how much content or information is updated. In Eq. 16, $r_{est}g_{ate}$ is similar to update gate, if the gate is set to zero, it reads input sequences and forget the previously calculated state. Further, $\tilde{h}_t$ shows the same functionality as in recurrent unit and h_t of GRU at time t represents the linear interpolation among the current $\tilde{h}_t$ and previous h_{t-1} activation states Eqs. 17 and 18.

2.5. COVID-19 dataset

Dataset of novel coronavirus is taken from the link [34]. The .csv file of confirmed cases, death cases and recovered cases of all countries is provided column wise. An individual file is created of these three categories from 22 January, 2020 to 27 June, 2020. Covid19 dataset contains number of confirmed cases, deaths and recovered cases of 158 samples and we have taken cases from 1/22/2020 to 5/10/2020 for training purpose and to predict cases from 5/11/2020 to 6/27/2020. For each country, data comprises given cases for 110 days and have to predict for next 48 days. The data is preprocessed before it is given to ML models for training.

2.6. Performance indices

Three performance measures are used to evaluate the performance of the proposed model, these are mean absolute error (MAE), root mean square error (RMSE) and r$_2$_score. C denotes the actual value and $\hat{C}$ for estimated value. The expected values of MAE is zero for the best model and is expressed mathematically as in Eq. 19.

$$MAE = \frac{1}{M} \sum_{i=1}^{M} \left|C - \hat{C}\right| \quad (19)$$

RMSE is well-defined in Eq. 20 as:

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M} (C - \hat{C})^2} \quad (20)$$

To demonstrates the variance between dependent and independent parameter, r$_2$_score is presented in Eq. 21 as:

$$\text{r}_2\text{\_score} = 1 - \frac{\sum |C - C'|}{\sum |C - \hat{C}|} \quad (21)$$

3. Experimental results and discussion

This paper aims at comparing prediction models from statistics, machine learning and deep learning on COVID-19 dataset for ten countries including Brazil, Germany, Italy, Spain, UK, China, India, Israel, Russia, and USA all over the globe. Statistical ARIMA and ML based SVR with polynomial and RBF kernels are implemented as base line regressors, while LSTM, GRU, and Bi-LSTM are taken as deep learning models for sequential predictions. COVID-19 dataset is divided into training samples of 110 days and 48 test days. Performance of all models have been compared on the basis of standard statistical performance measures in terms of MAE, RMSE, and r2_score. The simulations and experiments have been carried out on NVIDIA GTX1070 and encoded by python Keras.

The dataset comprises three features of confirmed cases, deaths and recovered cases. Unscaled data slows down the convergence process. MinMaxScaler subtracts the smallest value of feature and formerly divides by features range. The range is the difference between the original maximum and original minimum. MinMaxScaler reserves the shape of the original distribution of data. It does
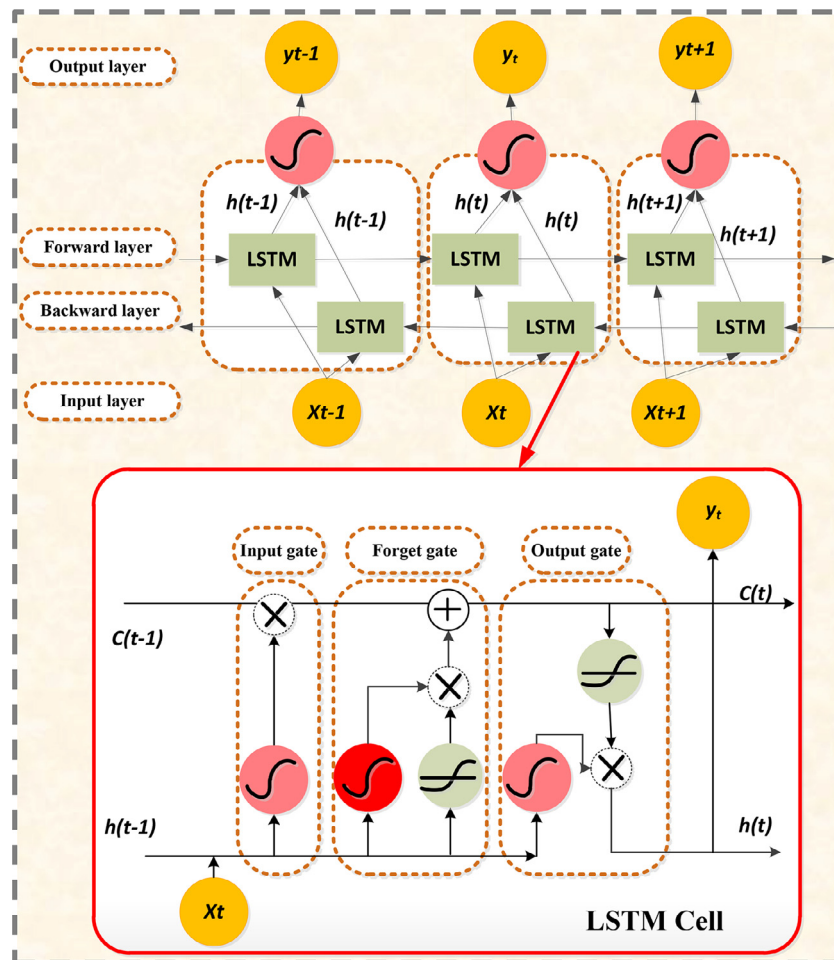
Fig. 2. Architecture of a single LSTM cell and Bi-LSTM.

Table 1
Proposed scheme with parameters and their values.

Method	Parameters	Values
SVR	C	3.0
	epsilon	0.0000001
	degree	3
	tolerance	0.000001
LSTM/Bi-LSTM/GRU	Layers	3
	No. of neurons	{16,32,64,128}
	Learning rate	0.001
	Optimizer	Adam
	Batch size	10
	Epochs	300
	Time step	3
ARIMA	(p, d, q)	(1,1,1)

not meaningfully change the information embedded in the original data and does not reduce the importance of outliers. Parameters with their values of SVR, ARIMA and LSTM is shown in Table 1, while results of actual and predicted cases in three categories in terms of performance measures are presented in Table 2. It can be observed from this table that none of the three models, ARIMA, SVR_Poly, and SVR_RBF fits the dataset very well and therefore does not generate consistent predictions. Observing the values of RMSE and MAE, for some countries and even for some feature, one predicts better and for others, another model gives better results. In terms of r2_score, mostly the values are negative and thereby depicting poorer performance of the models than linear regressors.

Therefore, it can be inferred that none of these models is able to give reliable and accurate predictions.

As a next step, deep learning techniques of LSTM, GRU and Bi-LSTM for three predicted categories are demonstrated in Fig. 3 in terms of MAE, RMSE and r2_score. It is worth mentioning here that parameter optimization of all methods has been carried out through trial and error and values enlisted in Table 1 have been used in generating all the results in this section. Prediction errors in terms of performance measures are plotted as bar charts for comparison among DL techniques. The smallest value of MAE is 20.79663 for Israel among ten countries for confirmed case. As the number of cases is much more for USA and Brazil, therefore the error measures are also higher for these countries as opposed to rest of other countries in actual figures.

Performance measure, r2_score, independently represent values very close to unity without any normalization and inverse transformation, which is a good sign of a consistent, efficient and accurate model for all countries and all cases. Normalized values of MAE and RMSE closer to zero along with r2_score closer to unity are the main criteria to prefer one model on another with lowest prediction error for one country to others. It is noteworthy here that DL models generate normalized error measures which are then transformed corresponding to actual numbers through inverse scalar transformation for more understanding of these cases in real world figures.

Keeping all three performance measures in view, it can be safely concluded on the basis of results that after parameter tuning, Bi-LSTM performs as best model giving highest accuracy. Pre-

Table 2
Comparison among statistical (ARIMA) and machine learning techniques (SVR_Poly, SVR_RBF) in terms of different error measures.

Countries	Models	Confirmed cases			Death cases			Recovered cases		
		MAE	RMSE	r2_score	MAE	RMSE	r2_score	MAE	RMSE	r2_score
Brazil	ARIMA	117494.33	152938.01	0.7949598	2142.3387	2458.0934	0.9670720	152352.88	207855.99	-0.038123
	SVR_Poly	331148.71	369559.04	-0.197225	16152.701	17650.760	-0.697830	145562.14	163739.68	0.3557835
	SVR_RBF	332941.33	379431.24	-0.262043	16404.688	18369.679	-0.838952	154417.29	181943.98	0.2045748
China	ARIMA	148.16735	180.63025	0.3197708	699.88646	846.27953	-849252.5	2663.9071	3334.3580	-1094.912
	SVR_Poly	468.90544	509.41525	-4.410260	1.4928239	1.5743235	-1.938987	51.518966	66.983256	0.5577334
	SVR_RBF	342.56828	392.10624	-2.205396	1.2710851	1.5344545	-1.792014	67.658164	76.626348	0.4212276
Germany	ARIMA	4608.3634	6255.8472	-0.128367	2342.6339	2693.3497	-51.85996	34938.078	39300.669	-19.12555
	SVR_Poly	5287.3782	5963.7363	-0.025451	529.42987	587.24466	-1.512924	11933.681	13226.521	-1.279498
	SVR_RBF	4554.2812	5540.6970	0.1148699	547.79387	620.47596	-1.805377	12276.062	13909.844	-1.521114
India	ARIMA	46506.574	66122.066	0.7540973	1355.1213	2181.2422	0.7373567	47949.109	66594.549	0.3438362
	SVR_Poly	114015.46	126633.59	0.0980794	3375.2273	3876.4407	0.1704841	74275.961	87042.154	-0.120970
	SVR_RBF	117452.96	134176.23	-0.012561	3504.8377	4072.5075	0.0844498	75041.682	87233.672	-0.125908
Israel	ARIMA	2430.2519	3628.0384	-2.544775	97.148087	111.43638	-55.36314	3479.3614	4203.4680	-11.52140
	SVR_Poly	558.90494	854.02038	0.8035819	17.261576	19.241420	-0.680410	1836.4625	2007.8061	-1.856809
	SVR_RBF	721.39707	944.20321	0.7599090	18.194071	20.712391	-0.947160	1912.0877	2196.0009	-2.417454
Italy	ARIMA	3353.8050	3612.8165	0.5674722	3215.3566	4081.6837	-11.95468	58359.342	73057.181	-8.427845
	SVR_Poly	6929.9818	7719.4938	-0.974693	1499.1804	1665.6623	-1.157357	32759.097	35555.505	-1.233059
	SVR_RBF	5790.5699	6922.6244	-0.588048	1373.0071	1578.5235	-0.937538	33467.462	37324.731	-1.460820
Russia	ARIMA	170638.90	216821.50	-2.269195	256.87862	299.66126	0.9804863	30966.059	35090.626	0.8923068
	SVR_Poly	163600.09	176450.88	-1.165129	2481.4820	2751.2542	-0.644894	141203.95	162676.94	-1.314503
	SVR_RBF	166420.94	183664.90	-1.345786	2515.7493	2854.5801	-0.770765	140895.07	162068.65	-1.297226
Spain	ARIMA	50841.745	63344.706	-110.3090	1061.7753	1185.2928	-3.308268	67778.550	78868.281	-669.4231
	SVR_Poly	5885.6532	6624.5861	-0.217383	515.62924	710.14321	-0.546476	8745.0743	9722.4457	-9.188148
	SVR_RBF	5092.2682	6074.6280	-0.023644	547.19528	741.46513	-0.685904	8929.7137	9976.7892	-9.728174
UK	ARIMA	83359.040	98881.484	-14.53098	7833.5872	9014.3431	-6.514077	398.43903	453.41127	-19.46613
	SVR_Poly	32152.976	35442.099	-0.995299	4169.6098	4612.5804	-0.967412	108.79110	121.57950	-0.471539
	SVR_RBF	33336.295	37554.320	-1.240211	4357.9775	4931.2685	-1.248665	113.03956	129.15120	-0.660535
USA	ARIMA	34867.611	61859.840	0.9622082	14838.374	17940.050	-1.111554	42347.136	58839.033	0.8180914
	SVR_Poly	244528.11	273851.39	0.2593555	14873.134	16377.824	-0.759816	158092.48	172130.93	-0.556824
	SVR_RBF	257046.10	298513.60	0.1199484	15795.817	17883.281	-1.098212	164032.79	187153.62	-0.840426
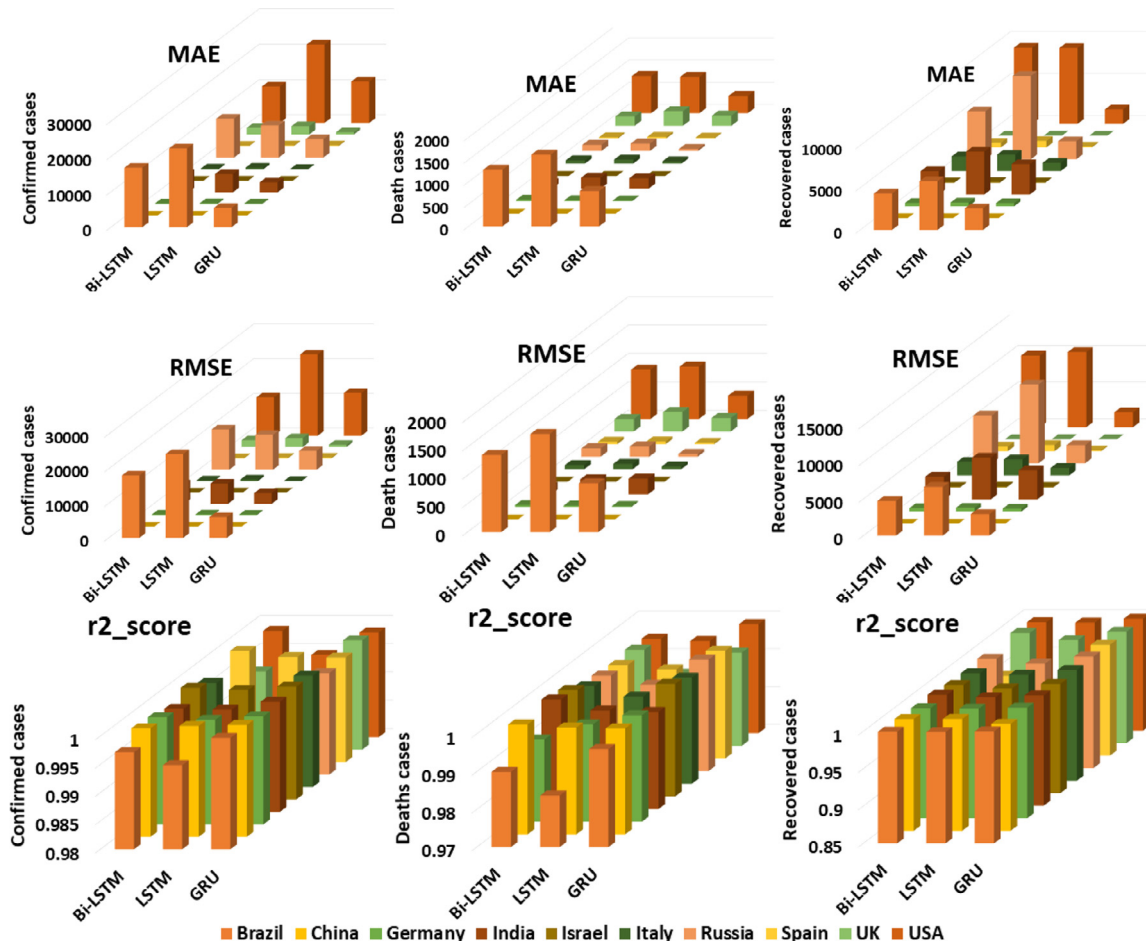
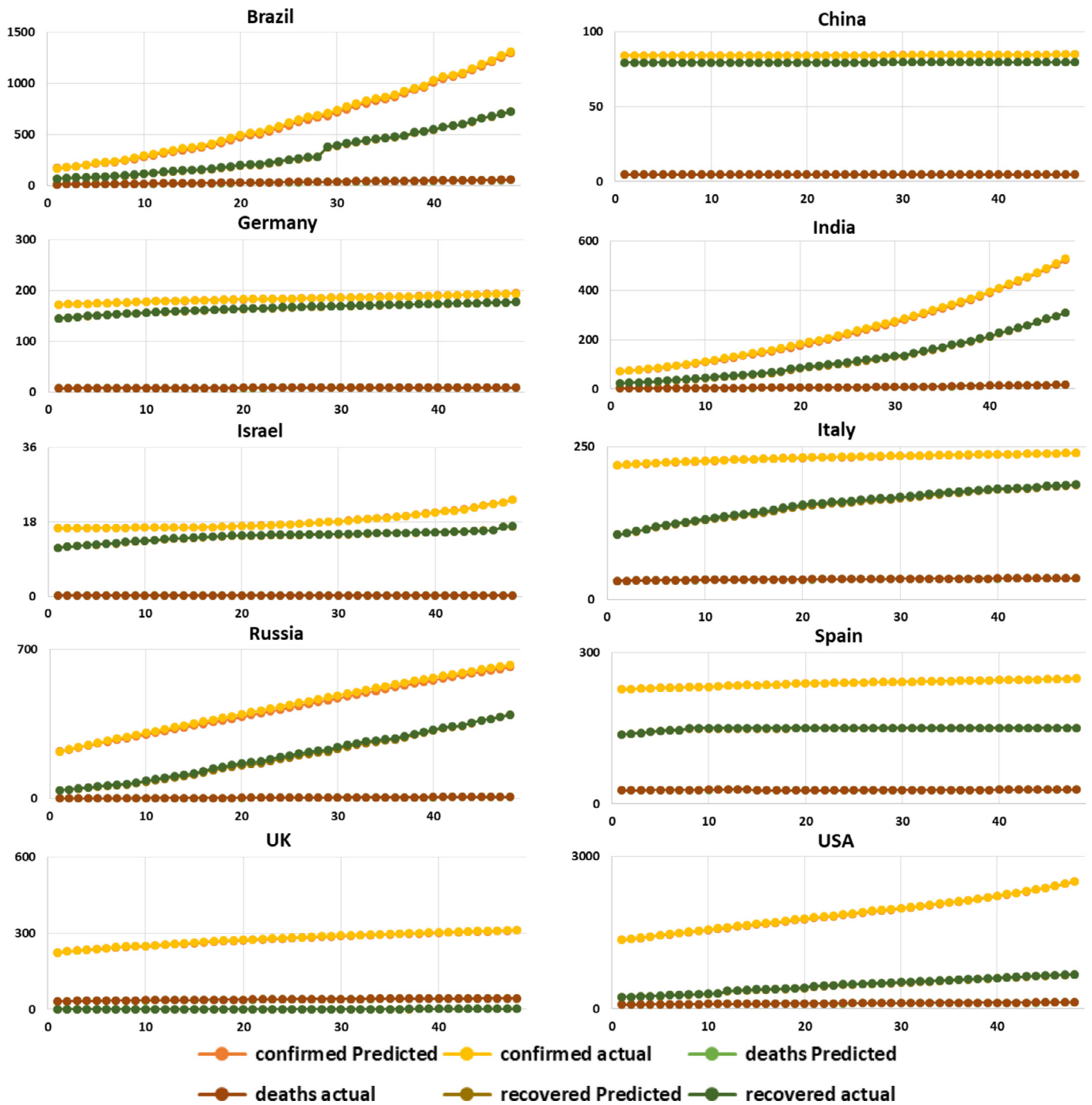**Fig. 3.** Comparison among deep learning techniques in terms of different error measures.

Fig. 4. Scatter plots of actual vs. prediction of ten countries of proposed Bi-LSTM technique.

dicted and actual plots of confirmed cases, death cases and recovered cases of Bi-LSTM are presented in Fig. 4. These scatter plots demonstrate a very good match of predicted cases against actual ones for all three techniques wit much better performance than baseline regressors. Furthermore, among DL models, Bi-LSTM performs very well and its predicted values completely overlap number of actual cases.

Without scaling, LSTM generates lowest MAE values for confirmed cases and deaths as 2.0463 and 0.0095, while RMSE values are 2.2428 and 0.0103, respectively, for China and Bets value for recovered cases is for UK. For LSTM, r2_score value of 0.9996 is the best value for recovered cases in UK. As far as results from

GRU are concerned, it performs best for China in all three categories. Whereas among all countries, Bi-LSTM predicts best three cases for China with highest accuracy among all methods for all countries.

Convergence of loss function for ten countries using GRU has been plotted against number of days for confirmed, death and recovered cases in Fig. 5. These logrithmic graphs demonstrate a smooth evolutionary plot towards converged value of fitness function. For each country this convergence value differs, but overall congverges very well and remains stable and consistent.
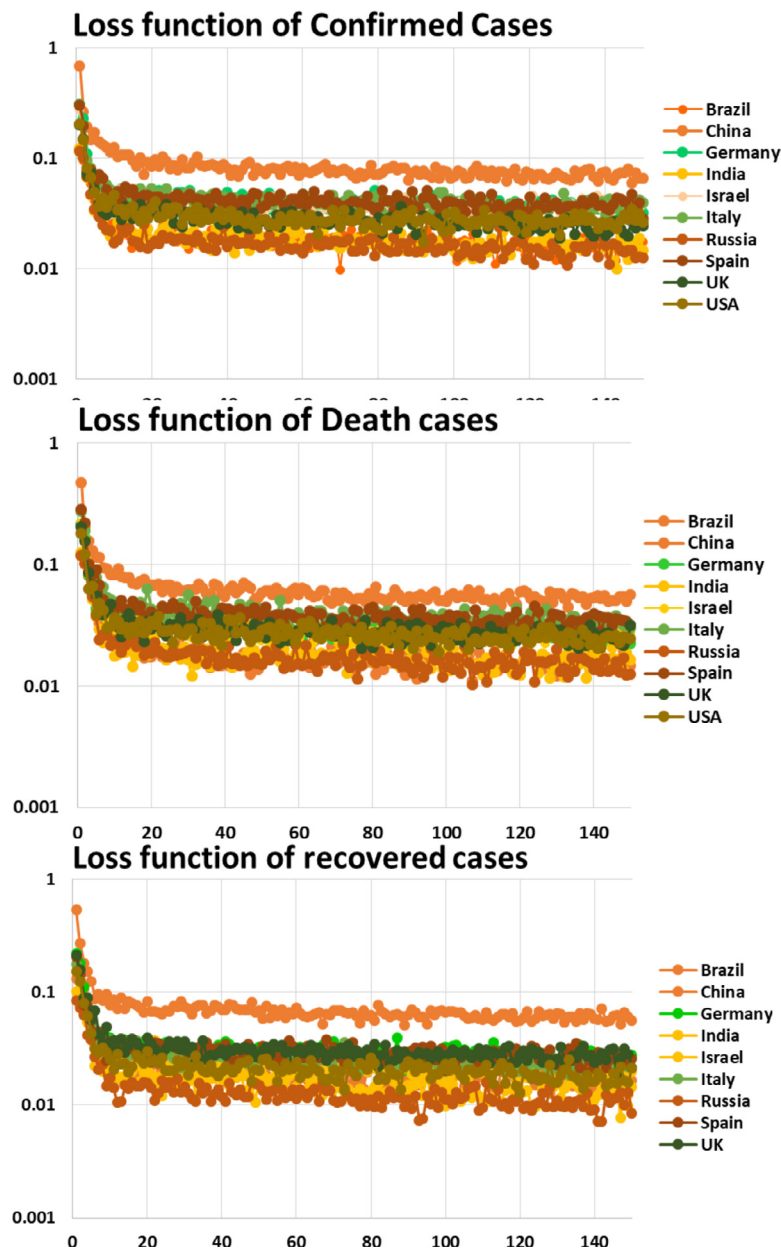
Fig. 5. Loss function of ten countries for three categories of GRU technique.

4. Conclusion

Inferences on the performance of proposed scehmes are listed as follows:

- COVID-19 dataset has been modelled using various regressors including ARIMA, SVR with polynomial and RBF kernels, LSTM, GRU and Bi-LSTM for future predictions on confirmed cases, deaths and recovered case for ten countries across the globe.
- Performance measures of MAE, RMSE and r2_score have been used to compare various models.
- ARIMA and SVR models are unable to follow the trend of these features with higher prediction error and negative values of r2_score.

Without scaling, LSTM generates lowest MAE values for confirmed cases and deaths as 2.0463 and 0.0095, while RMSE values are 2.2428 and 0.0103, respectively, for China and Best value for re-

covered cases is for UK. For LSTM, r2_score value of 0.9996 is the best value for recovered cases in UK.

As far as results from GRU are concerned, it performs best for China with MAE values of 2.8553, 0.0321, and 7.04867 and RMSE values of 3.3158, 0.0402, and 8.4009 for confirmed cases, deaths and recoveries, respectively.

Whereas among all countries, Bi-LSTM predicts best three cases for China with highest accuracy among all methods for all countries with lowest MAE and RMSE values of 0.0070 and 0.0077, respectively, for deaths in China. The best r2_score value is 0.9997 for recovered cases in China.

- LSTM, GRU and Bi-LSTM have shown robustness and much enhanced predictions when compared with actual numbers depicting lower prediction error, however, Bi-LSTM out performed among all models on the basis of three error measures.
- It can be concluded that Bi-LSTM is an appropriate predictor for such sequential data and capable of predicting with enhanced

accuracy for similar other datasets for appropriate planning and better management.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Farah Shahid: Validation, Investigation, Writing - original draft, Visualization, Methodology. **Aneela Zameer:** Conceptualization, Methodology, Writing - original draft, Project administration. **Muhammad Muneeb:** Investigation, Visualization, Methodology.

References

[1] Organization WH . Coronavirus disease 2019 (COVID-19) situation report51, Geneva, Switzerland: World Health Organization; 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200606-covid-19-sitrep-138.pdf?sfvrsn=c8abfb17_4.

[2] Bai Y, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA 2020;323(14):1406–7.

[3] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. In: Proceedings of the royal society of London. Series A, containing papers of a mathematical and physical character, 115; 1927. p. 700–21.

[4] Yasuhiro T, Wanbiao M, Edoardo B. Global asymptotic properties of a delay SIR epidemic model with finite incubation times [J]. Nonlinear Anal 2000;42(6):931–47.

[5] Sharomi O, et al. Mathematical analysis of the transmission dynamics of HIV/TB coinfection in the presence of treatment. Math Biosci Eng 2008;5(1):145.

[6] Willox R, et al. Epidemic dynamics: discrete-time and cellular automaton models. Physica A 2003;328(1-2):13–22.

[7] Knight GM, et al. Bridging the gap between evidence and policy for infectious diseases: how models can aid public health decision-making. Int J Infect Dis 2016;42:17–23.

[8] Fattah J, et al. Forecasting of demand using ARIMA model. Int J Eng Bus Manag 2018;10 p. 1847979018808673.

[9] Benvenuto, D., et al., Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in brief, 2020: p. 105340.

[10] Choi K, Thacker SB. Mortality during influenza epidemics in the United States, 1967-1978. Am J Public Health 1982;72(11):1280–3.

[11] Promprou, S., M. Jaroensutasinee, and K. Jaroensutasinee, Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand using ARIMA Models. 2006.

[12] Liu Q, et al. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect Dis 2011;11(1):218.

[13] Zifeng Yang ZZ, 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(3):165.

[14] Shahid F, et al. A novel wavenets long short term memory paradigm for wind power prediction. Appl Energy 2020;269:115098.

[15] Zameer A, et al. Bio-inspired heuristics for layer thickness optimization in multilayer piezoelectric transducer for broadband structures. Soft Comput 2019;23(10):3449–63.

[16] Hao B, et al. Predicting mental health status on social media International conference on cross-cultural design. Springer; 2013.

[17] Shi, F., et al., Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. arXiv preprint arXiv:2003.09860, 2020.

[18] Ardabili, S.F., et al., Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188, 2020.

[19] Frausto-Solis J, et al. The hybrid forecasting method SVR-ESAR for Covid-19. medRxiv 2020 p. 2020.05.20.20103200.

[20] Parbat D, Chakraborty M. A python based support vector regression model for prediction of COVID19 cases in India. Chaos, Solitons Fractals 2020;138:109942.

[21] Hao T. Prediction of coronavirus disease (covid-19) evolution in USA with the model based on the Eyring rate process theory and free volume concept. medRxiv 2020 p. 2020.04.16.20068692.

[22] Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons, Fractals 2020;135 109864-109864.

[23] Bandyopadhyay SK, Dutta S. Machine learning approach for confirmation of COVID-19 cases: positive, negative, death and release. medRxiv 2020 p. 2020.03.25.20043505.

[24] Huang C-J, et al. Multiple-input deep convolutional neural network model for COVID-19 forecasting in China. medRxiv 2020 p. 2020.03.23.20041608.

[25] Contreras J, et al. ARIMA models to predict next-day electricity prices. IEEE Trans Power Syst 2003;18(3):1014–20.

[26] Adhikari, R. and R.K. Agrawal, An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613, 2013.

[27] Santamaría-Bonfil G, Frausto-Solís J, Vázquez-Rodarte I. Volatility forecasting using support vector regression and a hybrid genetic algorithm. Comput Econ 2015;45(1):111–33.

[28] Yu R, et al. LSTM-EFG for wind power forecasting based on sequential correlation features. Fut Gener Comput Syst 2019;93:33–42.

[29] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735–80.

[30] Zaremba, W., I. Sutskever, and O. Vinyals, Recurrent neural network regularization. arXiv preprint arXiv:1409.2329, 2014.

[31] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. Trans Sig Proc 1997;45(11):2673–81.

[32] Chung, J., et al., Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.

[33] Rana, R., Gated recurrent unit (GRU) for emotion classification from noisy speech. arXiv preprint arXiv:1612.07778, 2016.

[34] Basemap, W.C.-D.C.W., https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/L20LOT.