Available online at sjuoz.uoz.edu.krd

journals.uoz.edu.krd

**Science Journal of University of Zakho**

*Vol. 11, No. 2, pp. 170 –178, April-Jumne, 2023*

p-ISSN: 2410-7549
e-ISSN: 2414-6943

# BAYESIAN DEEP LEARNING APPLIED TO LSTM MODELS FOR PREDICTING COVID-19 CONFIRMED CASES IN IRAQ

Dozdar Mahdi Ahmed [a]* and Masoud Muhammed Hassan[a]
a Faculty of Science, University of Zakho, Zakho, Kurdistan Region, Iraq
(dozdar.ahmed@stud.uoz.edu.krd, masoud.hassan@uoz.edu.krd)

***Received: 12 Oct., 2022 / Accepted: 12 Dec., 2022 / Published: 11 Apr., 2023*** *https://doi.org/10.25271/sjuoz.2023.11.2.1037*

**ABSTRACT**:

The COVID-19 pandemic has had a huge impact on populations around the world and has caused critical problems to medical systems. With the increasing number of COVID-19 infections, research has focused on forecasting the confirmed cases to make the right medical decisions. Despite the huge number of studies conducted to forecast the COVID-19 patients, the use of Deep Learning (DL) and Bayesian DL models are limited in this field in Iraq. Therefore, this research aims to predict the confirmed cases of COVID-19 in Iraq using classical DL models such as, Long-Short-Term-Memory (LSTM) and Bayesian LSTM models. In this study, Bayesian Deep Learning (BDL) using LSTM models was used to predict COVID-19 confirmed cases in Iraq. The motivation behind using BDL models is that they are capable to quantify model uncertainty and provide better results without overfitting or underfitting. A Monte Carlo (MC) Dropout, which is an approximation method, is added to the Bayesian-LSTM to create numerous predictions for each instance and evaluate epistemic uncertainty. To evaluate the performance of our proposed models, four evaluation measures (MSE, RMSE, R2, MAE) were used. Experimental results showed that the proposed models were efficient and provided an R2 of 0.93 and 0.92, for vanilla LSTM and Bayesian-LSTM, respectively. Furthermore, the two proposed models were optimized using ADAM and SGD optimizers, with the results revealing that optimizing with ADAM provided more accurate results. Thus, we believe that these models may assist the government in making critical decisions based on short-term predictions of confirmed cases in Iraq.

KEYWORDS: Deep learning, Bayesian Neural Networks, LSTM, Time-series data, forecasting, MC dropout.

## INTRODUCTION

The new coronavirus disease (COVID-19) is an acute respiratory ailment brought on by (SARS-CoV2) virus. The majority of healthy individuals infected by this illness will experience mild to moderate respiratory illness and recover without the need for special treatment, according to the World Health Organization (WHO) (*Coronavirus*, n.d.). However, early findings suggest that patients with underlying medical diseases such as heart disease (Bansal, 2020), diabetes (Hussain et al., 2020), chronic lung illness (Lai et al., 2020), obese (Abbas et al., 2020), and cancer (Moujaess et al., 2020) are much more likely to sustain major conditions when infected by COVID-19 (Su et al., 2020). Furthermore, the COVID-19 has the potential to cause significant and critical lung damage (Guan et al., 2020), putting patients' respiratory systems at risk. The development of effective and efficient forecasting models has a positive impact on the production of reasonably accurate success rates forecasts in the near future because it is crucial to comprehend the challenging epidemiological scenario for COVID-19 on a short-term horizon in order to mitigate the pandemic's effects. Therefore, precise model development will give health managers the best possible foundation for strategic planning and decision-making. Epidemiological models, which have been frequently employed by health researchers (Ndaïrou et al., 2020), (Barmparis & Tsironis, 2020), may be used for this purpose. Alternatively, hybrid forecasting models (Chakraborty & Ghosh, 2020), (Singh et al., 2020), and artificial intelligence (AI) techniques (Ribeiro et al., 2020), (Chimmula & Zhang, 2020) have proven to be excellent tools for predicting COVID-19 situations. The adaptability of AI algorithms for time series forecasting stem from their capacity to cope with a wide range of response variables, as well as their ability to learn data dynamical behavior, complexity, and accept nonlinearities (Dal Molin Ribeiro et al., 2019).

As COVID-19 data are sequential, using time series models to cope with their dynamic nature is strongly recommended for forecasting. Different time series models exist in the literature, including statistical models, such as Auto-Regressive (AR), Moving Average (MA), and Auto-Regressive Integrated Moving-Average (ARIMA). On the other hand, deep sequential models, such as Recurrent Neural Networks (RNN) and LSTM models, have been proven to provide more accurate and reliable forecasting for time series data, and the latter models are now widely used. For example, Bandyopadhyay et al. presented the gated recurrent neural network with LSTM (*Machine Learning Approach for Confirmation of COVID-19 Cases*, n.d.) to assess the predictions with confirmation, negative released, and mortality instances of COVID-19. Although deep sequential models produce outstanding forecasting results, they are incapable to quantify model uncertainty. To address this issue, Bayesian Deep Learning (BDL) models provide a solid framework for managing and quantifying the two main sources of uncertainty: Aleatoric (data uncertainty) and Epistemic (model uncertainty). In addition, BDL can deal with short sequences and small datasets without overfitting or underfitting. Therefore, Bayesian modelling is essential since it accounts for the model uncertainty, particularly noisy data, when drawing any conclusions on COVID-19 development.

For the prediction of COVID-19 data in Iraq, many researchers have used different methods, such as using ANNs (BF, NARX, FCM) (Yahya et al., 2021), Support Vector Machine (SVM), Decision Tree (DT) and Naive Bayes (Awlla et al., 2021). To predict the future trend of COVID-19 in Iraq, Gaussian Process regression was also used by (Aldeer et al., 2021), and Susceptible-infected-removed (SIR) epidemic models to predict susceptible populations to SARS-CoV-2 infection by

Corresponding author

(Mohammed et al., 2021). We were unable to find any research using LSTM and Bayesian LSTM in Iraq.

From the literature analysis, we may learn that there are numerous time-series prediction models, each of which performs under certain situations and has different limitations. Accurate predictions are often made with deep learning models, LSTM being one of them. This paper aims to use LSTM and Bayesian LSTM models to predict confirmed COVID-19 case in Iraq. In the case of Iraq, there are several constraints, and the literature revealed that the research in this field is quite limited. To the best of the author's knowledge, this is the first attempt to use Bayesian LSTM models for analyzing and forecasting COVID-19 data in Iraq. With our Bayesian LSTM model, we also aim to quantify the model uncertainty, particularly Epistemic uncertainty.

The data set of confirmed COVID-19 cases used in this experiment is accessible. Because of the complex nature of coronaviruses, DL models have been used to make early predictions. To anticipate COVID-19 confirmed cases, DL mechanisms of LSTM and Bayesian LSTM are suggested. Model accuracy is tested using four performance measures: MSE, MAE, RMSE, and R2 score.

The following parts of this paper are structured as follows. Section 2 addresses related work for LSTM-based COVID19 prediction. Section 3 outlines the background and our proposed methods. Section 4 provides our experimental results and discussion. Finally, the main conclusions drawn are reported in Section 5.

## RELATED WORKS

Recently, several studies have been conducted using statistical time-series models, deep learning models, LSTM and Bayesian deep learning to predict COVID-19 data. Some of such models have been developed to address the uncertainty in models that are applied on the different public datasets for COVID-19. For example, Kırbaş et al. (2020) used LSTM, ARIMA, and Nonlinear Autoregression Neural Network (NARNN) techniques to model COVID-19 instances from various countries, including Denmark, Belgium, Turkey. The most accurate model was chosen using six evaluation performance metrics (MSE, PSNR, RMSE, NRMSE, MAPE, and SMAPE). According to their findings, the LSTM model was shown to be the most accurate model.

In another study conducted by Chimmula and Zhang (2020), they employed LSTM models for COVID-19 transmission time series forecasting in Canada. When the transmission rates of Canada, Italy, and the United States were compared, the authors claimed that the outbreak would end around June 2020. Canada achieved its daily new case peak on May 2, 2020, and since then, the number of new cases has declined significantly. They concluded that their proposed method to describe the peak of COVID-19 in Canada was pretty accurate.

ArunKumar et al. (2021) suggested utilizing DL models with recurrent neural networks, LSTM and Gated Recurrent unit (GRU), for 60-day prediction of the COVID-19 pandemic based on cumulative confirmed cases, recovered cases, and mortality by country. The model with the lowest RMSE and MSE values were thought to be the best for forecasting. They conclude that, for confirmed instances, the LSTM model outperformed the GRU model in countries including the United States, Brazil, South Africa, Chile, Iran and Peru, while the GRU model outperformed in Russia, India, Mexico, and the United Kingdom. In terms of recovered cases, the LSTM model outperformed the GRU model in India, South Africa, Chile, the United Kingdom, and Iran, while the GRU model provided better results in the USA, Russia, Mexico, Brazil and Peru. For India, Brazil, South Africa, Russia, Iran and Mexico, the LSTM models beat the GRU in the prediction of death data. GRU models beat LSTM models for death data in the USA, Peru, Chile, and the UK.

Tomar and Gupta (2020) created an LSTM model for predicting COVID-19-positive cases 30 days in advance in India, where they also investigated the influence of preventative interventions on the spread of COVID-19. Their results demonstrated that with preventative measures and a lower transmission rate, its spread may be greatly curtailed.

Rao et al. (2020) used the climatic conditions of different states to enhance COVID-19 case prediction in several Indian states. The authors hypothesized that humidity levels in various states would result in differential virus transmission among the population. They proved that the LSTM model performed well at the medium and long-range predicting scales when climate data were included.

Nadler et al. (2021) created a prediction model that combines the epidemiological dynamics of compartmental models with the extremely nonlinear interactions learnt by an LSTM Network, and novel dynamic variables associated with the population transmission of Covid-19 are fitted to the SIR model. This is then implemented into a Bayesian recursive updating framework and combined with an LSTM network to anticipate Covid-19 instances. The model greatly outperforms basic univariate LSTM and SIR models in terms of forecast accuracy.

Gautam (2022) proposed using transfer learning on an LSTM network to understand patterns of COVID-19 in new cases and deaths using data in Italy and the United States, as well as forecasting for other countries. Single step and multi-steps forecasts from the constructed models were also tested in Germany, France, Brazil, India, and Nepal. Their results revealed that the suggested models can properly predict new cases and deaths.

Shetty and Pai (2021) demonstrated real-time prediction using a basic neural network for COVID-19 instances in the Indian state of Karnataka using the cuckoo search technique for parameter selection. According to their findings, the MAPE was reduced from 20.73% to 7.03%, and the proposed model was tested on the Hungary COVID-19 set with promising results.

Bodapati et al. (2020) also used LSTM networks to predict the daily COVID-19 cases, deaths, and recovered cases for the whole world, and datasets were collected from Johns Hopkins University's publicly accessible datasets, and encouraging experimental results have been achieved.

Rauf et al. (2021) used the most up-to-date deep learning methods LSTM, RNN, and (GRU) to estimate the severity of pandemics in India, Pakistan, Afghanistan and Bangladesh in the near future (10 days). The models' results have predicted an accurate rate of more than 90%, indicating that the suggested models were valid.

Istaiteh et al. (2020) examined the accuracy of ARIMA, multilayer perceptron, LSTM and CNN models for the global prediction of COVID-19 instances. According to their findings, DL models beat the ARIMA model, while CNN beats LSTM network and multilayer perceptron's.

It is believed that the LSTM models produce promising outcomes for forecasting time-series data based on the literature of COVID-19 experiments. As a result, we were even more motivated to use this model and create a Bayesian framework based on it to anticipate COVID-19 data in Iraq.

## BACKGROUND AND METHODOLOGY

### LSTM network for time series modelling

LSTM is one of the recurrent neural networks (RNN) used for sequential data. To overcome the issue of vanishing gradients in RNN networks, (Hochreiter & Schmidhuber, 1997) suggested the LSTM structure, which contains memory cells in each memory block. Each memory block is provided with input and output gates to control the inflow of information. The mathematical equations of the LSTM model are given below, and the architecture of this model is shown in Fig. 1.

$$f_t = \sigma(w_f[v_{t-1}, x_t] + b_f) \quad (1)$$
$$m_t = \sigma(w_f[v_{t-1}, x_t] + b_m) \quad (2)$$
$$N_t = tanh(w_f[v_{t-1}, x_t] + b_n) \quad (3)$$
$$C_t = C_{i-1}f_t + N_t m_t \quad (4)$$
$$Q_t = \sigma(w_q[v_{t-1}, x_t] + b_q) \quad (5)$$
$$v_t = Q * tanh(C_n) \quad (6)$$

Where $f_t$is the forget gate at time $t$, $m_t$ is the input gate , $Q_t$ is output gate, $C_t$ is the cell stat, $N_t$ is the update layer, $v_t$ is the hidden state, $w_f$ is weights, $\sigma$ is sigmoid funtion, $b$ is bias , and $x_t$ input at time $t$.

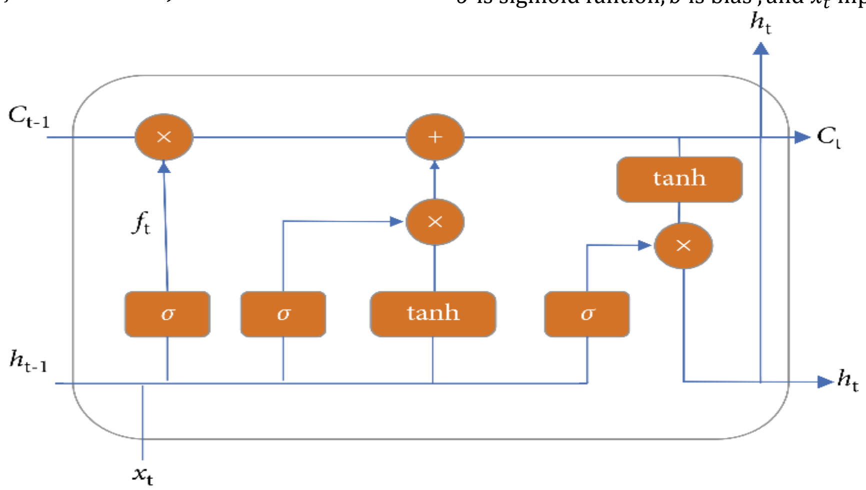


Fig. 1: Vanilla LSTM model architecture (D. Ahmed et al., 2022). $x_t$ illustrate input data, $h_{t-1}$ is previous hidden state, $C_{t-1}$ is previous cell state in this layer, $f_t$ is the forget gate, which pass information to input gate then data passed throw sigmoid and tanh activation function to find the $C_t$ new hidden cell, and $h_t$ is the new hidden state.

**Bayesian LSTM with MC dropout Approximation Method**

To give prediction probabilities, Bayesian models, such as Bayesian neural networks (BNNs) have attracted a lot of attention. In contrast to single point predictions, BNNs produce predictive distributions where the posterior distribution of weights are estimated by combining the likelihood of data with the prior distributions of weights (Mullachery et al., 2018). Making inference about those parameters and estimating posterior distributions is a challenging issue. Especially, when dealing with a complex model with a large number of parameters, such as neural networks, the posterior distribution is intractable analytically and it is computationally expensive. To deal with this issue, researchers developed approximation methods, such as Monte-Carlo dropout (MC-dropout) and Variational Inference (VI). The MC-dropout technique, which is computationally far more efficient than sampling method for BNNs, has recently seen a revival of attention as an evolution of Bayesian methods (Alarab et al., 2021). The use of Monte Carlo dropout has grown for evaluating model uncertainty for neural network outputs.

Given some training data $\mathcal{D} = (x_i, y_i)$ where $\mathbf{x} = \{x_1, ..., x_n\}$ are input and their corresponding outputs $y = \{y_1, ..., y_n\}$, with their weights $\Theta = \{\theta_1, ..., \theta_n\}$, the MC-dropout can be used to find a predictive distribution

$$\mathcal{P}(y^*|\mathcal{D}, \Theta) \quad (7)$$

When a predictive distribution is obtained, the variance may be examined to reveal uncertainty. One method for learning a predictive distribution involves learning a distribution over functions, or equivalently, a distribution over the weights (the parametric posterior distribution)

$$\mathcal{P}(\Theta|\mathcal{D}) \quad (8)$$

The Monte Carlo (MC) dropout method proposed by Gal and Ghahramani (Gal & Ghahramani, 2016) offers a scalable method for acquiring knowledge of a predictive distribution. MC-dropout functions by randomly deactivating neurons in a neural network, therefore regularizing the network. Each dropout configuration corresponds to a distinct sample from the approximation of the parametric posterior distribution, $q(\Theta|\mathcal{D})$, with

$$\Theta_t \sim q(\Theta|\mathcal{D}) \quad (9)$$

If $\Theta_t$ is a dropout configuration, a simulation samples are taken from its parametric posterior $q(\Theta|\mathcal{D})$ distribution as shown in following equation. Sampling from the estimated posterior $q(\Theta|\mathcal{D})$ permits Monte Carlo integration of the model's likelihood, which reveals the predictive distribution in the following manner:

$$\mathcal{P}(y|x) \approx \int_{\Omega} \underbrace{\mathcal{P}(y|x,\Theta)}_{likelihood} \underbrace{q(\Theta|\mathcal{D})}_{p.\ posterior} d\Theta \quad (10)$$

$$MC \approx \frac{1}{T}\sum_{t-1}^{T} \mathcal{P}(y|x,\Theta_t), \quad s.t. \Theta_t \sim q(\Theta|\mathcal{D}) \quad (11)$$

The probability distribution may be assumed to have a Gaussian distribution for the sake of simplicity.

$$\mathcal{P}(y|x,\Theta) = \mathcal{N}(f(x,\Theta), s^2(x,\Theta)) \quad (12)$$

With the mean $f(x,\Theta)$ and variance $s^2(x,\Theta)$ functions of the parameters, which are produced via Monte Carlo dropout in BNN simulations as follows:

$$f(x,\Theta), s^2(x,\Theta) \sim MC - Dropout(x)$$

Fig. 2 depicts the MC dropout. Multiple forward passes with various dropout settings result in a predictive distribution over the mean $p(f(x,\Theta))$. The number of forward passes over the data should be quantified, but 30-100 is a suitable range to examine.

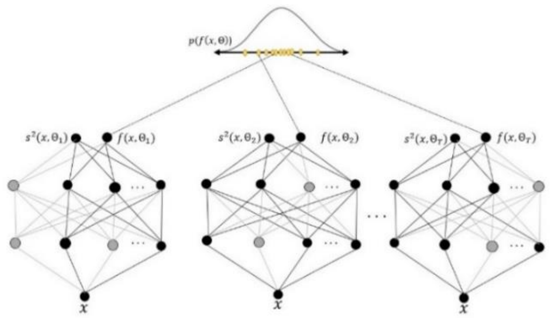

Fig. 2: MC dropout (Davis et al., 2020) with two types of circles: black and gray. Each one represents a new output by randomly switching off (grey circles) and activating neurons (black circles) during each forward propagation. Numerous forward passes with various dropout settings produce in a predictive distribution over the mean $\boldsymbol{p}(\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\Theta}))$.

During the training phase for each repeat or epoch, the dropout approach involves turning off or deleting certain neurons in a particular layer with a specified probability (Gal & Ghahramani, 2016), and this process is known as MC-dropout. It is worth noting that this approximation method is comparable to how the posterior distribution is estimated using Variational Inference (VI). MC-dropout technique is quicker than VL since it contains less parameters and needs less time for the models to be converged (Abdullah et al., 2022).

**Model Performance**

In time series prediction, the model is evaluated using the regression metric (D. Ahmed et al., 2022). In this work, four different evaluation metrics are used for checking the performance of our predictive models. These metrics are: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared ($R^2$), and their formulas are:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (14)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (15)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (16)$$

Where $n$ is the number of data points, $y_i$ and $\hat{y}_i$ are observed and predicted data points, respectively.

## RESULT AND DISCUSSION

**Dataset**

For all the experiments in this section, the univariate time series data, which are confirmed cases data for Covid19 in Iraq, were used for modeling. The dataset used in this study is obtained from the Johns Hopkins University, Center for Systems Science and Engineering (CSSE). These data are collected from around the world and updated daily. The data we used are collected from 22-01-2020 until 15-06-2022 in Iraq. The first two months were skipped due to the data being abnormal. The cumulative data is shown in Fig. 3.

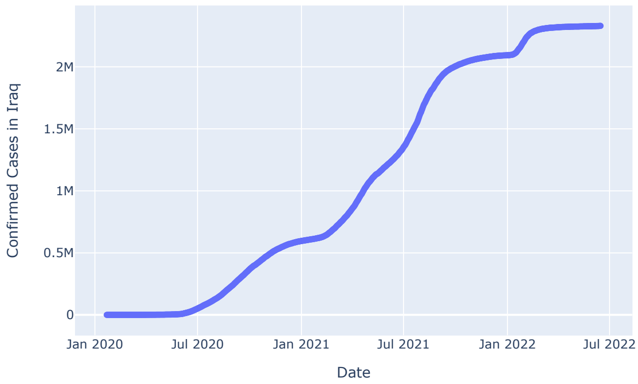


Fig. 3: Cumulative curve of confirmed cases of COVID-19 in Iraq.

To extract confirmed cases, we used the difference function to undo the accumulation, as shown in Fig. 4.

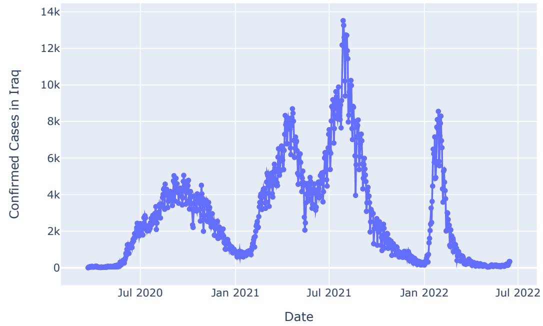


Fig. 4: Confirmed cases of COVID-19 in Iraq.

**Data Pre-processing**

Before applying our data for analyzing and predicting neural network models, they must be normalized to be in the same scale. For this purpose, Min-Max scaler is used to normalize the data, and Eq. 17 represents Min-Max scaler mathematically.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (17)$$

Where $X$, $X_{scaled}$ , $X_{min}$, $X_{max}$ denotes the real time-series data, the normalized data, and the minimum and maximum values, respectively. The Min-Max scaler transforms each observation inside the feature so that its lowest and maximum values fall within the range of (0 to 1). The benefit of utilizing Min-Max scaler is that after converting data, Min-Max scaler preserves the original shape of the data distributions and does not affect the data's contained patterns. Normalized data were then used to construct training and validation sets, where 80% were employed for training set and 20% for the validation set. The training time series data is separated into numerous data sequences, each of length 7. Each sequential data point is provided to the LSTM model (i.e., the data from days 1 to7) to predict the day 8 data point, and so on. The data analyses were implemented in Python using the TensorFlow 2.8.0 package for deep learning written in Jupiter notebook.

**Experimental Results**

In this work, we applied two models of LSTM. The first was vanilla LSTM and the second was Bayesian (probabilistic) LSTM with MC-dropout.

**Vanilla LSTM**

Vanilla LSTM was applied in two architectures; the structure of LSTM-1 is made up of three layers. Each layer has 50 neurons with a dropout probability of 0.3 and is optimized with the SGD optimizer. The SGD optimizer has a hyperparameter learning rate of 0.1 and a momentum of 0.6. 20% of data were used for validation and the remainder were used for training. The model run for different epochs, and after 100 epochs the model learned to converge with the MSE loss function compiled, as shown in Fig. 5. The structure of LSTM-2 is made up of two layers. Each layer has 100 neurons with a dropout of 0.10 and is optimized with the ADAM optimizer. 20% of data were used for validation and the remainder for training. The model learned to converge after 200 epochs with the MSE loss function compiled, as shown in Fig. 6.

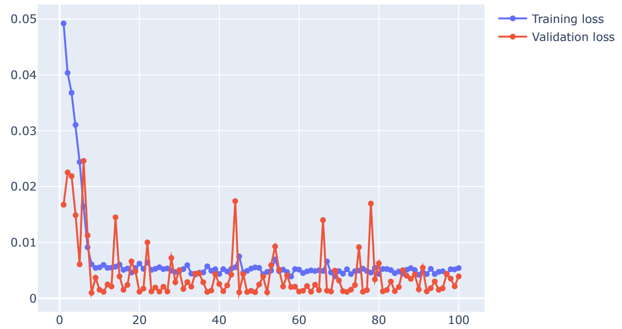


Fig. 5: Training LSTM model with SGD optimizer.

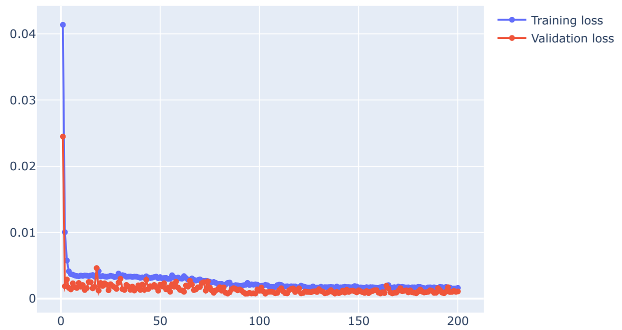


Fig. 6: Training LSTM model with ADAM optimizer.

It can be seen that both Figures 5 and 6 are well converged after we find the appropriate weights. In Figure 5, the vanilla LSTM with SGD optimizer learning in hundred epochs, our model gets a large peak in the first few epochs and some peaks in forty and eighty epochs. There are a few peaks at the beginning of Figure 6, then the model converges with ADAM optimizer.

The trained models were then applied to test the data and predict confirmed cases, as shown in Figures 7 and 8.

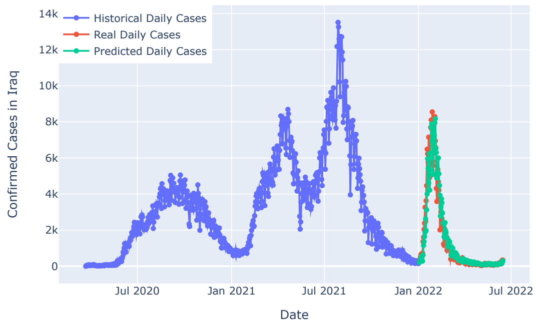


Fig. 7: Testing LSTM model with ADAM optimizer.

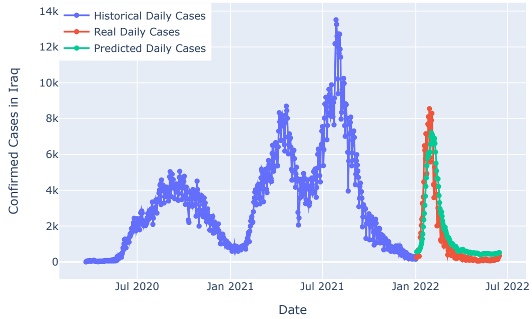


Fig. 8: Testing LSTM model with SGD optimizer.

Comparing the two figures, 7 and 8, it was demonstrated that LSTM model with ADAM optimizer performed better and provided more accurate prediction results according to the performance evaluation shown in Table 1.

After simulation and data fitting to our models, we then forestated 10-day prediction of confirmed cases, and the results are illustrated in Figures 9 and10.

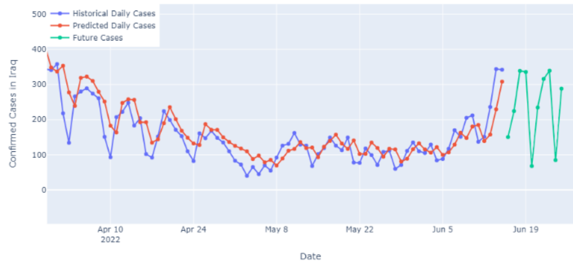


Fig. 9: Forecasting 10-day prediction based on LSTM model with ADAM optimizer.

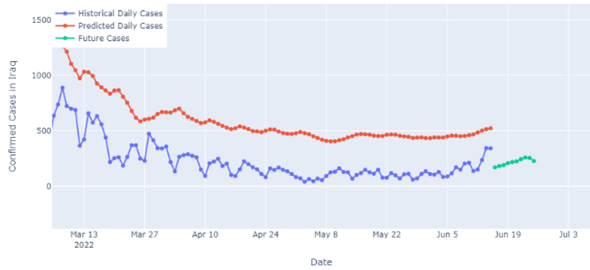


Fig. 10: Forecasting 10-day prediction based on LSTM model with SGD optimizer.

**Bayesian LSTM**

In probabilistic Bayesian LSTM, the same structures as in Vanilla LSTM were used, and we added Monte Carlo dropout simulation by using various random samples from the distribution to make inference about the posterior distribution and quantify model uncertainty from the determined values.

The probabilistic LSTM training model with SGD optimizer is depicted in Fig. 11, and with ADAM optimizer in Fig. 12, after the model converges.

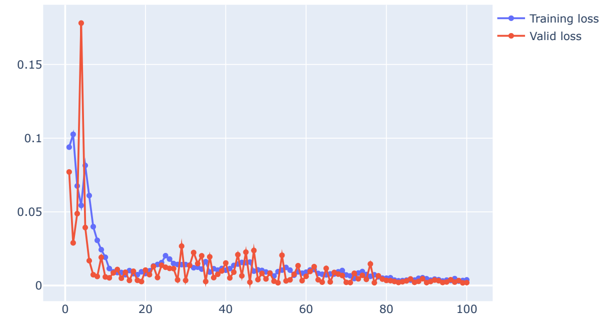


Fig. 11: Bayesian-LSTM model trained with SGD optimizer.

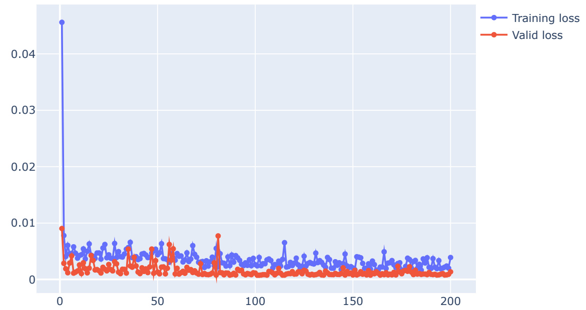


Fig. 12: Bayesian-LSTM model trained with ADAM optimizer.

To quantify the model uncertainty, epistemic uncertainty, the posterior distribution of the model parameters is inferenced based on Bayesian framework which was used in Bayesian LSTM model. This is especially challenging in neural networks because of the non-conjugacy often caused by nonlinearities, in which we approximated the posteriors using MC-dropout method. In this way, dropout was implemented at both training and validation sets. At validation, the data were transmitted across the network many times, with various parameters being discarded at each run. The results were then averaged across the number of runs to produce posterior samples and hence make inference.

Again, the MC-dropout method was estimated to quantify model uncertainty with both ADAM and SGD optimizers. Figures 13 and 14 show the model uncertainty during training data that covering 99, 97, and 95 percent of the uncertainty. In Figures 15 ,16 it is shown how the models perform on test data by looking at epistemic uncertainty which cover 99, 97, and 95 percent of the uncertainty.

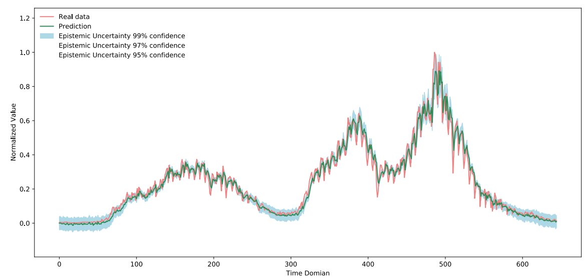


Fig. 13: Epistemic uncertainty of Bayesian LSTM model for training data using the SGD optimizer.

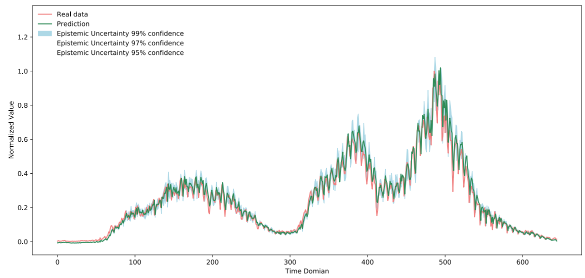


Fig. 14: Epistemic uncertainty of Bayesian LSTM model for training data using the ADAM optimizer.

For the testing data, different sampling outputs were used in predictions for every given test dataset. Then, we found the means and standard deviation of the posterior distribution to find epistemic uncertainty, as shown in Figures 15 and 16.

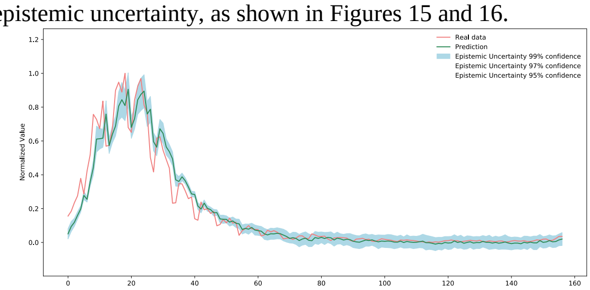


Fig. 15: Epistemic uncertainty of Bayesian LSTM model for testing data using the SGD optimizer.

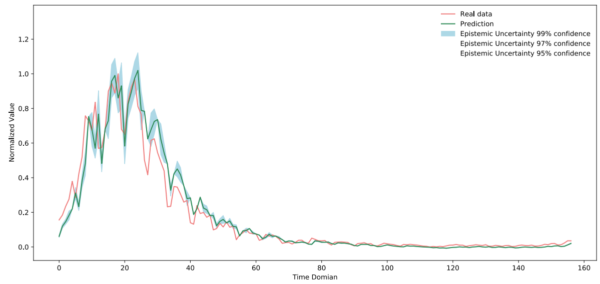


Fig. 16: Epistemic uncertainty of Bayesian LSTM model for testing data using the ADAM optimizer.

Table 1: Performance measures of LSTM and Bayesian LSTM models with ADAM and SGD optimizers.

| Model | | Vanilla LSTM | Bayesian LSTM | Vanilla LSTM | Bayesian LSTM |
|---|---|---|---|---|---|
| Optimizer | | ADAM | | SGD | |
| Training evaluation | *RMSE* | 0.037 | 0.039 | 0.056 | 0.045 |
| | *MSE* | 0.001 | 0.002 | 0.003 | 0.002 |
| | *MAE* | 0.024 | 0.027 | 0.037 | 0.032 |
| | $R^2$ | 0.96 | 0.96 | 0.91 | 0.94 |
| Testing evaluation | *RMSE* | 0.066 | 0.07 | 0.107 | 0.073 |
| | *MSE* | 0.004 | 0.005 | 0.011 | 0.005 |
| | *MAE* | 0.035 | 0.039 | 0.061 | 0.042 |
| | $R^2$ | 0.93 | 0.92 | 0.81 | 0.92 |

The two models investigated in this work performed well on training and testing, and results in Table 1 shows that there was no overfitting or underfitting in models. When using the ADAM optimizer instead of the SGD, the Bayesian LSTM model performed better. The Bayesian LSTM was able to quantify model uncertainty in addition to model predictions. Therefore, we believe that for forecasting COVID-19 data in Iraq, LSTM and Bayesian LSTM models are effective.

**Comparison with the state-of-the-art studies**

In our search for the state-of-the-art studies of COVID-19 prediction in Iraq, a variety of classical methods that were used in each study were found, which are shown in Table 2. However, Deep Learning (DL) and Bayesian DL models have not been used for the prediction of COVID-19 in Iraq before. This is why we used BDL method as a unique method in Iraq.

Table 2: Comparison of research related to COVID-19 forecasts in Iraq

| References | Models | Data source | Results |
|---|---|---|---|
| (Bouhamed, 2020) | LSTM | ecdc.europa.eu web site | The best prediction among the various countries (including Iraq) was obtained in France's with $R^2$ of 0.9. |
| (Ali et al., 2021) | Hybrid of K-Means and Partitioning Around Medoids (PAM) | Many Iraqi clinics | They discovered that K-MP is more effective than K-Means and PAM in determining patient status. MSE of K-MP=0.7. |
| (Mustafa & Fareed, 2020) | ARIMA | Iraqi Ministry of Health | The ARIMA (2,1,5) model was determined to be an effective and suitable model for sequence data. |
| (A. Ahmed et al., 2020) | Euler's method, Runge–Kutta method of order two (RK2) and of order four (RK4). | WHO | The model was reasonable in describing this pandemic illness to make future projections about the number of infected, susceptible, and recovered patients. |
| (Ibrahim & Al-Najafi, 2020) | logistic regression, and Gaussian models. | Worldometer website | The models are a reasonable fit for the incidence data. |
| (Yahya et al., 2021) | ANNs (RBF, NARX, FCM) | Iraqi Ministry of Health | Results show that the spread severity will intensify in this short term by 17.1%, and the average death cases will increase by 8.3%. |
| (Awlla et al., 2021) | Support Vector Machine (SVM), Decision Tree (DT) and Naive Bayes (NB) | Hospitals in Sulaymaniyah | It is shown that SVM, DT, and NB algorithms can classify COVID-19 patients, and DT was the best one with an accuracy of 96.7 %. |
| (Aldeer et al., 2021) | Gaussian Process regression | Johns Hopkins Coronavirus Resource Center | Project the future trend of COVID-19 in Iraq. |
| (Mohammed et al., 2021) | Susceptible-infected-removed (SIR) epidemic models | KRI official website | The model can predict susceptible populations to SARS-CoV-2 infection |
| Proposed Method | LSTM and Bayesian B-LSTM | Johns Hopkins Coronavirus Resource Center | $R^2$ = 0.93 (LSTM)<br>$R^2$ = 0.92 (B-LSTM)<br>MSE = 0.004 (LSTM)<br>MSE = 0.005 (B-LSTM) |

From Table 2, we can see that the only study that used the LSTM method was the study of Bouhamed (Bouhamed, 2020), which predicted COVID-19 in several countries, including Iraq. His results for the case of Iraq were not discussed, while he observed the best prediction obtained in France. The results of our proposed models for confirmed cases COVID-19 in Iraq were efficient and provided an $R^2$ of 0.93 and 0.92 for vanilla LSTM and Bayesian-LSTM, respectively. Furthermore, the lowest MSE of 0.004 and 0.005 was obtained with ADAM optimizer for the two proposed models vanilla LSTM and Bayesian-LSTM, respectively.

## CONCLUTION

COVID-19 pandemic was brought on by severe coronavirus mutations and had a significant effect on people's lives all across the world. Researchers have focused their efforts on determining the risk that COVID-19 infections would ultimately result in patient mortality as the number of severe COVID-19 cases increased globally. It has been discovered that forecasting COVID-19 patients can help scientists and doctors better grasp the disease's severity, amount of risk, and most crucially, the type of medical care each patient will need. Different statistical and DL models are used for this purpose, yet there is a lack of research regarding using Bayesian deep learning models.

In this study, we proposed using probabilistic Bayesian DL with LSTM models for predicting COVID-19 confirmed cases in Iraq. Each model was optimized with two optimizers, ADAM and SGD. The vanilla LSTM models were utilized to predict 10 days ahead of confirmed cases, while the probabilistic LSTM models were used to quantify epistemic uncertainty in addition to data predictions. Our experimental results showed that forecasting with Bayesian LSTM model is more effective as it provides good prediction with the model uncertainty. Based on different evaluation metrics used, results revealed that optimizing our proposed model with ADAM, can provide more accurate results, and LSTM and Bayesian LSTM obtained an $R^2$ of 0.93 and 0.92, respectively. For future work, we aim to apply our proposed models on different datasets with more complex patterns to gain more understanding of the model behaviors.

**REFERENCES**


Abbas, A. M., Fathy, S. K., Fawzy, A. T., Salem, A. S., & Shawky, M. S. (2020). The mutual effects of COVID-19 and obesity. *Obesity Medicine*, *19*, 100250. https://doi.org/10.1016/j.obmed.2020.100250

Abdullah, A. A., Hassan, M. M., & Mustafa, Y. T. (2022). A Review on Bayesian Deep Learning in Healthcare: Applications and Challenges. *IEEE Access*, *10*, 36538–36562. https://doi.org/10.1109/ACCESS.2022.3163384

Ahmed, A., Salam, B., Mohammad, M., Akgül, A., H. A. Khoshnaw, S., 1 Department of Mathematics, College of Basic Education, University of Raparin, Kurdistan Region of IRAQ, & 2 Department of Mathematics, Art and Science Facalty, Siirt University, Siirt, TURKEY. (2020). Analysis coronavirus disease (COVID-19) model using numerical approaches and logistic model. *AIMS Bioengineering*, *7*(3), 130–146. https://doi.org/10.3934/bioeng.2020013

Ahmed, D., Hassan, M., & Mstafa, R. (2022). A Review on Deep Sequential Models for Forecasting Time Series Data. *Applied Computational Intelligence and Soft Computing*, *2022*. https://doi.org/10.1155/2022/6596397

Alarab, I., Prakoonwit, S., & Nacer, M. I. (2021). Illustrative Discussion of MC-Dropout in General Dataset: Uncertainty Estimation in Bitcoin. *Neural Processing Letters*, *53*(2), 1001–1011. https://doi.org/10.1007/s11063-021-10424-x

Aldeer, M., Hilli, A. A., & Ismail, I. S. (2021). Projecting the Short-Term Trend of COVID-19 in Iraq. *Digital Government: Research and Practice*, *2*(1), 1–7. https://doi.org/10.1145/3431769

Ali, N. G., Abed, S. D., Shaban, F. A. J., Tongkachok, K., Ray, S., & Jaleel, R. A. (2021). Hybrid of K-Means and partitioning around medoids for predicting COVID-19 cases: Iraq case study. *Periodicals of Engineering and Natural Sciences (PEN)*, *9*(4), 569. https://doi.org/10.21533/pen.v9i4.2382

ArunKumar, K. E., Kalaga, D. V., Kumar, Ch. M. S., Kawaji, M., & Brenza, T. M. (2021). 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*, *146*, 110861. https://doi.org/10.1016/j.chaos.2021.110861

Awlla, A. H., Muhammed, B. T., Murad, S. H., & Ahmad, S. N. (2021). Prediction of CoVid-19 mortality in Iraq-Kurdistan by using Machine learning. *UHD Journal of Science and Technology*, *5*(1), 66–70. https://doi.org/10.21928/uhdjst.v5n1y2021.pp66-70

Bansal, M. (2020). Cardiovascular disease and COVID-19. *Diabetes & Metabolic Syndrome*, *14*(3), 247–250. https://doi.org/10.1016/j.dsx.2020.03.013

Barmparis, G. D., & Tsironis, G. P. (2020). Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. *Chaos, Solitons, and Fractals*, *135*, 109842. https://doi.org/10.1016/j.chaos.2020.109842

Bodapati, S., Bandarupally, H., & Trupthi, M. (2020). COVID-19 Time Series Forecasting of Daily Cases, Deaths Caused and Recovered Cases using Long Short Term Memory Networks. *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, 525–530. https://doi.org/10.1109/ICCCA49541.2020.9250863

Bouhamed, H. (2020). *Covid-19 Cases and Recovery Previsions with Deep Learning Nested Sequence Prediction Models with Long Short-Term Memory (LSTM) Architecture*. 7.

Chakraborty, T., & Ghosh, I. (2020). Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons, and Fractals*, *135*, 109850. https://doi.org/10.1016/j.chaos.2020.109850

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons, and Fractals*, *135*, 109864. https://doi.org/10.1016/j.chaos.2020.109864

*Coronavirus*. (n.d.). Retrieved May 29, 2022, from https://www.who.int/health-topics/coronavirus

Dal Molin Ribeiro, M., Gomes da Silva, R., Fraccanabbia, N., Mariani, V., & Coelho, L. (2019). *Forecasting Epidemiological Time Series Based on Decomposition and Optimization Approaches*. https://doi.org/10.21528/CBIC2019-18

Davis, J., Jason Zhu, J., & Oldfather, J. (2020). AWS Prescriptive Guidance—Quantifying uncertainty in deep learning systems. *Https://Docs.Aws.Amazon.Com/Prescriptive-Guidance/Latest/Ml-Quantifying-Uncertainty/Mc-Dropout.Html*, 25.

Gal, Y., & Ghahramani, Z. (2016). *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning* (arXiv:1506.02142). arXiv. https://doi.org/10.48550/arXiv.1506.02142

Gautam, Y. (2022). Transfer Learning for COVID-19 cases and deaths forecast using LSTM network. *ISA Transactions*, *124*, 41–56. https://doi.org/10.1016/j.isatra.2020.12.057

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., … Zhong, N. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England Journal of Medicine*, NEJMoa2002032. https://doi.org/10.1056/NEJMoa2002032

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hussain, A., Bhowmik, B., & do Vale Moreira, N. C. (2020). COVID-19 and diabetes: Knowledge in progress. *Diabetes Research and Clinical Practice*, *162*, 108142. https://doi.org/10.1016/j.diabres.2020.108142

Ibrahim, M. A., & Al-Najafi, A. (2020). Modeling, Control, and Prediction of the Spread of COVID-19 Using Compartmental, Logistic, and Gauss Models: A Case Study in Iraq and Egypt. *Processes*, *8*(11), 1400. https://doi.org/10.3390/pr8111400

Istaiteh, O., Owais, T., Al-Madi, N., & Abu-Soud, S. (2020). Machine Learning Approaches for COVID-19 Forecasting. *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, 50–57. https://doi.org/10.1109/IDSTA50958.2020.9264101

Kırbaş, İ., Sözen, A., Tuncer, A. D., & Kazancıoğlu, F. Ş. (2020). Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos, Solitons & Fractals*, *138*, 110015. https://doi.org/10.1016/j.chaos.2020.110015

Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., & Hsueh, P.-R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, *55*(3), 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924

*Machine learning approach for confirmation of COVID-19 cases: Positive, negative, death and release*. (n.d.). Periodikos. Retrieved May 29, 2022, from http://www.iberoamericanjm.periodikos.com.br/journal/iberoamericanjm/article/doi/10.5281/zenodo.3822623

Mohammed, D. A., Tawfeeq, H. M., Ali, K. M., & Rostam, H. M. (2021). Analysis and Prediction of COVID-19 Outbreak by a Numerical Modelling. *Iraqi Journal of Science*, 1452–1459. https://doi.org/10.24996/ijs.2021.62.5.8

Moujaess, E., Kourie, H. R., & Ghosn, M. (2020). Cancer patients and research during COVID-19 pandemic: A systematic review of current evidence. *Critical Reviews in Oncology/Hematology*, *150*, 102972. https://doi.org/10.1016/j.critrevonc.2020.102972

Mullachery, V., Khera, A., & Husain, A. (2018). *Bayesian Neural Networks*.

Mustafa, H. I., & Fareed, N. Y. (2020). COVID-19 Cases in Iraq; Forecasting Incidents Using Box—Jenkins ARIMA Model. *2020 2nd Al-Noor International Conference for Science and Technology (NICST)*, 22–26. https://doi.org/10.1109/NICST50904.2020.9280304

Nadler, P., Arcucci, R., & Guo, Y. (2021). *A Neural SIR Model for Global Forecasting*. 13.

Ndaïrou, F., Area, I., Nieto, J. J., & Torres, D. F. M. (2020). Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons, and Fractals*, *135*, 109846. https://doi.org/10.1016/j.chaos.2020.109846

Rao, K., PATRA, G., Mopuri, R., & Mutheneni, S. R. (2020). *A deep learning approach for prediction of SARS-CoV-2 cases using the weather factors in India*. https://doi.org/10.22541/au.160275979.91541585/v1

Rauf, H. T., Lali, M. I. U., Khan, M. A., Kadry, S., Alolaiyan, H., Razaq, A., & Irfan, R. (2021). Time series forecasting of COVID-19 transmission in Asia Pacific countries using deep neural networks. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-020-01494-0

Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & Coelho, L. dos S. (2020). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons, and Fractals*, *135*, 109853. https://doi.org/10.1016/j.chaos.2020.109853

Shetty, R. P., & Pai, P. S. (2021). Forecasting of COVID 19 Cases in Karnataka State using Artificial Neural Network (ANN). *Journal of The Institution of Engineers (India): Series B*, *102*(6), 1201–1211. https://doi.org/10.1007/s40031-021-00623-4

Singh, S., Parmar, K. S., Kumar, J., & Makkhan, S. J. S. (2020). Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. *Chaos, Solitons, and Fractals*, *135*, 109866. https://doi.org/10.1016/j.chaos.2020.109866

Su, H., Yang, M., Wan, C., Yi, L.-X., Tang, F., Zhu, H.-Y., Yi, F., Yang, H.-C., Fogo, A. B., Nie, X., & Zhang, C. (2020). Renal histopathological analysis of 26 postmortem findings of patients with COVID-19 in China. *Kidney International*, *98*(1), 219–227. https://doi.org/10.1016/j.kint.2020.04.003

Tomar, A., & Gupta, N. (2020). Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Science of The Total Environment*, *728*, 138762. https://doi.org/10.1016/j.scitotenv.2020.138762

Yahya, B. M., Yahya, F. S., & Thannoun, R. G. (2021). COVID-19 prediction analysis using artificial intelligence procedures and GIS spatial analyst: A case study for Iraq. *Applied Geomatics*, *13*(3), 481–491. https://doi.org/10.1007/s12518-021-00365-4