

A Deep Learning LSTM Approach to Predict COVID-19 Deaths in North Africa

Sami Khedhiri, PhD¹ 

Introduction

The countries of Tunisia, Algeria, and Morocco are part of the North Africa region, also called the Maghreb. As of October 16, 2022, there are 52413 COVID-19 related deaths reported for these countries despite the significant progress in vaccination.¹ A notable interest by scholars has emerged recently to model and to forecast the spread and the lethality of the pandemic and its impact in the Middle East and North Africa region. In a recent study, spatial panel-data models were used to identify the factors for the spike of COVID-19 infections in North Africa.² In another study, a statistical analysis was performed based on zero-inflation models and autoregressive conditional count models to forecast death counts with evidence from Tunisian data.³ Furthermore, quantitative analyses including statistical modeling and deep learning methods have also been performed to forecast the pandemic outbreak in different parts of the world.⁴ For instance, some authors presented long short-term memory (LSTM) based models to predict novel infections of the coronavirus in India, whereas in other studies deep learning methods were used to forecast new COVID-19 cases and death rates in Australia and Iran.⁵ Using COVID-19 datasets of several countries including Brazil, Germany, Italy, Spain, United Kingdom, China, India, Israel, Russia, and United States, alternative deep learning methods were studied and their results were compared in terms of forecast performance.⁶ In this study, we contribute to this ongoing literature, by conducting a statistical analysis with publicly available data on the coronavirus death counts for the Maghreb countries, and we show that the method of deep learning with LSTM network outperforms time series autoregressive integrated moving average (ARIMA) models in terms of forecast accuracy for the pandemic deaths.

Methods

There are several deep learning methods in the literature, including LSTM neural network, Gated Recurrent Unit, Convolutional Neural Network, Deep Neural Network, Extreme Learning Machine and Multilayer Perceptron.⁷ The main components of the LSTM are its gates which are given by the input gate, the forget gate, and the output gate. The

input gate will control the inflow of new information into the cell. The forget gate will control the content of the memory, that is, the forget gate will decide if we want to forget a piece of information so we can store new information. The output gate will control when the information is used in the output from the cell.

In this study, we limit our empirical analysis to only one state-of-the-art method which is based on LSTM networks to predict COVID-19 deaths in the Maghreb. In general, an LSTM network solves the gradient vanishing problem which characterizes recurrent neural networks. It does so by modeling the long-term dependencies of a time series with an optimal lag length, and by allowing memory unit of the process to decide, remember, and forget information, accordingly, which can create connections between present and past data observations, and can compute mapping between input and output sequences.

To apply the LSTM methodology, each country data are divided into two subsets, one for in-sample training and it includes 90% of the data, and the other 10% is for out of sample prediction. The adaptive moment estimation (ADAM) optimization algorithm is applied to the training data which is standardized to have zero mean and unit variance. The algorithm computes an exponential moving average of the gradient and its square with specified parameter values to control the decay rates. ADAM optimizer is popular in deep learning applications, and with the new updates to the learning rate, scholars addressed the shortcomings of its original algorithm and made it a more reliable optimizer.⁸ In this paper, we compute the prediction values of deaths using the *predictAndUpdateState* MATLAB function. The final step is to compute the root mean square error of the forecasts and to plot the observed and the predicted values of COVID-19-related deaths for each North African country.

Next, we fit the data to time series ARIMA models with R software and we calculate the root mean square error (RMSE)

Asia Pacific Journal of Public Health

2023, Vol. 35(1) 53–55

© 2022 APJPH



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/10105395221141590

journals.sagepub.com/home/aph



¹University of Prince Edward Island, Charlottetown, PE, Canada

Corresponding Author:

Sami Khedhiri, University of Prince Edward Island, 550 University Avenue, Charlottetown, PE C1A 4P3, Canada.

Email: skh.ypei@gmail.com

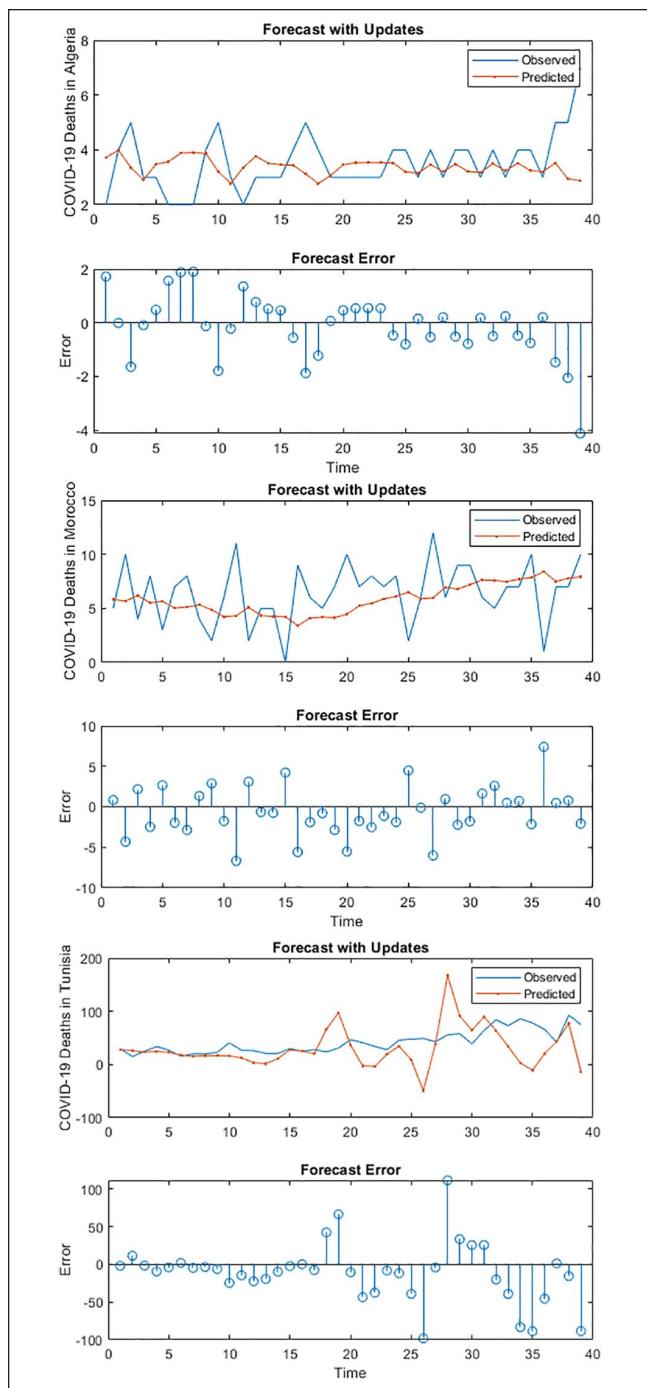


Figure 1. Observed and predicted COVID-19 deaths in North Africa.

to compare the forecast accuracy of these models with deep learning LSTM models. An autoregressive integrated moving average model can be represented as follows:

$$\varphi(B)(1-B)^d(y_t - \mu) = \theta(B)\varepsilon_t \quad (1)$$

Where $\psi(B)$ and $\theta(B)$ are polynomial functions in the backshift operator B , y_t is the time series, d is the order of integration, and ε_t is a white noise process.

Results

We collected publicly available data on death counts related to the pandemic for each of the three North African countries. The data sets which have no missing values, cover the period from March 24, 2020, to April 21, 2021, and they can be obtained online.⁹ The computations for the LSTM model predictions are performed with MATLAB programs, and for ARIMA models we applied R software coding. *Auto.arima* function in R shows that the best fit for the pandemic death data is given by an ARIMA model of order (0,1,1) for Algeria, (1,1,2) for Morocco, and (2,2,3) for Tunisia. We run these models to compute postsample predictions and root mean square error for each country data.

Figure 1 displays the observed and updated forecast values of LSTM models for each of the Maghreb countries, based on the last 10% of data, which is used for prediction, whereas the first subsample of 90% of the data is used for training. There are 394 daily observations in total for each COVID-19 death data and therefore the figure lists the last 39 observed and predicted values and the forecast errors obtained from deep learning LSTM method, from March 14, 2021, to April 21, 2021.

Table 1 lists RMSE for both ARIMA and LSTM models and shows clear evidence of better forecast accuracy of the deep learning method compared to times series ARIMA models. The root mean square errors for each of Algeria, Morocco, and Tunisia COVID-19 death data are much smaller with long short-term memory networks than with the autoregressive integrated moving average time series methods.

Discussion and Conclusion

As we enter the third year of the pandemic, COVID-19 deaths have so far exceeded 6 million worldwide, and people are still struggling to return to normalcy. Like other parts of the world, the North African countries have been negatively impacted socially and economically by the pandemic outbreak and active research has emerged to model and to predict accurately the outcomes of COVID-19 in the region. As a contribution to the ongoing literature on reliable statistical modeling of the pandemic outbreak, we present an empirical study based on deep learning LSTM methods to forecast the pandemic lethality in North Africa and to compare the forecast accuracy of these models with time series ARIMA models. Our study finds that methods based on deep learning networks provide more accurate forecasts than time series

Table 1. RMSE for LSTM and ARIMA Model Forecasts of COVID-19 Deaths.

Country	COVID-19 death count (as of October 16, 2022)	RMSE	
		Deep learning LSTM network	ARIMA model
Algeria	6881	1.185	2.698
Morocco	16278	3.056	6.054
Tunisia	29254	25.440	31.212

Abbreviations: ARIMA, autoregressive integrated moving average; LSTM, long short-term memory; RMSE, root mean square error.

autoregressive integrated moving average models, with lower root mean square forecast errors. It is very important for health official and healthcare professionals to have access to accurate forecasts of COVID-19 lethality to implement measured and effective health policies.

The findings of the paper show that deep learning networks have more accurate predictions of deaths related to the pandemic than time series models. In a related study based on COVID-19 data from the Gulf countries, it was found that state space models outperform LSTM networks in terms of forecast accuracy in presence of highly complex surveillance data.¹⁰

One limitation of this study is that it does not include a comprehensive complexity analysis in order to verify whether the superiority of deep learning LSTM models over ARIMA models in terms of more accurate forecasts with lower root mean square errors may be explained by the notion of data complexity. Also, this paper used conventional recurrent neural networks which are only capable of training a single model, whereas a bidirectional LSTM network allows the information to be processed from a sequence of input data and from the reverse of that sequence, and therefore better predictions may be expected from bidirectional LSTM than from the conventional LSTM model. This could be a direction for future empirical research on the topic.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Approval

This research uses publicly available data and thus ethical approval is not applicable.

ORCID iD

Sami Khedhiri  <https://orcid.org/0000-0001-7496-3182>

Data Availability Statement

The data used in this paper is publicly available online.

References

1. Worldometers. COVID-19 Coronavirus pandemic. Published 2022. Accessed October 16, 2022. www.worldometers.info/coronavirus/
2. Khedhiri S. A spatiotemporal analysis of the COVID-19 pandemic in North Africa. *Geohealth*. 2022;6(7):e2022GH000630. doi:10.1029/2022GH000630
3. Khedhiri S. Statistical modeling of COVID-19 deaths with excess zero counts. *Epidemiol Methods*. Published online October 8, 2021. doi:10.1515/em-2021-0007
4. Chandra R, Jain A, Singh Chauhan D. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLoS ONE*. 2022;17(1):e0262708. doi:10.1371/journal.pone.0262708
5. Ayoobi N, Sharifrazi D, Alizadehsani R, et al. Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. *Results Phys*. 2021;27:104495. doi:10.1016/j.rinp.2021.104495
6. 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. doi:10.1016/j.chaos.2020.110212
7. Oshinubi K, Amakor A, James Peter O, Rachid N, Demongeot J. Approach to COVID-19 time series data using deep learning and spectral methods. *Bioengineering*. 2022;9(1):1-21. doi:10.3934/bioeng.2022001
8. Wang Y, Xiao Z, Cao G. A convolutional neural network method based on Adam optimizer with power-exponential learning rate for bearing fault diagnosis. *J Vibroengineering*. 2022;24(2):666-678. doi:10.21595/jve.2022.22271
9. Statista. Coronavirus (COVID-19) deaths worldwide per one million population as of July 13, 2022, by country. Published 2022. Accessed October 16, 2022. www.statista.com/statistics/1104709/coronavirus-deaths-worldwide-per-million-inhabitants/
10. Khedhiri S. Forecasting COVID-19 infections in the Arabian Gulf region. *Model Earth Syst Environ*. 2022;8(3):3813-3822. doi:10.1007/s40808-021-01332-z