Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/180113

Comparative analysis of time series prediction models for visceral leishmaniasis:based on SARIMA and LSTM

Fathelrhman EL Guma[1,2]

[1]Department of Mathematics, Faculty of Science and Arts, Al-Baha University, Baljurashi 1988, Saudi Arabia
[2]Department of Statistical Study, College of Economics and Community Development, Peace University, West Kordofan, Sudan

Received: 9 Oct. 2023, Revised: 26 Nov. 2023, Accepted: 9 Dec. 2023
Published online: 1 Jan. 2024

Abstract: Visceral leishmaniasis, a severe health disorder, is attributed to the microscopic parasite Leishmania. The parasitic illness possesses the capacity to pose a significant risk to human life and exhibits a variable prevalence across people worldwide. Using time series prediction techniques for VL might offer valuable insights to aid public health professionals in strategizing and implementing effective measures for VL prevention. This study presents a comparative analysis of time series forecasting techniques, specifically focusing on two methods: SARIMA and LSTM recurrent neural networks. Forecast performance evaluation involves utilizing monthly VL data acquired from district health offices from January 2000 to December 2021. An assessment of the model's performance is conducted to ascertain its efficacy. According to the evaluation conducted using three metrics, namely mean average precision (MAP), root mean square (RMS), and mean absolute error (MA), the findings indicate that the LSTM model outperforms the SARIMA model in terms of forecasting monthly conditions. The discovery implies that the LSTM approach may be better suited for predicting VL incidents and has the potential to contribute to the formulation of efficient preventive measures. Furthermore, it is suggested that future studies should investigate the possibility of integrating SARIMA and LSTM techniques to improve VL forecasts' precision.

Keywords: Prediction; LSTM; SARIMA; VL pandemic

1 Introduction

Visceral Leishmaniasis (VL), commonly called Kala-Azar, is a grave health condition caused by a microscopic organism called Leishmania. This parasitic infection can potentially be life-threatening and affects many individuals globally, ranging from 2 to 4000 cases. The transmission of this disease occurs through the bite of the phlebotomine female sandfly, a tiny insect approximately one-third the size of a mosquito. The transmission of the parasite occurs through biting, which enables promastigotes to enter macrophages and other mononuclear phagocytic cells beneath the skin. This process plays a crucial role in the parasite's lifecycle and its ability to establish infection within the host. Leishmania, a parasitic infection, can have detrimental effects on the cells of its host, particularly impacting vital organs such as the liver, spleen, bone marrow, and lymph nodes. Common symptoms of this condition may include persistent and irregular fever, cough, decreased appetite, diarrhoea, vomiting, jaundice, swelling, nosebleeds, and anaemia. Individuals who test positive for VL (Visceral Leishmaniasis) encounter a range of symptoms within 3 to 6 months after being infected. PKDL, or post-kala-azar dermal leishmaniasis, is a skin condition characterized by maculopapular or nodular rashes. It is essential to be aware that this condition can potentially contribute to the spread of the underlying disease [1,2,3,4]. Sudan is among the six nations endemic to visceral VL, accounting for 90% of global VL cases. Sudan has the highest reported prevalence of post-kala-azar dermal leishmaniasis (PKDL) worldwide [3]. Zijlstra et al. [1] reported the findings of population-based investigations, which revealed an annual incidence rate of 38 cases per 1000 individuals and a case fatality rate of up to 20.5%. The nation has encountered several epidemics throughout the preceding two decades. From the western bank of the White Nile to the shared border between Sudan and Ethiopia, the illness is endemic throughout a sizable geographic area. The geographical scope of this region encompasses the southern territories of the Central and Eastern provinces. Instances have been documented in Darfur, a region located in western Sudan, and in the

* Corresponding author e-mail: fathyelrhman@gmail.com

Nuba Mountains area in Kordofan [4,5,6,7,8,9,10,11]. Related Work Epidemiological research has extensively used ARIMA and LSTM models to predict illness incidence, analyze trends, and understand disease dynamics. The following are instances of previous papers that have used these models: Y et al. [12] employed LSTM and ARIMA models to predict COVID-19 data in Germany. They evaluated the effectiveness of these models against conventional prediction methods, such as multiple linear regression models. Zhao D et al. [13] employed ARIMA, GM (1,1), and LSTM models to forecast mainland China TB cases and compare their performance. The LSTM model was determined to be the most effective model for predicting TB cases. Shaman and Karspeck [14] employed ARIMA models to predict the frequencies of influenza-like illness (ILI). The ARIMA model can forecast future outbreaks by evaluating previous ILI rates, assisting in prompt public health interventions. The study by Muhammad Riaz et al. [15] focuses on predicting the number of malaria cases in Pakistan using statistical models. From 2011 to 2022, data was gathered from the Ministry of Health in Rahim Yar Khan. The Holt-Winter multiplicative model demonstrated superior performance compared to the ARIMA and SARIMA models, exhibiting the lowest values for RMSE, MAPE, and MAE. The forecasted data indicated that there would be a minimum of 586.75 cases in June 2022 and a maximum of 1281.93 cases in October 2022. The findings indicate that the government of Pakistan should strengthen its immunization efforts to decrease the incidence of malaria. Chimmula and Zhang [16] presented a deep learning approach using long short-term memory (LSTM) networks to predict the COVID-19 outbreak in Canada and globally. The study found that the possible ending point for the outbreak was around June 2020. The researchers also compared Canada's transmission rates with those of Italy and the USA and presented predictions for successive days until March 31, 2020. Watad, A. et al.[17] utilized LSTM to forecast vector-borne illnesses, specifically the West Nile virus. These models facilitate comprehension of the dissemination patterns and enable the implementation of preventive measures. Seasonal considerations play a significant role in influencing the occurrence of VL. The SARIMA model is a widely used methodology in time series analysis to identify trends and make forecasts based on seasonal data. This approach demonstrates superior precision in forecasting seasonal time series data. The SARIMA model has been effectively utilized in several domains over the last thirty years; nevertheless, it includes specific limitations. The applicability of the SARIMA model is limited to linear time series data models, as it cannot effectively handle nonlinear patterns. A novel methodology for deep learning algorithms has been proposed, specifically designed to address the challenges posed by non-linearity and complexity in the context of time series prediction tasks. LSTM is a deep learning technique that enables the processing of significantly longer temporal sequences. Several empirical studies have demonstrated that LSTM models perform better than conventional algorithms like the SARIMA model. The present study aims to discuss the contribution of the research findings in the following manner:

1. The study enhances our understanding of the mechanisms underlying the transmission of visceral leishmaniasis. The importance of this issue is especially apparent in regions such as Gadaref State, Sudan, where the disease is highly prevalent.
2. This study compares the SARIMA and LSTM models' ability to predict visceral leishmaniasis data. The purpose of this investigation is to gain an understanding of the relative effectiveness of both approaches in accurately capturing inherent patterns and trends within the dataset.
3. This investigation employs suitable error measurements, namely MAP (mean absolute percentage error), RMS (root mean square error), and MA (mean absolute error), to assess the effectiveness of the models. This approach guarantees a thorough evaluation of their precision and dependability, enabling a full comparison.
4. This study can potentially improve the methodology of time-series analysis in epidemiology. Specifically, it aims to investigate the application of LSTM, a neural network technique, for predicting illness patterns. This approach could potentially have wider implications for predicting trends in various diseases. By demonstrating the effectiveness of data-driven methodologies in accurately predicting the spread of diseases, this research significantly contributes to the broader field of healthcare analytics.

The current article is structured as follows: Section 2 describes the data utilized and outlines the suggested materials and methods. Section 3 provides a comprehensive overview of the study's data and the recommended approach. In conclusion, Section 4 of this study offers a comprehensive summary of the research findings and an analysis of the limitations inherent in the presented models. Additionally, this section outlines potential avenues for future research and exploration.

2 Materials and Methods

2.1 VL data

According to data provided by the officials of the Health Information Center, Ministry of Health, Gadaref State, Sudan, we have obtained the monthly number of visceral leishmaniasis cases recorded for the State of El Gadaref, Eastern Sudan, spanning from January 2020 to December 2021. The comprehensive dataset utilized in this research comprised a total of 25 months.

2.2 Auto regressive integrated moving average (ARIMA)

The ARIMA model, a well-established technique for time-series prediction, is a comprehensive framework that integrates the autoregressive (AR) and moving average (MA) processes [18,19,20,21,22]. The ARIMA (p,d,q) model is suitable for predicting future values of stationary time series data. The parameters of the model, denoted as (p,d,q), may be precisely characterized as follows: the auto-regression (AR) term (p) and the moving average (MA) term (q). ARIMA models are constructed by integrating two components of the models through the utilization of differencing terms $(dorI)$. The autoregressive (AR) term refers to regressing a particular variable against its past values to predict the variable of interest. The MA term in forecasting utilizes the error terms from a previous time step to predict the value of a variable at a subsequent time step. The equation denoted as Equation (1) provides a generalization of the pth-order autoregressive (AR) model and the qth-order moving average (MA) model.

$$\begin{aligned}\psi_t = \delta + \phi_1\psi_{t-1} + \phi_2\psi_{t-2} + \cdots + \phi_p\psi_{t-p},\\ + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}.\end{aligned} \tag{1}$$

Here ψ_t and ε_t refer to the actual value of a time series and the random error (white noise) at time t, respectively. Moreover, δ, ψ_i and θ_j are a constant and the coefficients of the AR and MA parts, respectively. In this context, δ is intercept, ψ_t represents the true value of a time series at a given time t, whereas ε_t represents the random error, often known as white noise, at the same time t. Furthermore, ϕ_i and θ_i are fixed values representing the coefficients of the autoregressive (AR) and moving average (MA) components, respectively.

Seasonal ARIMA Model The SARIMA $[(p,d,q)]$ model is an expanded iteration of the ARIMA model, specifically designed to accommodate time series data that exhibit seasonal patterns. The SARIMA model incorporates three new superparameters, namely P,D, and Q, and an additional seasonal periodic parameter s, building upon the ARIMA model. The components of SARIMA, denoted as P,D,Q, and s, correspond to seasonal autoregression, seasonal integration, seasonal moving average, and seasonal period length, respectively, as given by Equation (2) [22,23].

$$\phi_P(\beta^m)\phi_P(\beta)(1-\beta^m)^s(1-\beta)^d\psi_t = \theta_Q(\beta^m)\theta_q(\beta)\varepsilon_t. \tag{2}$$

The non-seasonal autoregressive polynomial, denoted as ϕ_P, represents the non-stationary time series and the Gaussian white noise process. On the other hand, the non-seasonal Moving average polynomial, denoted as θ_q,, is employed for the same purpose. The seasonal difference term, S,, is discussed in reference [24].

2.3 LSTM Networks for Modeling Time Series

The LSTM is a variant of the recurrent neural network (RNN) that can learn and capture relationships of varying durations, including both short-term and long-term contexts. The technique is employed to address the issue of the bursting or disappearing gradient phenomenon encountered in deep recurrent neural networks. LSTM was first introduced by S. Hochreiter and J. Schmidhuber [25]. The Long Short-Term Memory (LSTM) architecture incorporates a distinctive memory cell that enables it to selectively store or discard information over a duration, rendering it highly proficient in activities that need the acquisition and preservation of crucial patterns or sequences. The LSTM's efficacy in managing extended dependencies has rendered it a widely favoured option across many domains, including but not limited to speech recognition, natural language processing, and time series analysis.

The architectural structure of a Long Short-Term Memory (LSTM) model consists of three distinct gates, namely the forget gate, input gate, and output gate, as described in reference [26].

The forget gate denoted as ω, determines the exact information to be removed or erased from the memory cells, also known as the cell state. Gates commonly employ a sigmoid activation function as their activation mechanism, yielding an output bounded within the range of 0 to 1. If the output is equal to 1, it may be inferred that all information will be preserved without any loss or alteration. If the value is zero, all the data will be disregarded. The above information is provided by Equation (3).

$$Y_t = \rho(\pi_\lambda P_{t-1} + \pi_\lambda X_t). \tag{3}$$

In the given context, the symbol "π_λ" represents the weight of the forget gate. The symbol "P_{t-1}" represents the previous state or the state at time t. The symbol "X_t" signifies the input at time $t-1$. Lastly, the symbol "s" represents the sigmoid activation function.

The input gate V_t identifies the information incorporated into the cell state (S_t). This procedure is divided into two separate steps. During the initial stage of the procedure, the calculation of the candidate value Y, which can be subsequently included in the cell states, is performed. In the procedure's second stage, the activation values V_t of the input and gates are calculated. The input gate weight is denoted as π_v , while the cell state weight is denoted as π_s. The two procedures are designated as

follows:

$$V_t = \sigma(\pi_v P_{t-1} + \pi_v X_t),$$
$$\phi = \tanh(\pi_s P_{t-1} + \pi_s X_t).$$

The updated cell states (si) are computed by leveraging the outputs of the preceding computational steps. The formula can be decomposed into its constituent components in the following manner:

$$S_t = G_t S_{t-1} + V_t * \phi,$$
$$m_t = \sigma(\pi_t P_{t-1} + \pi_m X_t),$$
$$n_t = m_t * \tanh(S_t).$$

The memory cell undergoes input gate and forget gate operations before generating an output gate, represented as m_t. This gate determines specific regions within the memory cell for output, ultimately generating the outcome. The computational process involves element-wise multiplication of the current memory cell state with a sigmoid activation function. The output gates have two implemented gates, enhancing system functionality and efficiency. The initial gating mechanism uses a sigmoid activation layer to determine specific portions of the cell state for transmission as output. The hyperbolic tangent activation function and multiplication operation regulate information flow within the memory cell, enhancing the model's capacity to capture long-term dependencies and generate precise predictions.

2.4 Strengths and weaknesses of LSTM and SARIMA models:

LSTM models are appropriate for complicated forecasting jobs because they can capture long-term dependencies and non-linear interactions in time-series data. They can be trained end-to-end without the requirement for feature engineering, and they can manage noisy inputs and missing data. One of the limitations of LSTM models is their high computational cost and the need for a substantial quantity of training data to achieve optimal performance. Overfitting may occur when the model is excessively intricate, or the training data is inadequate [27,28,29]. The implementation of the SARIMA model offers notable strengths and weaknesses. SARIMA models excel in their ability to effectively capture and analyze seasonal patterns and trends within time-series data. This makes them particularly well-suited for forecasting tasks that exhibit distinct and recurring seasonal patterns. Furthermore, these models are characterized by their relative simplicity and ease of implementation, making them highly accessible for practical applications. Additionally, they exhibit the remarkable capability to be trained effectively even with limited data. SARIMA models, short for Seasonal Autoregressive Integrated Moving Average models, operate assuming that the underlying time-series data is stationary. However, it is important to note that this assumption may not hold for all types of time-series data. Machine learning models can exhibit sensitivity to outliers and noise present in the data, thereby potentially compromising their accuracy [31,32,33].

3 Results and Discussion

This section presents the experimental data that compare the predicted outcomes of SARIMA and LSTM.

3.1 SARIMA Forecasting

The training data for the SARIMA model consists of 90% of the visceral leishmaniasis dataset, while the remaining 10% is allocated for testing the model's predictions. The SARIMA model was developed using a training data set. The plot of the training data set is illustrated in Figure 1, showing the time series pattern over a specific period. The SARIMA model considers the seasonal and non-seasonal components of the data, allowing for accurate forecasting and prediction. In comparison, Figure 2 represents the seasonal decomposition of the training data set using the SARIMA model. The decomposition shows clear patterns of seasonality, trend, and residual components. These components will inform the forecasting process and improve the accuracy of future predictions. The autocorrelation function (ACF) and partial autocorrelation function (PACF) showed clear seasonal patterns. Figure 3 shows that the series is not stationary. Figure 4 represents ACF and PACF for a first-order difference and a first seasonal difference time series. The Augmented Dickey-Fuller (ADF) test yielded a $t = (-2.65)$, indicating a statistically significant result with a P value of 0.005. Hence, it may be concluded that the initial sequence had characteristics of both non-stationary and non-randomness. The original series achieved stationary using a first-order difference and a first-seasonal difference. Because the Augmented Dickey-Fuller (ADF) test yielded a $t = (-4.88)$, indicating a statistically significant result with a P = 0.00005. Ultimately, the SARIMA $(p,1,q)$ $(P,1,Q)$ model was tentatively chosen. An exhaustive search was conducted using a Python programming grid search method to identify the optimal model based on the best Akaike Information Criterion (AIC) value criteria. After careful analysis, it has been determined that the SARIMA $(0,1,1)(2,1,2)$ model exhibits superior performance, as evidenced by its remarkably low AIC value of 2748.17. The SARIMA $(0,1,1)(2,1,2)$ model was the most suitable for forecasting VL cases. Figure 5 illustrates a comparative fit. The SARIMA model's predicted data and the test data set are shown in conjunction, illustrating a strong resemblance between the two. This suggests that the

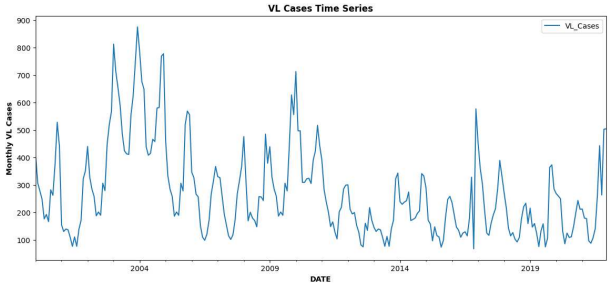

Fig. 1: Plot of the monthly incidence of visceral leishmaniasis in Gadaref State, Sudan, from 2000 to 2022.

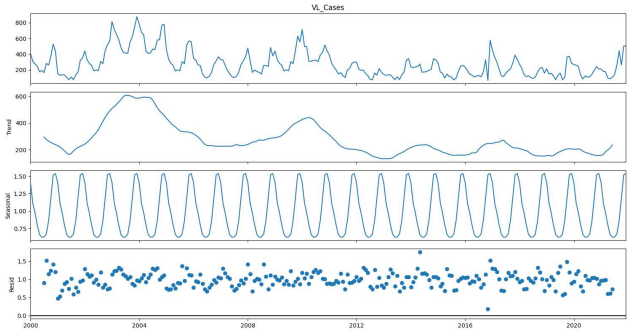

Fig. 2: Seasonal decomposition of visceral leishmaniasis in Gadaref State, Sudan, from 2000 to 2021.

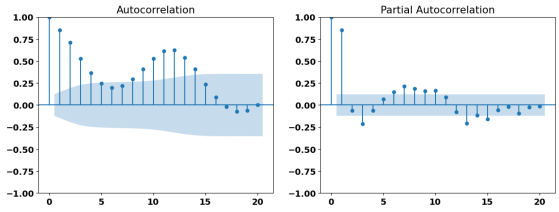

Fig. 3: Autocorrelation and partial autocorrelation plots: The original time series.

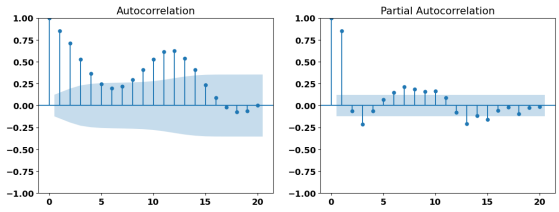

Fig. 4: Autocorrelation and partial autocorrelation plots: A first-order difference and a first seasonal difference time series

SARIMA model can effectively capture VL case patterns and variations.

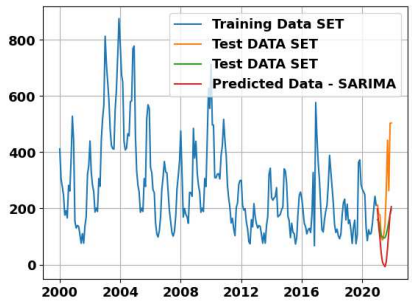

Fig. 5: Comparing predicted SARIMA and test data.

3.2 LSTM Result

The investigation utilizes an artificial recurrent neural network (RNN) model to forecast instances of visceral leishmaniasis. This is accomplished by leveraging a dataset consisting of 240 monthly cases. The dataset has been subjected to compression and partitioning, resulting in distinct subsets for training and testing. Specifically, the training dataset encompasses 90% of the overall instances of visceral leishmaniasis cases. The input data undergoes a conversion process to be represented as a NumPy array, which is then utilized as input for the LSTM model. The Adam optimizer minimizes the mean squared error, augmenting the model's efficacy. The dynamic algorithm uses an adaptive learning rate mechanism to expedite convergence and increase precision. The Mean Squared Error (MSE) is a statistical metric that quantifies the average discrepancy between predicted and observed values. It is particularly well-suited for tasks involving regression analysis. The training involves utilizing a batch size of 1 and a single epoch value while the model is evaluated on a scaled testing dataset. The model exhibits a mean squared error (MSE) of 0.05, suggesting a moderate average deviation between the predicted and observed values. This indicates the model's robustness in effectively forecasting the dependent variable within regression analyses. The model is subsequently evaluated using the scaled testing dataset, yielding the ensuing outcomes: Figure (9) depicts the testing dataset for visceral leishmaniasis, showcasing the predictions generated by the LSTM model. The visual representation illustrates that the predicted values exhibit higher proximity to the training data. The observed results suggest that the LSTM model has effectively captured the underlying patterns and trends within the testing dataset, resulting in precise predictions for visceral leishmaniasis. The strong correlation between the predicted values and the actual data implies the robustness and efficacy of the model in prognosticating this ailment.

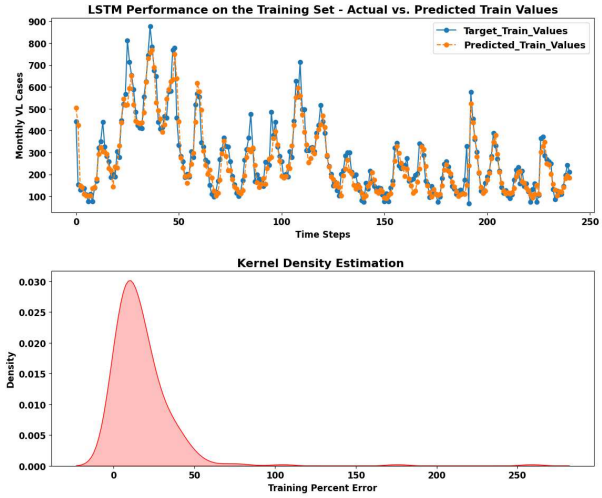

Fig. 6: A comparative analysis between the training dataset and the predicted values generated by the LSTM model.

3.3 Model comparison

The metrics often employed for evaluating and comparing forecasting techniques on a singular time series are MAP (mean absolute percentage error), RMS (root mean square error), and MA (mean absolute error). The utilization of MAP as the key performance statistic in this study is justified by its interpretability. The mean absolute percentage mistake (MAP) is a metric that quantifies the absolute mistake as a percentage of the true value. This measure facilitates the convenient comparison of diverse time series characterized by differing scales and units. Furthermore, the mean absolute percentage error (MAP) exhibits a higher degree of resistance to the influence of outliers when compared to both the root mean square error (RMS) and the mean absolute error (MAP). This characteristic renders MAPE a reliable option for assessing the accuracy of predictions. Equation 4 provides the formal mathematical expressions for the Mean Absolute Percentage Error (MAP), Root Mean Square Error (RMS), and Mean Absolute Error (MAP). In this equation, Z_i represents the observed values, $\hat{Z}_i$ represents the projected values, and T is the total number of data points used for testing.

$$\begin{aligned} \text{MAP} &= \frac{1}{T}\sum_{i=1}^{T}\frac{|Z_i-\hat{Z}_i|}{Z_i}\times 100\%, \\ \text{RAM} &= \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Z_i-\hat{Z}_i)^2}, \\ \text{AM} &= \frac{1}{T}\sum_{i=1}^{T}|Z_i-\hat{Z}_i|. \end{aligned} \tag{4}$$

Table 1: A comparative analysis of the performance between two models.

Methods	MPA	RMS	MA
SARIMA	66.69	198.94	160.5
LSTM	40.10	183.685	127.5

As shown in Table 1, the MAP, RMS, and MA of the test data set for the LSTM model were 40.10, 183.685, and 127.5, respectively, which are significantly lower than those of the SARIMA model. These results indicate that the LSTM model outperforms the SARIMA model regarding accuracy, precision, and mean error. The lower values of MAP, RMS, and MA suggest that the LSTM model provides more accurate predictions for the test data set than the SARIMA model.

3.4 Discussion

This work provides a comparative analysis of SARIMA and LSTM models to forecast the time series of VL states. Forecast accuracy assessment entails using monthly VL data counts, covering the period from January 2000 to December 2021. The SARIMA model (0,1,1)(2,1,2) was identified as the optimal model for predicting the monthly number of VL cases. The ensuing study utilized the LSTM model with a single hidden layer layout. Based on empirical data, both SRIMA and LSTM have positively predicted future outcomes. The LSTM model demonstrates superior prediction accuracy compared to the SRIMA model. In light of the unpredictability and dynamic character of epidemics, it is critical to acknowledge the need for future methods in the conclusion of this study comparing ARIMA and LSTM models for epidemic forecasting. The following are a few instances of potential future research areas to improve the efficiency of epidemic forecasting models: The accuracy of the models might be improved by adding more epidemiological variables, such as vaccination rates, public health initiatives, and demographic data. By utilizing the advantages of each model, combining many model types (ARIMA, LSTM, and SEIR models) in an ensemble method may result in more reliable predictions, extending the models to examine cross-border dynamics and geographical dissemination, especially for illnesses that are likely to spread via migration and travel.

Acknowledgement

The author would like to thank everyone who helped compile the study's data set from the Health Information Center under the Gedaref State Ministry of Health. He would also like to thank the anonymous reviewers for their invaluable contributions.

References

[1] E. Zijlstra and A. El-Hassan, "Leishmaniasis in sudan. 3. visceral leishmaniasis," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 95, no. Supplement_1, pp. S27–S58, 2001.

[2] J. van Griensven, T. P. Dorlo, E. Diro, C. Costa, and S. Burza, "The status of combination therapy for visceral leishmaniasis: an updated review," *The Lancet Infectious Diseases*, 2023.

[3] F. Chappuis, S. Sundar, A. Hailu, H. Ghalib, S. Rijal, R. W. Peeling, J. Alvar, and M. Boelaert, "Visceral leishmaniasis: what are the needs for diagnosis, treatment and control?," *Nature reviews microbiology*, vol. 5, no. 11, pp. 873–882, 2007.

[4] A. Clem, "A current perspective on leishmaniasis," *Journal of global infectious diseases*, vol. 2, no. 2, p. 124, 2010.

[5] Guma, F., Badawy, O., Musa, A., Mohammed, B., Abdoon, M., Berir, M. & Salih, S. Risk factors for death among COVID-19 Patients admitted to isolation Units in Gedaref state, Eastern Sudan: a retrospective cohort study. *Journal Of Survey In Fisheries Sciences*. **10**, 712-722 (2023).

[6] Elshoubary, E., Abdoon, M., Bahatheg, A. & Albeladi, M. Stability analysis and numerical simulation of fractional model of Leishmaniasis. *Results In Nonlinear Analysis*. **6**, 9-17 (2023).

[7] Abdoon, M., Saadeh, R., Berir, M., Guma, F. & Others Analysis, modeling and simulation of a fractional-order influenza model. *Alexandria Engineering Journal*. **74** pp. 231-240 (2023).

[8] R. G. Wamai, J. Kahn, J. McGloin, and G. Ziaggi, "Visceral leishmaniasis: a global overview," *Journal of Global Health Science*, vol. 2, no. 1, 2020.

[9] Guma, F., Badawy, O., Berir, M. & Abdoon, M. Numerical Analysis of Fractional-Order Dynamic Dengue Disease Epidemic in Sudan. *Journal Of The Nigerian Society Of Physical Sciences*. pp. 1464-1464 (2023).

[10] A. O. A. Abdalla, A. A. Mohammed, M. A. M. Khaier, A. A. Zainaldeen, *et al.*, "Prevalence and characterization of leishmania donovani in human hospital cases and domestic dogs in gadarif state, sudan," 2023.

[11] D. K. Almutairi, M. A. Abdoon, S. Y. M. Salih, S. A. Elsamani, F. E. Guma, and M. Berir, "Modeling and analysis of a fractional visceral leishmaniosis with caputo and caputo–fabrizio derivatives," *Journal of the Nigerian Society of Physical Sciences*, pp. 1453–1453, 2023.

[12] Y.-T. Tsan, D.-Y. Chen, P.-Y. Liu, E. Kristiani, K. L. P. Nguyen, and C.-T. Yang, "The prediction of influenza-like illness and respiratory disease using lstm and arima," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1858, 2022.

[13] Zhao, D., Zhang, H., Cao, Q., Wang, Z., He, S., Zhou, M. & Zhang, R. The research of ARIMA, GM(1,1), and LSTM models for prediction of TB cases in China. *PLOS ONE*. **17** pp. e0262734 (2022,2)

[14] Shaman, J. & Karspeck, A. Forecasting seasonal outbreaks of influenza. *Proceedings Of The National Academy Of Sciences*. **109**, 20425-20430 (2012)

[15] Riaz, M., Hussain Sial, M., Sharif, S., Mehmood, Q. & Others Epidemiological Forecasting Models Using ARIMA, SARIMA, and Holt–Winter Multiplicative Approach for Pakistan. *Journal Of Environmental And Public Health*. **2023** (2023)

[16] Chimmula, V. & Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*. **135** pp. 109864 (2020)

[17] Watad, A., Watad, S., Mahroum, N., Sharif, K., Amital, H., Bragazzi, N., Adawi, M. & Others Forecasting the West Nile virus in the United States: an extensive novel data streams–based time series analysis and structural equation modeling of related digital searching behavior. *JMIR Public Health And Surveillance*. **5**, e9176 (2019)

[18] Dubey, A., Kumar, A., Garcia-Diaz, V., Sharma, A. & Kanhaiya, K. Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies And Assessments*. **47** pp. 101474 (2021)

[19] Aryal, S., Nadarajah, D., Rupasinghe, P., Jayawardena, C. & Kasthurirathna, D. Comparative analysis of deep learning models for multi-step prediction of financial time series. *Journal Of Computer Science*. **16**, 1401-1416 (2020)

[20] Aghelpour, P., Mohammadi, B. & Biazar, S. Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. *Theoretical And Applied Climatology*. **138**, 1471-1480 (2019)

[21] Zhao, Z., Zhai, M., Li, G., Gao, X., Song, W., Wang, X., Ren, H., Cui, Y., Qiao, Y., Ren, J. & Others Study on the prediction effect of a combined model of SARIMA and LSTM based on SSA for influenza in Shanxi Province, China. *BMC Infectious Diseases*. **23**, 71 (2023)

[22] Adeyeye, J. & Nkemnole, E. Predicting Malaria Incident Using Hybrid SARIMA-LSTM Model. *International Journal Of Mathematical Sciences And Optimization: Theory And Applications*. **9** (2023)

[23] Zhao, R., Liu, J., Zhao, Z., Zhai, M., Ren, H., Wang, X., Li, Y., Cui, Y., Qiao, Y., Ren, J. & Others A hybrid model for tuberculosis forecasting based on empirical mode decomposition in China. *BMC Infectious Diseases*. **23**, 665 (2023)

[24] T. Mills, *A very British affair: Six Britons and the development of time series analysis during the 20th century*. Springer, 2012.

[25] K. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of gated recurrent units (gru), long short-term memory (lstm) cells, autoregressive integrated moving average (arima), seasonal autoregressive integrated moving average (sarima) for forecasting covid-19 trends," *Alexandria engineering journal*, vol. 61, no. 10, pp. 7585–7603, 2022.

[26] J. Yang, L. Yang, G. Li, J. Du, L. Ma, T. Zhang, X. Zhang, J. Yang, L. Feng, W. Yang, *et al.*, "Comparative study on influenza time series prediction models in a megacity from 2010 to 2019: Based on sarima and deep learning hybrid prediction model," 2022.

[27] Amshi, A., Prasad, R. & Sharma, B. Forecasting cholera disease using SARIMA and LSTM models with discrete wavelet transform as feature selection. *Journal Of Intelligent & Fuzzy Systems*., 1-13

[28] A. H. Amshi and R. Prasad, "Time series analysis and forecasting of cholera disease using discrete wavelet transform and seasonal autoregressive integrated moving average model," *Scientific African*, vol. 20, p. e01652, 2023.

[29] H. Liu, C. Li, Y. Shao, X. Zhang, Z. Zhai, X. Wang, X. Qi, J. Wang, Y. Hao, Q. Wu, *et al.*, "Forecast of the trend

in incidence of acute hemorrhagic conjunctivitis in china from 2011–2019 using the seasonal autoregressive integrated moving average (sarima) and exponential smoothing (ets) models," *Journal of infection and public health*, vol. 13, no. 2, pp. 287–294, 2020.

[30] A. K. L. Bezerra and É. M. C. Santos, "Prediction the daily number of confirmed cases of covid-19 in sudan with arima and holt winter exponential smoothing," *International Journal of Development Research*, vol. 10, no. 08, pp. 39408–39413, 2020.

[31] L. Mamudu, A. Yahaya, and S. Dan, "Application of seasonal autoregressive integrated moving average (sarima) for flows of river kaduna," *Niger. J. Eng*, vol. 28, no. 2, 2021.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[33] M. Othman, R. Indawati, A. A. Suleiman, M. B. Qomaruddin, and R. Sokkalingam, "Model forecasting development for dengue fever incidence in surabaya city using time series analysis," *Processes*, vol. 10, no. 11, p. 2454, 2022.

Fath Elrhman EL Guma holds the associate professor position in applied statistics at the Faculty of Art and Science in Baljurshi, AL-BAHA University. In 2012, he obtained a Doctor of Philosophy degree in applied statistics from the University of Sudan University of Science and Technology. The primary area of his research focuses on the study of epidemiological modeling and statistical predictions. He serves as a supervisor for graduate students pursuing master's degrees and doctorates in statistics.