International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 3 ǁ March 2024 ǁ PP. 365-373

The Design of Ensemble Deep Learning Model for the Prediction of Lassa Fever Outbreak using Multiple Data Sources

Oluwadare, Samuel Adebayo[1], Adegun, Iyanu Pelumi[1], [*]Orogun, Okunola Adebola[3], Boyinbode, Olutayo Kehinde[1], Oyekanmi, Ezekiel Olufunminiyi[2]

[1]Computer Science Department, School of Computing, The Federal University of Technology, Akure, Nigeria

[2]Computer Science Department, School of Applied and Natural Sciences, Achievers University, Owo, Nigeria

*[*3]Digital technologies, Pearson, Global Banking School, Birmingham, United Kingdom*
**Corresponding Author*

Abstract
Lassa fever is a deadly viral disease that is prevalent in West Africa. Unlike other viral diseases such COVID19 and measles, Lassa fever has no known vaccine for its prevention. Hence, the need to have a means of predicting its outbreak to sensitise stakeholders such as governments, health authorities and communities. This could help to mitigate the disaster risks associated with Lassa fever outbreak. This research investigates the relationship between Lassa fever and weather. The relationship is combined with Lassa fever incidence cases to design an ensemble Deep Learning (DL) model to predict the outbreak. Ensemble DL model is good at combining datasets from multiple sources and formats. The stack ensemble model used in this research consists of Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). Performance evaluation was achieved using standard metrics. The result obtained would assist stakeholders in developing appropriate preventive and response mechanisms against the outbreak.

Key words: *Lassa Fever, Prediction, Deep Learning, Stack Ensemble, Machine Learning.*

Date of Submission: 12-03-2024 Date of acceptance: 27-03-2024

I. INTRODUCTION

According to reports from the United States Centre for Disease Control and Prevention (CDC), annually, Lassa fever infects approximately 100,000 and 300,000 individuals in West Africa, resulting in about 5,000 deaths (John-Ugwuanya, 2021). Unfortunately, Lassa fever is one of the viral diseases that has no vaccine to prevent its outbreak (Tambo et al., 2018). This situation calls for a means by which its occurrence could be detected and predicted so that the stakeholders could prepare adequate response measures to prevent or mitigate the disaster risks. Artificial Intelligence (AI) and its related technologies have enabled useful and insightful predictions to be made from such large amount of data (Takyar, 2020). One of the unique capabilities of AI, is its ability to perform predictive modelling of infectious diseases, thereby preventing millions of lives from diseases and death (Wong *et al*., 2019). The Sustainable Development Goals (SDG) 3d aptly captured the need for the deployment of Early Warning Systems that would help to mitigate and reduce risks associated with allowing communicable diseases to assume epidemic dimension (SDG, 2015). This can be achieved by developing appropriate models that could assist in predicting the outbreak of Lassa fever.

Over the years, there have been several attempts to develop appropriate models to ensure accurate prediction of diseases outbreak. These approaches include mathematical models, statistical models, time series models, machine learning models, deep learning models and hybrids of two or more of these models. Lassa fever outbreak is one of the major challenges facing the public health system in Nigeria and some parts of West Africa because of its high fatality rate. Most previous studies that have been carried out on the prediction of Lassa fever outbreak use mathematical models. Machine learning and deep learning models are Artificial Intelligence models that have shown promising results based on their use for other known infectious diseases like measles, cholera, influenza, dengue and COVID-19 (Ahmad et al., 2021**)**. As much as we know, only very few studies have been conducted on prediction of Lassa fever outbreaks in Nigeria using machine learning or deep learning techniques. Many of the predictive models developed for infectious diseases are based on traditional linear models such as Moving Average model, Auto-Regressive Moving Average (ARIMA) model, Seasonal ARIMA model, Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX) model and Vector Autoregression (VAR) model. These traditional models can only handle linear relationships in datasets which limit their predictive ability and may not be ideal for real-life applications (Jia *et al*., 2019). These models assume a linear relationship between the past and future, which may not be applicable in complex and non-linear situations, thereby limiting their effectiveness. Another limitation of traditional models is that their performance of highly sensitive to their specifications which may include the choice of lag lengths in ARIMA models. Incorrect specification may lead to poor forecasts and minimal model performance (Barrera-Animas et al., 2022.

While the simplicity of traditional models can be an advantage, it can also be a drawback when handling complex data such as time series data. More sophisticated models like machine learning techniques may offer enhanced performance (Cai et al., 2019; Siami-Namini et al., 2018). More recently, the development of deep learning techniques has also shown a promising alternative to statistical/traditional approaches while handling complex data. Deep learning models allow computers to solve more complex problems having very diverse, unstructured, non-linear, and inter-related datasets when compared with statistical and traditional ML models (Abdelhafid Zeroual et al., 2020).Some recent studies conducted on infectious disease forecasting used machine learning and deep learning models obtained better results (Kim and Ahn, 2021**).** Also, there is a rising need to use multiple models that harness the strength of a combination of deep learning models to predict disease outbreaks. In this research, we propose the use of deep learning stacked ensemble model which combines three models to produce a single prediction that takes advantage of the strengths of the individual models. Most traditional and conventional models, make use of case data for prediction which could make their results inaccurate. Hence, this research proposes a combination of meteorological data (rainfall, temperature, and humidity) and case data (confirmed cases and death cases) for the prediction model because of the potential relationship between these two categories of data in the outbreak of Lassa fever.

II. LITERATURE REVIEW

Infectious diseases are among the leading causes of death worldwide. These include HIV/AIDS, COVID19, ebola, Lassa fever, measles, influenza, dengue and many others. A review conducted by Olumade et al. (2020) showed that in recent times, Monkey Pox, Yellow Fever, Poliomyelitis, Ebola Virus Disease (EVD) and Lassa Fever, are among the top five emerging and re-emerging infectious diseases outbreak that have caused a major public health challenge in Nigeria. In 2020, COVID-19 emerged as one of the top causes of death worldwide. COVID-19 is caused by the SARS-CoV-2 virus, according to the (CDC, 2020). The major means of preventing these infectious diseases is the development of appropriate vaccines and development of early warning systems because there is no vaccine yet for preventing Lassa fever. This makes the development of predictive models for prediction of its occurrence very important.

Infectious Diseases Forecasting Models

Infectious diseases forecasting falls under the category of time series forecasting problems or regression tasks in machine learning. It involves using mathematical and computational methods to predict unknown or future events. Predictive modelling refers to the process of using mathematical or computational techniques to forecast future events. This method involves creating a model that associates a target variable with several predictor variables. (Selam, 2012). These models are used to predict future disease outbreaks by adjusting the input parameters. Prediction of infectious disease outbreaks are often classified as “time-series forecasting problem” which involves the use of time-based (days, hours, weeks, months, year) data to find the hidden patterns to enhance decision making (Parmezan*et al*., (2019).

Machine Learning Prediction Models

Most previous studies that have been carried out on the prediction of Lassa fever outbreak use mathematical models. Machine learning and deep learning models are Artificial Intelligent models that have shown promising results based on their use for other known infectious diseases like Cholera, Influenza, Dengue and COVID-19 (Ahmad et. al., 2021**).** Many of the predictive models developed for infectious diseases are based on traditional linear models such as Auto-Regressive Moving Average (ARIMA) model. Examples of the use of mathematical, statistical or ARIMA models used for predicting infectious diseases are found in the works of (Kane *et al.*,2014; Shi *et al*., 2016; Yu *et al*., 2014; Collins and Duffy, 2023; Abdulhamid *et a*l., 2022). Some works also focused on comparative studies of some infectious diseases.For example, Tahmo et al. (2023) applied SARIMA and Poisson Regression Model techniques to evaluate the effects of Lassa Fever, COVID-19, and Cholera by analysing their interactions throughout 2021. The research found a significant dependency of Lassa fever incidents on the quantity of confirmed COVID-19 cases. However, there is room for a significant improvement in performance of the models if machine learning or deep learning models are explored.

Some recent studies conducted on infectious disease forecasting used Machine Learning (ML) and Deep Learning (DL) models obtained better results (Kim and Ahn, 2021**).** There are lots of works in this direction in recent years. Many papers published adopt different machine learning algorithms and their hybrids.

The examples of research publications that used ML, hybrid ML algorithms, DL and ensemble DL algorithms could be found in ((Zhang and Nawata, 2017; Tapak*et al*., 2019; Badkundri *et al*., 2019; Akhtar *et al*., 2019, Wang *et al*., 2020; Yang *et al*., 2020; Ribeiro *et al*., 2020; ArunKumar *et al*., 2021; Maaliw*et al*., 2021; Babski-Reeves *et al*., 2023).

It is also noted that most of these works were on COVID19 due to the pandemic nature of its outbreak. Some of these publications emphasised the need to investigate the dependency of outbreak of infectious diseases on seasonality and meteorological factors such as rainfall, humidity and temperature (Yang *et al*., 2020; Mussimeci *et al*., 2020; Watore*et al*., 2023). It was found that inclusion of meteorological data and seasonality improves prediction accuracy. During the literature search, we observed that there are few publications on the use of ML, DL and ensemble DL algorithms on the prediction of Lassa fever outbreak compared to other infectious diseases. Also, from available literature, no study has been reported on prediction of Lassa fever outbreak using multiple data sources. This research is aimed at addressing these limitations.

III. METHODOLOGY

Systems Design

The proposed system architecture presented in Figure 1 describes the major stages that are pivotal to the Lassa fever outbreak prediction process.

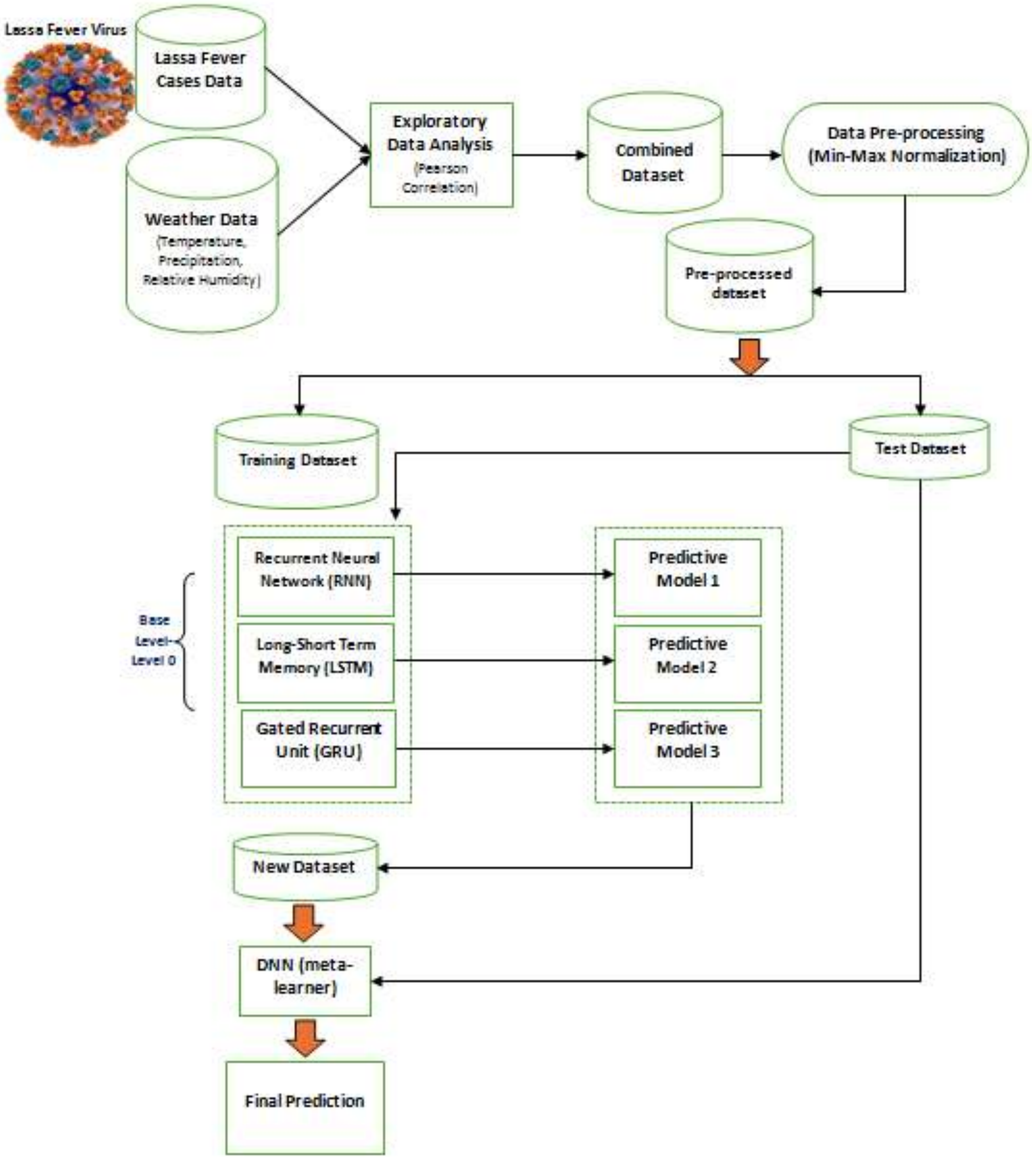

Figure 1: Architecture for the Stacked Deep Learning Ensemble (SDLE) Model

These stages are discussed in this section:

Data Collection:
This involves gathering of the relevant data namely Lassa fever case data and weather data. The structure of the data sourced from the are described in Tables 1 and 2 respectively.

Table 1: Description of Attributes of the Lassa Fever Dataset

S/N	Variable Name	Data Type	Description
1.	WeekNum	integer	Week of occurrence
2.	Date	datetime	Date of occurrence
3.	Year	datetime	Year of occurrence
4.	NumOfCases	integer	Number of confirmed cases reported for a specified week

Table 2: Description of Attributes of the Weather Dataset

S/N	Name	Data Type	Description
1.	WeekNum	integer	Week of occurrence
2.	Date	datetime	Date of occurrence
3.	Year	datetime	Year of occurrence
4.	MeanTemperature	Float	Mean Weekly temperature
5.	Meanrecipitation	Float	Mean Weekly rainfall
6.	MeanRHumidity	Float	Mean Weekly Relative Humidity

Exploratory Data Analysis
Exploratory Data Analysis (EDA) involves analysing the relationship between Lassa fever incidence data and weather data. This involves visualization of the datasets using graphical representations such as histograms, scatter plots, and heatmaps. Correlation analyses are also conducted between Lassa fever cases and weather variables - temperature, precipitation, and humidity - to understand inherent associations. The two sources of data are then integrated into a single dataset.

Data Pre-processing
Data pre-processing is a crucial step in refining the raw data into a usable format for modelling. In this research, the following significant pre-processing activities are carried out:

Data Cleaning: The purpose of data cleaning is to enhance the quality of the dataset by removing irrelevant or noisy data and handling missing values. In our dataset, identified missing values among the Lassa fever and weather data were replaced with 0.

Normalization: This involves rescaling the original data to ensure that values in the dataset are within a specified range, thereby improving the model's performance. Min-max normalization technique was selected to rescale the data to the range [0,1]. This is described mathematically in equation 1.

Given a dataset D, with has R rows (entries) and S columns (features), the normalized value of D can be computed as follows:

$$D_{norm} = \frac{D - D_{min}}{D_{max} - D_{min}} \qquad (1)$$

where D represents the original value before normalization, D_{min} represents the minimum (smallest) value in the feature that D belongs to, D_{max} represents the maximum (largest) value of that feature in the feature that D belongs to while D_{norm} is the normalized value of D after applying the normalization formula. Applying this formula to the dataset used for this research will give the following sets of equations.

$$\mathrm{L}_{norm} = \frac{\mathrm{L} - \mathrm{L}_{min}}{\mathrm{L}_{max} - \mathrm{L}_{min}} \qquad (2)$$

$$\mathrm{T}_{norm} = \frac{\mathrm{T} - \mathrm{T}_{min}}{\mathrm{T}_{max} - \mathrm{T}_{min}} \qquad (3)$$

$$\mathrm{P}_{norm} = \frac{\mathrm{P} - \mathrm{P}_{min}}{\mathrm{P}_{max} - \mathrm{P}_{min}} \qquad (4)$$

$$\mathrm{H}_{norm} = \frac{\mathrm{H} - \mathrm{H}_{min}}{\mathrm{H}_{max} - \mathrm{H}_{min}} \qquad (5)$$

where L_{norm} *represents normalized Lassa fever cases feature (column),* T_{norm} *represents temperature feature (column),* P_{norm} *represents precipitation feature (column) and* H_{norm} *represents relative humidity feature (column) in the dataset.*

The pre-processed datasets are combined by using date and week number as common identifier or timestamp into a single dataset. The combined dataset is split into 80% training set and a 20% validation (test) set for model training and evaluation.

Base Models Training

Each of the three base models namely Recurrent Neural Network (RNN), Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are trained on the training set. These base models should be able to capture patterns from all data sources.

Generate Predictions from Base Models

For each base model, predictions are made on the validation set. Thereafter, a new dataset is created where each row represents a sample in the validation set, and each column corresponds to a prediction from one of the base models.

Meta-Learner Training

The meta-learner used is this research is Deep Neural Network (DNN). It is trained on the dataset which emanates from the predictions (outputs) of the three base models.

Making Final Predictions

The trained meta-learner is used to make the final prediction which is then passed to the mobile alert system for dissemination to the stake holders via email and/or text messages.

Evaluation

The performance of the final predictions from the stacked ensemble model is evaluated by using appropriate evaluation metrics which include MAE, MSE, RMSE, MAPE.

Model Refinement

Based on the results of the performance evaluation, the hyperparameters of the base models and meta-learner are fine-tuned. This could involve testing different combinations of base-models that would produce optimal meta-learner which will ultimately lead to an efficient ensemble.

Mathematical Formulation of the Stacked Deep Learning Ensemble (SDLE) Model

A more detailed mathematical description of the proposed SDLE model is described is this section.

Notations

L_t: represent the Lassa fever cases at time t
T_t: represent the temperature at time t
P_t: represent the precipitation at time t
H_t: represent the relative humidity at time t
t: represent the time step.
w: represent the number of time steps ahead for which forecast is to be performed (for instance, if $w = 2$, the forecast is two weeks ahead)

Base Model Training

Each base model f_i $(for\ i = 1, 2, \dots, m)$ is trained on the merged data and predictions made to get the number of cases for the next w weeks. Predictions L_{t+w} that are generated are shown in equation (3.10)

$$L_{t+w} = f_i(L_{t+w-1}, T_{t+w-1}, P_{t+w-1}, H_{t+w-1})$$
$$(for\ k = 1,2, \dots, s) \qquad (10)$$

where L_{t+w} represents the predicted number of cases w weeks ahead as predicted by the i^{th} base model. f_i represents a function that performs the actual forecast and i is the index that runs through all the base models in the ensemble.L_{t+w-1}represent the number of Lassa fever cases one week prior to the forecast week (i.e., the specific future week for which prediction is to be made), T_{t+w-1}represents temperature value one week prior to the forecast weekP_{t+w-1}represents precipitation value one week prior to the forecast week and H_{t+w-1} represents relative humidity value one week prior to the forecast week.

Constructing Meta-features

Predictions from the three base models are then combined to form meta-features represented by X_{t+w} as seen in equation (11).

$$X_{t+w} = [L_{t+w,1}, L_{t+w,2}, \dots, L_{t+w,s}] (11)$$

where X_{t+k} is a vector that stacks predictions from all the base models for the time $t + k$. If there are s base models, then the vector will have s components.

Meta-model Predictions

A Deep Neural Network (DNN) model will act as a meta-model g that takes the meta-features to generate the ensemble's final prediction. This is shown in Equations 11 and 12.

$$L_{t+w}^{meta} = g(X_{t+w}) \quad (12)$$

where L_{t+w}^{meta} is the ensemble's final prediction for $t + w$.

Base-level Deep Learning Algorithms

Three base-level deep learning algorithms were used for the purpose of this study. They are: RNN, LSTM and GRU. These models are described in the next sub-sections.

Recurrent Neural Network

RNNs can be referred to as a robust type of feed-forward neural network (Yang et al., 2020). Because of their internal memory, RNNs can remember important information about whatever input they receive, which allows them to be very precise in predicting a next action. In a feed-forward neural network, information moves in a single direction which is from the input layer, through the hidden layers, to the output layer. The information moves straight through the network and never have contact with a node more than once. They have no memory of the input they receive and cannot predict the next input. On the other hand, RNN allow the information to cycle through a loop. To make a decision, it considers the current input and also what it has learned from the inputs received previously (Kamal et al., 2020). The RNN output calculation is based on an iterative calculation of the output of Equations 13 and 14:

$$h_t = \sigma_h(-W_h x_t + U_h y_{t-1} + b_n) \quad (13)$$

$$y_t = \sigma_y\left(W_y h_t + b_n\right) \quad (14)$$

where x_t is the vector of the input layer, h_t is the vector of the hidden layer, y_t is the vector of the output layer, W, U and b are the matrix or vector of the weight parameter and σ_h and σ_y are the activation functions.

Long Short-Term Memory

Long short-term memory networks can be referred to as an extension of recurrent neural networks, which basically extends the memory. LSTM enables RNNs to remember inputs over a long period of time. This is because LSTM stores information in a memory, similar to a computer memory. The LSTM has the ability to read, write and delete information from its memory (Donges, 2019). It is made up of a set of recurrently connected blocks, known as memory blocks. Each block has one or more recurrently connected memory cells and the input, output and forget gates that provide continuous read, write and reset operations (Graves et al., 2005). LSTM cells act as a memory to write, read and erase information based on the decisions made by the input, output and forget gates respectively (Venna et al., 2018). The mathematical description is given as follows:

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (15)$$

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \quad (16)$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (17)$$

$$\tilde{c}_t = tanh(Wht - 1 + Ux_t + b) \quad (18)$$

$$c_t = (f_t \,.\, c_{t-1} + i_t \,.\, \tilde{c}_t) \quad (19)$$

$$h_t = o_t \,.\, Tanh(c_t) \quad (20)$$

$$y_t = h_t \quad (21)$$

where i_t represents the input gate which decides if the current input should be incorporated into the memory state, f_t represents the forget gate that determines if the forget gate has been previously retained in memory. The memory state o_t represents the output gate tasked with assessing the amount of memory output passed into the hidden layer (Yang *et al*., 2020). $\tilde{c}_t$ represents a candidate hidden state formulated based on the previous hidden state and the present input, c_t symbolises the unit's internal memory while h_t stands for the output of the hidden state, derived from the product of the memory and the output gate (Goytom & Sankaran, 2019).

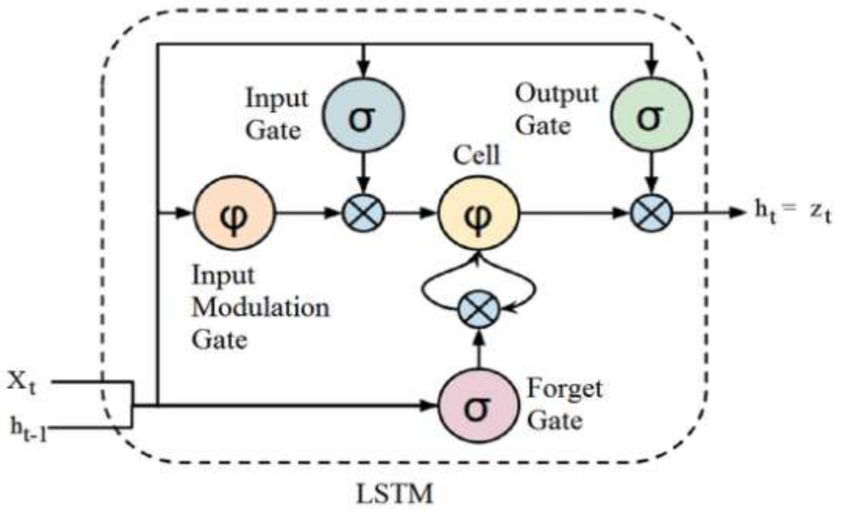

Figure 4: LSTM Diagram

Gated Recurrent Units

Gated Recurrent Units (GRU) is a more recent version of RNN like LSTM with a simpler cell architecture (Kamal et al., 2020). They are designed to learn with long-range dependencies and are similar to LSTM in gating units; the information flow occurs with short memory and few gates (two gates, reset and update gate). The function of the update gate is like that of the forget and input gates of an LSTM. It takes the decision on what information to keep, add or discard. Meanwhile, the reset gate regulates the amount of past information that should not be remembered. Gated Recurrent Units are little faster to train than their LSTM counterpart because of fewer gates. The mathematical formula is described as follows:

$$z_t = \sigma(W_{xz} x_t + W_{dz} d_{t-1} + b_z \qquad (21)$$

$$r_t = \sigma(W_{xr} x_t + W_{dr} d_{t-1} + b_r \qquad (22)$$

$$d_t = Tanh(Wxdx_t + W_{dd}(r_t \,.\, d_{t-1}) + b_d \qquad (23)$$

$$\tilde{d}_t = z_t \,.\, d_{t-1} + (1 - z_t).\tilde{d}_t \qquad (24)$$

where z_t represents the update gate, r_t represents the reset gate, x_t represents the current input, d_{t-1} is the hidden state of the previous input, W_{xr}, W_{xz} and W_{dr} are weight parameters while b_r and b_z represents biases (Yao et al., 2022).

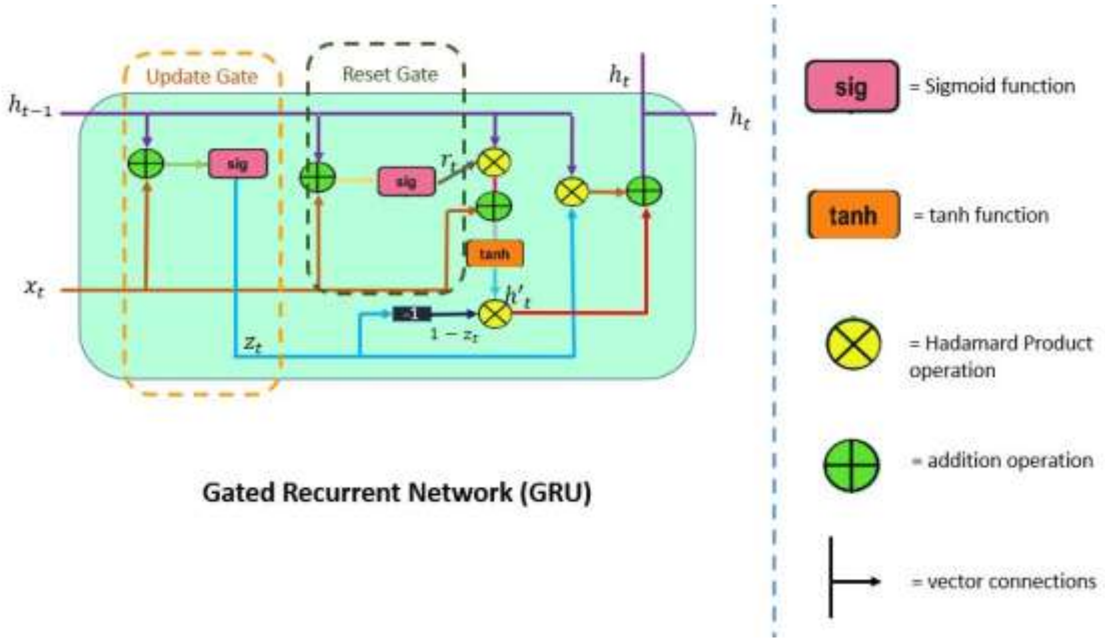

Figure 5: Gated Recurrent Unit Network Diagram

Performance Metrics

To assess how well our model predicts Lassa fever confirmed cases Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE; and Mean Absolute Percentage Error (MAPE), denoted in Equations 25 to 28 respectively will be employed.

MAE, standing for Mean Absolute Error, calculates the average of the absolute discrepancies between forecasted and true values.

Mean Absolute Error (MAE):MAE is used to calculate the average absolute difference between predicted values and actual values. The mathematical formula is shown in Equation 25.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (25)$$

Mean Squared Error (MSE): MSE is a metric that used to measure the average of the squares of the differences between the actual values and the values predicted by the model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i|^2) \qquad (26)$$

Root Mean Squared Error (RMSE):RMSE is a metric that assesses the prediction of a model by calculating the square root of the average squared differences (standard deviation) between the actual values and the predicted values.. The mathematical formula is shown in Equation (3.27)

oot Mean Squared Error (RMSE) is a metric that assesses the accuracy of a prediction model by calculating the square root of the average squared differences between the predicted and actual values.

$$RMSE = \sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]} \qquad (27)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \qquad (28)$$

where n is the number of observations, y is the vector of observed values and $\hat{y}$ is the vector of predicted values.

IV. Conclusion and Recommendation

This paper presents the design of an ensemble DNN Model that could predict Lassa fever outbreak using case data and weather (temperature, rainfall and relative humidity) data. It is established in epidemiological studies of some viral diseases such as COVID19, Influenza, Denque fever and so on, that diseases outbreak is associated with weather conditions. It therefore, expected that the design proposed in this research would produce more accurate prediction. The inclusion of weather data in this research should provide significant contribution to knowledge in the use of Machine Learning algorithms for Lassa fever outbreak prediction. The next phase of our research will be the implementation of this design and the development of Mobile Alert System to make predictions on future outbreak available to stakeholders. This will improve the preparedness and response capacities of stakeholders in mitigating disaster risks associated with Lassa fever outbreak. It is observed that there is a dearth of an established dataset for Lassa fever research. Our future work will also involve the establishment of a Lassa fever dataset.

Acknowledgement

This research was sponsored by Tertiary Education Fund, National Research Fund (GRANT AWARD_NRF_SETI_ICT_00058, 2021) of the Federal Republic of Nigeria".

References

[1]. Abdulhamid, A., Hussaini, N., Musa, S. S., & He, D. (2022). Mathematical analysis of Lassa fever epidemic with effects of environmental transmission. Results in Physics, 35, 105335.

[2]. Ahmad, H. F., Khaloofi, H., Azhar, Z., Algosaibi, A. and Hussain, J. (2021). An Improved COVID-19 Forecasting by Infectious Disease Modelling Using Machine Learning.

[3]. Appl. Sci. 2021, 11(23), 11426; https://doi.org/10.3390/app112311426, Retrieved: June 17, 2022.

[4]. Akhtar, M., Kraemer, M. U., & Gardner, L. M. (2019). A dynamic neural network model for predicting risk of Zika in real time. BMC medicine, 17(1), 171.

[5]. ArunKumar, K. E., Kalaga, D. V., Kumar, C. 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.

[6]. Babski-Reeves, K., Eksioglu, B., & Hampton, D. Analysis of Machine Learning and Ensemble Methods for Forecasting COVID-19 Cases in California (2023. InProceedings of the Institute of Insdustrial and Systems Engineer (IISE) Annual Conference & Expo 2023, 91-96.

[7]. Badkundri, R., Valbuena, V., Pinnamareddy, S., Cantrell, B., & Standeven, J. (2019).

[8]. Forecasting the 2017-2018 Yemen Cholera Outbreak with Machine Learning. arXiv preprint arXiv:1902.06739.

[9]. Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Machine Learning with Applications, 7, 100204.

[10]. Cai, M., Pipattanasomporn, M., & Rahman, S. (2019). Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Applied energy, 236, 1078-1088.

[11]. Centre for Disease Control and Prevention. Lassa Fever. Retrieved from http://www.cdc.gov/vhf/lassa/

[12]. Collins, O. C., & Duffy, K. J. (2023). Using Data of a Lassa Fever Epidemic in Nigeria: A Mathematical Model Is Shown to Capture the Dynamics and Point to Possible Control Goytom, I., & Sankaran, K. Forecasting Maxima in Climate Time Series (2019, October). In 9th International Workshop on Climate Informatics (IWCI) (pp. 1-5).

[13]. Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850. Methods. Mathematics, 11(5), 1181.

[14]. Jia, W., Wan, Y., Li, Y., Tan, K., Lei, W., Hu, Y., Ma, Z., Li, X. & Xie, G. (2019). Integrating Multiple Data Sources and Learning Models to Predict Infectious Diseases in China. AMIA Summits on Translational Science Proceedings, 2019, 680.

[15]. John-Ugwuanya A. G, Ifunanya J. E. and Nnenna U. (2021). Epidemiological trends of Lassa

[16]. fever in Nigeria from 2015-2021: A review. Therapeutic Advances in Infectious Disease 2021, Vol. 8: 1–7 DOI: 10.1177/ 20499361211058252

[17]. Kamal, I. M., Bae, H., Sunghyun, S., & Yun, H. (2020). DERN: Deep Ensemble Learning Model for Short-and Long-Term Prediction of Baltic Dry Index. Applied Sciences, 10(4), 1504.

[18]. Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC bioinformatics, 15(1), 276.

[19]. Kim, J. and Ahn, I. (2021). Infectious disease outbreak prediction using media articles with machine learning models. https://www.nature.com/articles/s41598-021-83926-2, Retrieved: June 17, 2022.

[20]. Maaliw, R. R., Ballera, M. A., Mabunga, Z. P., Mahusay, A. T., Dejelo, D. A., & Seño, M. P. (2021, October). An ensemble machine learning approach for time series forecasting of COVID-19 cases. In 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 0633-0640). IEEE.

[21]. Milinovich, G. J., Williams, G. M., Clements, A. C., & Hu, W. (2014). Internet-based surveillance systems for monitoring emerging infectious diseases. The Lancet infectious diseases, 14(2), 160-168.

[22]. Olumade, T.J., Adesanya, O.A., Fred-Akintunwa, I.J., Babalola, D.O., Oguzie, J.U.,

[23]. Ogunsanya, O.A., George, U.E., Akin-Ajani, O.D. and Osasona, D.G (2020). Infectious disease outbreak preparedness and response in Nigeria: history, limitations and recommendations for global health policy and practice. AIMS Public Health, 7(4), 736.

[24]. Parmezan, A. R. S., Souza, V. M., & Batista, G. E. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Information Sciences, 484, 302-337.

[25]. Rees, E. E., Ng, V., Gachon, P., Mawudeku, A., McKenney, D., Pedlar, J., Parmely, Y. J. & Knox, J. (2019). Early detection and prediction of infectious disease outbreaks. CCDR, 45, 5.

[26]. Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & dos Santos Coelho, L. (2020). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos, Solitons & Fractals, 135, 109853.

[27]. Selam, A. (2012). Predicting the Occurrence of Measles Outbreak in Ethiopia Using Data Mining Technology (Doctoral dissertation, Addis Ababa University).

[28]. SDG (2015). Transforming the World: The 2030 Agenda for Sustainable Development https://sustainabledevelopment.un.org/content/documents/ 21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf,retrieved: June 10, 2022

[29]. Shi, Y., Liu, X., Kok, S. Y., Rajarethinam, J., Liang, S., Yap, G., Chong, C.S., Lee, K.S., Tan, S.S., Chin, C.K.Y. & Lo, A. (2016). Three-month real-time dengue forecast models: an early warning system for outbreak alerts and policy decision support in Singapore. Environmental health perspectives, 124(9), 1369-1375.

[30]. Siami-Namini, S., Tavakoli, N., &Namin, A. S. (2018, December). A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE international conference on machine learning and applications (ICMLA) (pp. 1394-1401). IEEE.

[31]. Tahmo, N. B., Wirsiy, F. S., & Brett-Major, D. M. (2023). Modeling the Lassa fever outbreak synchronously occurring with cholera and COVID-19 outbreaks in Nigeria 2021: A threat to Global Health Security. PLOS Global Public Health, 3(5), e0001814.

[32]. Tambo, E., Adetunde, O. T., &Olalubi, O. A. (2018). Re-emerging Lassa fever outbreaks in Nigeria: Re-enforcing" One Health" community surveillance and emergency response practice. Infectious diseases of poverty, 7(02), 74-80.

[33]. Takyar, A. (2020). How AI helps in early infectious disease surveillance? LeewayHertz. https://www.leewayhertz.com/ai-infectious-disease-surveillance/

[34]. Tapak, L., Hamidi, O., Fathian, M., & Karami, M. (2019). Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from sentinel sites of healthcare centers in Iran. BMC research notes, 12(1), 353.

[35]. Wang, L., Adiga, A., Venkatramanan, S., Chen, J., Lewis, B., & Marathe, M. (2020). Examining deep learning models with multiple data sources for COVID-19 forecasting. In 2020 IEEE International conference on big data (Big Data), 3846-3855, IEEE.

[36]. Wathore, R., Rawlekar, S., Anjum, S., Gupta, A., Bherwani, H., Labhasetwar, N., & Kumar, R. (2023). Improving performance of deep learning predictive models for COVID-19 by incorporating environmental parameters. Gondwana Research, 114, 69-77.

[37]. Wong, Z. S., Zhou, J., & Zhang, Q. (2019). Artificial intelligence for infectious disease big data analytics. Infection, disease & health, 24(1), 44-48.

[38]. Yang, C. T., Chen, Y. A., Chan, Y. W., Lee, C. L., Tsan, Y. T., Chan, W. C., & Liu, P. Y. (2020). Influenza-like illness prediction using a long short-term memory deep learning model with multiple open data sources. The Journal of Supercomputing, 1-27.

[39]. Yao, L., Hao, K., & Wei, B. (2022, March). Dynamic Feature Selection Model Based on Gated Recurrent Units for Time Series Forecasting. In 2022 2nd International Conference on Computer, Control and Robotics (ICCCR) (pp. 139-145). IEEE.

[40]. Yu et al. (2014) developed a new hybrid model by combining both linear and non-linear models for forecasting HFMD in China.

[41]. Zhang, J., & Nawata, K. (2017). A comparative study on predicting influenza outbreaks. Bioscience trends, 11(5), 533-541.