

# Weather Forecasts Using Deep Learning Models for Urban Areas in Nigeria

Tayo P. Ogundunmade<sup>1\*</sup>; Opeyemi Adewale Okunoye<sup>2</sup>

<sup>1,2</sup>Computational Unit, Department of Statistics, University of Ibadan, Ibadan, Nigeria.

Corresponding Author: Tayo P. Ogundunmade<sup>1\*</sup>

Publication Date: 2026/03/18

**Abstract:** The data includes information from ten years, from 2014 up to 2023. Weather data is recorded monthly from the meteorological station at the Forestry Research Institute in Nigeria. Important data recorded includes temperature in Celsius, humidity, wind speed in km/h, and relative humidity. This data is recorded in seven columns and 120 rows. It includes temperature, humidity, wind speed, and relative humidity. Three models are applied in solving this problem. These models are time-aware long short-term memory networks, gated recurrent units, and the Transformer. These models have been improved with an attention-based approach to interpretability, inspired by RETAIN. The GRU model can forecast data up to six months into the future. This data shows an inverse correlation with temperature and relative humidity. Relative humidity goes down to  $72\% \pm 5$ , indicating pre-rainy conditions. This occurs while temperatures peak at  $31^\circ\text{C} \pm 0.8$  in Month 4. In Month 6, temperatures drop to  $28.2^\circ\text{C} \pm 0.5$ , and relative humidity rises to  $85\% \pm 3$ , indicating that rain is on the way. Wind speeds decrease to 9.8-10.5 km/h in Months 3 and 4, when temperatures are at their peak and relative humidity is at its lowest. The weather forecasting model has shown how GRU, Time Aware, and even the Transformer can be applied in solving weather problems in Nigeria.

**Keywords:** Weather Prediction, Long Short-Term Memory (LSTM) Networks, Gated Recurrent Units (GRUs), Transformer Models, Deep Learning.

**How to Cite:** Tayo P. Ogundunmade; Opeyemi Adewale Okunoye (2026) Weather Forecasts Using Deep Learning Models for Urban Areas in Nigeria. *International Journal of Innovative Science and Research Technology*, 11(3), 1089-1100. <https://doi.org/10.38124/ijisrt/26mar796>

## I. INTRODUCTION

Weather forecasts have progressed from simple sky gazing to big data and intelligent machines. In today's world, accurate weather forecasts are no longer a luxury but a necessity for farming, flying, power supplies, water management, and preparing for disasters. In Nigeria, however, accurate forecasts are more critical due to rapid urbanization and changing climate conditions. For Nigerian cities, accurate forecasts are now more critical than ever. In contrast to this situation, old-fashioned methods used in advanced countries involve extensive observation networks and powerful computers. These are not readily available in Nigeria. As a solution to this problem, researchers are increasingly relying on Machine Learning (ML) and Deep Learning (DL), especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to improve forecasts for Nigerian cities.

ML is revolutionizing meteorology on a large scale. Unlike conventional methods that rely on detailed physical equations to analyze atmospheric conditions, ML is able to recognize intricate relationships between atmospheric variables on its own. This is particularly useful for countries

such as Nigeria, where accurate forecasts are difficult to obtain due to a lack of weather stations. As Ali et al. [2] stated, ML is able to learn from various data sources such as satellite images, radar data, and ground-based observations to accurately predict temperature, humidity, rainfall, and wind speed. Mohammed et al. [3] added that data-driven methods allow for flexible frameworks that combine multiple data sources to minimize prediction uncertainties. By automatically recognizing intricate relationships between variables over space and time, ML-based forecasts are found to be more accurate than conventional methods, especially for short-term and medium-term forecasts.

A more sophisticated type of machine learning, termed deep learning, has also been found to possess immense potential in improving weather prediction outcomes. This type of machine learning utilizes sophisticated neural network architectures that are capable of automatically learning complex patterns present in large, multidimensional weather data sets [4]. Among various types of deep learning algorithms, Convolutional Neural Networks (CNNs) are found to be particularly good at recognizing patterns that are spatial in nature. These networks analyze weather maps, satellite imagery, and other data sets to identify patterns such

as cloud movement and rain patterns. On the other hand, Recurrent Neural Networks (RNNs) and their variants, including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are found to be particularly good at recognizing patterns that are temporal in nature, including sequences of atmospheric data sets, such as hourly or daily readings of atmospheric data sets ([5]; [6]).

Around the world, CNNs and RNNs have been employed in weather forecasting, resulting in remarkable outcomes. To cite an example, Shi et al. [7] were among the first few researchers who employed convolutional long short-term memory networks in short-term rainfall prediction. They demonstrated that deep learning methods can be employed to estimate rain intensity hours in advance. Expanding on this work, Sønderby et al. (2020) introduced MetNet, a deep neural network developed by Google that offers very detailed precipitation forecasts for North America. In a related area of study, Rasp et al. [8] demonstrated that neural networks can be employed to simulate atmospheric processes in climate models, resulting in increased accuracy and significantly reduced computational costs. These developments have inspired other researchers worldwide to employ similar deep learning methods to tackle local and regional weather forecasting problems in Africa.

For Africa, deep learning methods are increasingly being employed in meteorology, owing to enhanced access to remote-sensing data. In this regard, Ogutu et al. [9] demonstrated that machine learning methods employing satellite data improved rainfall prediction accuracy in East Africa, extending prediction times significantly. In a related area of study, Fashae and Ajayi [10] demonstrated that deep learning methods employing satellite data were effective in predicting short-term temperature variations in tropical West Africa. These studies demonstrate that AI-based methods are easily adaptable to regions that offer limited data sources for training machine learning models. In Nigeria, Eze et al. [11] and Olowe et al. [12] demonstrated that deep learning methods are effective in enhancing localized weather forecasting in regions such as Lagos and Ibadan, where unique microclimates are formed owing to increased levels of urban heat and pollution.

Urbanization plays a significant role in the modification of local climatic conditions observed in Nigerian metropolitan areas. This leads to the formation of the Urban Heat Island (UHI) phenomenon, where temperatures are higher, humidity levels are low, and the pattern of precipitation is altered compared to the surrounding areas. However, existing meteorological models are usually not sensitive to such local-scale variations. According to Adelekan [14], flooding observed in Nigerian metropolitan areas is usually associated with the absence of adequate early warning systems and low-resolution weather forecasts. However, to address the aforementioned issues, machine learning (ML) and deep learning (DL) techniques can be employed to generate more accurate and detailed weather forecasts. Such sophisticated techniques can help to address the aforementioned challenges associated with existing models.

Convolutional Neural Networks (CNNs) are found to be highly effective tools for the analysis of spatial meteorological data. This tool can be used to analyze multi-channel satellite imagery to identify the formation of cloud structures and convective areas associated with the formation of rainfall. This tool can be used to predict rainfall over Nigerian metropolitan areas. Ogunjo & Akinbobola [15] have shown the effectiveness of the CNN tool to predict rainfall over Nigerian metropolitan areas using satellite imagery. This tool can be used to identify local-scale features such as coastal convection and the formation of storms associated with the urban heat island phenomenon.

Recurrent Neural Networks (RNN) and their sophisticated variants, such as Long Short-Term Memory Networks (LSTM Networks), provide a robust methodology for understanding the dynamics in the changing weather patterns in Nigeria. The LSTM model has the capability to identify long-term dependencies and patterns in the data. For example, Aremu and Oladimeji [16] successfully applied the LSTM model for temperature and humidity prediction in Northern Nigeria. The study concluded that the model was superior in performance compared to the conventional statistical model, such as Autoregressive Integrated Moving Average (ARIMA).

Currently, in the global community, researchers have successfully applied a hybrid model that combines CNN and RNN for accurate prediction in the changing weather patterns. The model combines the strengths of both CNN and RNN in accurately identifying the spatiotemporal dynamics in the changing weather patterns. Leinonen et al. [17] successfully applied a hybrid model that combines CNN and RNN for precipitation prediction in Europe. The study concluded that the model was stable and accurate in producing results. The study by Mahmud et al. [18] concluded that a hybrid model that combines CNN and LSTM for tropical cyclone prediction in the Bay of Bengal was highly effective. The study concluded that the model was effective since the region's weather patterns were similar to those in West Africa. The model has a high potential for being applied in Nigeria for accurate prediction in the changing weather patterns. The model can be effective in Nigeria since the region can benefit from the integration of satellite images and time-series data for accurate prediction of the changing weather patterns.

In addition to that, the importance of machine learning in the weather and climate systems in Nigeria is quite significant. For instance, organizations like the Nigerian Meteorological Agency (NiMet) and the National Space Research and Development Agency (NASRDA) are improving their data collection systems using ground-based weather stations, remote sensing systems, and inter-country research collaborations. The integration of machine learning (ML) and deep learning (DL) into these systems has the potential to greatly improve early warning systems for flooding, droughts, and stormy weather. This will greatly improve preparedness for natural disasters, thereby protecting lives and property. According to Okafor et al. [19], these systems will greatly help in designing the country's

climate adaptation policies, help the farming population improve their planning for better harvests, and reduce economic losses resulting from natural calamities.

Moreover, this is also in line with the overall aspirations of the country pertaining to digital transformation and sustainable development. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are examples of ML approaches that are cost-effective and can be implemented on open-source platforms with minimal computing capabilities. The development of cloud-based services, including Google Earth Engine, is also allowing Nigerian researchers greater access to effective data processing capabilities. This is also taking the development of ML-based weather prediction systems specific to particular cities and regions of the country into a positive direction.

The recent advancements in AI-driven weather forecasting systems also highlight the overall accuracy of these contemporary approaches. For example, Pathak et al. [20] proposed FourCastNet, a neural operator-based model that is capable of delivering accurate high-resolution global forecasts that are comparable to the overall accuracy of the best Numerical Weather Prediction (NWP) models, but at significantly faster speeds. Similarly, Lam et al. [21] also proposed transformer architectures for medium-range forecasting that also demonstrated high accuracy in terms of consistency and spatial details of the overall predictions for the entire globe. These advancements are indicative of the overall transformative capabilities of DL in the field of meteorology on a global scale.

A particularly exciting opportunity exists in the use of transfer learning and data fusion techniques to adapt machine learning algorithms from a global context to a local context. Reichstein et al. [22] have shown that DL algorithms can be adapted for local use by fine-tuning them using local datasets. Such an approach would be highly relevant in Nigeria since the use of data fusion techniques to incorporate local datasets and global reanalysis datasets such as the ERA5 model could greatly improve the accuracy and reliability of weather forecasts in Nigeria. Apart from short-term weather forecasts, CNN and RNN algorithms also offer exciting opportunities for environmental monitoring and long-term climate studies. Their potential for handling complex data makes them highly suitable for analyzing complex spatiotemporal data such as those obtained from environmental and climate studies. For instance, Vandal et al. [23] have successfully used DL algorithms to upscale coarse-resolution global climate data and create high-resolution regional climate simulations. Such an approach in Nigeria could help in the development of insights for use in climate impact studies. Such studies could help in the monitoring and assessment of long-term rainfall patterns in Nigeria, temperature extremes in Nigeria, and flood risks in Nigeria. According to the literature, machine learning algorithms greatly improve the accuracy and efficiency of weather forecasts compared to conventional methods. Kadow et al. [24] have shown that neural network-based algorithms perform better than statistical interpolation techniques in the reconstruction of climate patterns. Such results from Africa and Nigeria suggest that the use of CNN

and RNN algorithms in Nigeria could greatly improve short- and medium-term weather forecasts. Such an approach could help Nigeria overcome the technology gap that has long limited the precision and accuracy of weather forecasts in Nigeria.

To summarize, the analysis of the machine learning (ML) and deep learning (DL) models for weather prediction within the urban environments of Nigeria represents a significant amalgamation of computational innovation, climate adaptation, and scientific inquiry. The use of both convolutional neural networks (CNNs) and recurrent neural networks (RNNs), individually or collectively, represents a powerful tool for the analysis of the complex meteorological systems within the country. The use of AI-powered weather prediction systems not only represents the potential for more accurate weather predictions relevant to the country but also for the achievement of sustainability goals. With the increasing urbanization of the country, the use of AI-powered systems for weather prediction represents the potential for not only improving the country's meteorological capabilities but also for addressing the challenges of climate change.

## II. MATERIALS AND METHODS

In this research, the problem of weather forecasting is addressed with the application of various deep learning models with an experimental research approach. The research design is based on a straightforward step-by-step approach that includes data collection, data preprocessing, model development, model training, model evaluation, and results interpretation. The research is quantitative in nature, as it is based on the application of various data related to weather patterns. The research is based on the application of various weather-related data with a focus on time-aware LSTM, GRU, and transformer models with the integration of attention-based interpretations based on the RETAIN approach.

### A. Data Collection and Description

This dataset records monthly weather readings obtained over a period of over ten years, spanning from 2014 to 2023. The data was obtained from the meteorological station of the Forestry Research Institute of Nigeria. The data recorded includes temperature in Celsius, humidity, wind speed in kilometers per hour, and relative humidity, arranged in seven columns and 120 rows of data. The data was obtained through automated weather stations that are calibrated for accuracy in data collection. To ensure data accuracy, timestamps, geographical information, and comments on the measuring devices used in obtaining the data are also recorded. The data was obtained from a private source and later uploaded onto Kaggle. The link to access this data on Kaggle is <https://www.kaggle.com/datasets/faruqtaiwo/weather-data>.

### ➤ Data Preprocessing

Preprocessing is a major part of ensuring that the data is clean for deep learning. For data gaps less than three months, we used linear interpolation for filling in missing data due to sensor glitches or transfer hiccups. For data gaps greater than three months, we used seasonal decomposition for filling in

the gaps. We also dropped any feature that had over 15% missing data points to prevent biasing the model. We used min-max scaling for all numerical features, which scales the data between 0 and 1. This helps the model train smoothly. Irregularly spaced timestamps were resampled into regular monthly intervals. Values are then averaged over these intervals. We also engineered features by creating lag features (1-month lag, 3-month lag, 12-month lag) and rolling statistics like a 3-month moving average for helping the model detect seasonal patterns [28].

*B. Model Architectures and Training*

In order to understand the way things evolve over a period of time and to extract useful patterns from the data, we utilized various architectures of deep learning. Time-Aware LSTM helped to deal with the problem of irregular time gaps by adjusting its memory according to the time elapsed. RETAIN is a transparent RNN that helped to understand the exact time steps that were involved in the predictions. Clinical BERT is a transformer architecture that is typically fine-tuned for dealing with categorical or event-based data. This helped to add a layer of context to the analysis. GRU is a variant of the RNN architecture that is typically utilized for dealing with the time dependencies of the data. This is a leaner variant of the LSTM architecture that is typically faster. All these architectures helped to deal with the complex nature of the data from the clinical as well as the weather-related domains.

➤ *Time-Aware LSTM Model*

The Time-Aware LSTM (T-LSTM) is designed to handle the irregular time gaps between data samples, unlike the traditional LSTM that assumes all data samples are at evenly spaced time intervals [29]. The T-LSTM is designed to adjust the memory cell based on the time since the last observation was made. The TA-LSTM is an extension of the traditional LSTM that incorporates time interval features to improve the ability to detect patterns in weather data. The mathematical equations for the meteorological data inputs are:

- **Input Representation:** The Time-Aware LSTM combines both meteorological variables and temporal information, where  $x_t \in \mathbb{R}^4$  represents the normalized weather features (temperature, humidity, wind speed, and relative humidity) at time step  $t$ , while  $\Delta_t \in \mathbb{R}$  captures the elapsed time since the previous observation, addressing irregularities in the dataset's sampling frequency. Let  $x_t \in \mathbb{R}^4$  be the input vector at time  $t$  (containing the 4 normalized weather variables), and let  $\Delta_t \in \mathbb{R}$  be the time interval since the last observation.
- **Time-Augmented Input:** combines raw weather variables with engineered time features (like logarithmic/reciprocal time intervals) to help the model explicitly track irregular observation patterns. This allows the LSTM to distinguish between regular monthly readings and irregular gaps while maintaining temporal awareness. It is given by:

$$z_t = [x_t; \phi(\Delta_t)] \in \mathbb{R}^5 \tag{1}$$

Where  $\phi(\Delta_t) = \left[ \log(\Delta_t + 1), \frac{1}{(\Delta_t + 1)} \right]$  is a time featured transformation

- **TA-LSTM Gates;**  
For each time step with memory cell  $c_t$  and hidden state  $h_t$  ;

Input gate:  $i_t = \sigma(W_i \cdot [h_{t-1}^{-1}; z_t] + b_i)$

Forget gate:  $f_t = \sigma(W_f \cdot [h_{t-1}^{-1}; z_t] + b_f)$  (2)

- **Time –Aware Modulations:**

$$\begin{aligned} \checkmark \quad g_{t1} &= \tanh(W_g \cdot [h_{t-1}^{-1}; z_t] + b_g) \\ \checkmark \quad g_t &= i_t \odot g_{t1} + \alpha \Delta_t \odot \tanh(W_\Delta \cdot \Delta_t + b_\Delta) \end{aligned}$$

Where,  $\alpha \Delta_t = \sigma(W_\alpha \cdot \Delta_t + b_\alpha)$  is a time –gating parameter

Cell state update:  $c_t = f_t \odot c_{t-1} + g_t$

- **Output Gate:**

$$\begin{aligned} \checkmark \quad o_t &= \sigma(W_o \cdot [h_{t-1}^{-1}; z_t] + b_o) \\ \checkmark \quad h_t &= o_t \odot \tanh(c_t) \\ \checkmark \quad \text{Prediction Output } (\hat{y}_t^{+1}) &\text{ in Time – Aware LSTM} \end{aligned}$$

The prediction output is the final step where the time-aware LSTM generates forecasts for future weather variables. Mathematically, it is computed as:

$$\hat{y}_t^{+1} = W_y \cdot h_t + b_y \tag{3}$$

➤ *Implementation of Time-Aware LSTM Model on the Weather Data*

This section explains the process of how a Time-Aware LSTM (TA-LSTM) is developed for the prediction of weather variables, which are temperature, humidity, wind speed, and relative humidity, using the Forestry Research Institute of Nigeria Meteorological Station dataset from 2014 to 2023. The TA-LSTM model addresses the issue of time gaps between the samples, which improves the prediction accuracy. Once the dataset is cleaned, the missing values are handled, and the normalization step is completed, the dataset is now ready for modeling. From here, the dataset can either be normalized using the Python normalize function or the standard normalization equation.

$$x_{norm} = \frac{x - \min(X)}{\max(X) - \min(X)}$$

The result after normalizing the data is given below.

Table 1 Normalized Data

Normalized Weather Data:							
S/N	YEARS	MONTH	TEMPRATURE	HUMIDITY	WIND	RELATIVE	HUMIDITY
1	2014	JAN	0.015732	0.128571	0.123377		0.547619
2	2014	FEB	0.018315	0.134884	0.136465		0.380952
3	2014	MAR	0.383608	0.130303	0.124834		0.619048
4	2014	APR	0.188118	0.125974	0.123126		0.738095
5	2014	MAY	0.630998	0.121818	0.121122		0.785714
6	2014	JUN	0.270340	0.115736	0.121122		0.952381
7	2014	JUL	0.171429	0.109524	0.118577		1.000000

➤ *Gated Recurrent Unit (GRU) Model*

Gated Recurrent Units (GRUs) are a type of streamlined version of recurrent neural networks (RNNs) that have been proposed for handling issues of long-term dependencies in sequences by using a less complex architectural framework compared to that of long short-term memory (LSTM) units. In this regard, the GRU model utilizes a gating mechanism involving both update and reset gates for controlling the flow of information throughout the RNN. In this regard, the update gate is used for controlling the extent of forgetting or remembering the previous hidden state information by the RNN model, whereas the reset gate controls the extent of previous hidden state information incorporated into the activation candidate. Due to the less complex architectural framework used by the GRU model compared to LSTM units, it has the added advantage of increased computational efficiency. Based on empirical results from a number of machine learning tasks, it has been observed that the GRU model has the potential for outperforming the LSTM model in a number of scenarios. In addition to dynamic risk analysis, the GRU model has shown promising results for a number of different tasks, including speech recognition, splice site recognition, and forecasting. In addressing the vanishing gradient problem for identifying long-range dependencies in sequences, the GRU model utilizes a gating mechanism for controlling the flow of information throughout the RNN model.

The GRU layer uses a number of mathematical formulas to update and regulate its internal states. Let's designate the hidden state at timestep t-1 as  $h_{t-1}$ , the input to the GRU layer at timestep t as  $x_t$ , and the updated hidden state at timestep t as  $h_t$ . How much of the new candidate activation  $h_t$  to take into account and how much of the prior concealed state to keep is decided by the updates gate  $z_t$ . It is computed in Equation 4.

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{4}$$

Where  $\sigma$  stands for the sigmoid activation function, and  $W_z$  and  $U_z$  are weight matrices connected to the update gate. The amount of the new candidate activation to update the hidden state and the amount of the old hidden state to forget

are decided by the reset gate  $r_t$ . Equation 5 is used to calculate it.

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{5}$$

Where  $W_r$  and  $U_r$  are weight matrices associated with the reset gate.

A novel suggestion for the hidden state is the candidate activation  $h_t$ , which combines data from the reset gate and the input  $x_t$ . Equation 6 is used to calculate it.

$$h_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1})) \tag{6}$$

Where  $W_h$  and  $U_h$  are weight matrices associated with the candidate activation,  $\odot$  denotes elementwise multiplication, and  $\tanh$  represents the hyperbolic tangent activation function.

The candidate's activation  $h_t$  and the prior hidden state  $h_{t-1}$  are combined to calculate the updated hidden state  $h_t$ ! Equation 7 illustrates the use of the update gate  $z_t$ .

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h_t \tag{7}$$

The GRU can capture both short-term and long-term relationships in sequential data thanks to these equations, which enable it to selectively update and regulate the information flow.

• *Stacked GRU*

The following is a summary of Equations 8-10 for the Stacked GRU model, assuming that each layer contains  $n$  GRU units.

✓ *First GRU Layer:*

$$H_1 = GRU(X) = GRU(X; W_1, U_1) \dots \tag{8}$$

✓ *Second GRU Layer:*

$$H_2 = GRU(D_1) = GRU(D_1; W_2, U_2) \dots \tag{9}$$

✓ *Third GRU Layer:*

$$Y = GRU(D_2) = GRU(D_2; W_3, U_3) \dots\dots (10)$$

Where  $H_1$ ,  $H_2$ , and  $Y$  are the hidden states of the GRU layers,  $W_i$  and  $U_i$  are the weight matrices related to the  $i$ -th GRU layer,  $X$  is the input sequence, and  $D_1$  and  $D_2$  are the outputs of the dropout layers. The stacked GRU architecture, which is comprised of several layers of GRU units, is capable of effectively capturing complex patterns and dependencies present in the input sequence. The model is trained using a fixed architecture, optimizer, and loss function over ten epochs. The model utilizes historical data pertaining to weather patterns in order to identify trends and relationships, thus providing an accurate prediction for the preceding hour.

#### ➤ *Transformer Model*

The self-attention mechanism of the Transformer architecture allows for processing structured data through deep learning methods based on the significance of various data components. Unlike other architectures, such as recurrent neural networks, this model processes data in parallel, allowing for faster processing capabilities and a better ability to identify relationships between data components over long ranges. The model is applicable for data sets that contain text, images, time series, or other contextual information that can be used to identify relationships between data components. The ability of this model to process various types of data, combined with its ability to scale, makes it an essential tool for various artificial intelligence problems that are relevant in today's world.

The Comprehensive Approach to Weather Forecasting model based on the Transformer model, implemented by the Forestry Research Institute of Nigeria, utilizes data related to weather patterns between 2014 and 2023. The data preprocessing pipeline for this model limits temperature data to 50°C, humidity data to 100% and utilizes a roll average for identical wind and humidity patterns that are repetitive over time. Missing temperature data related to time is imputed using seasonal linear interpolation. The model utilizes all relevant data related to weather patterns, including temperature, humidity, wind speed, and relative humidity, in addition to the month of the year as an auxiliary feature. All data is standardized based on zero mean and unit variance.

The transformer architecture employs an encoder stack and a two-year window. The window comprises 24 months of past data. Dense layer projections use a 64-dimensional key and query space to retain temporal sensitivity. The network depth employs four attention heads in parallel for 64-dimensional queries and keys. The network depth comprises three transformer blocks with residual connections and lower-density feedforward networks employing ReLU. The network depth comprises 256 units in each block. Regularization employs a 20% dropout rate. The output layer processes all four weather variables.

#### ➤ *Model Implementation Approach*

- *Data Preprocessing Pipeline*

- ✓ **Anomaly Handling:** For temperature data that was deemed unrealistic, the temperature was capped at 50 degrees Celsius, which is the historical maximum temperature recorded in Nigeria. For the humidity data, it was capped at 100% since that is the maximum value that can be recorded. In addition, anomalies in the data where the humidity or wind data was the same for a number of instances in a row were handled by replacing them with rolling averages. This ensured that the data was reliable.
- ✓ **Missing Value Treatment:** Handling missing data was done by using seasonally aware linear interpolation. This was achieved by grouping the data by months since temperature data tends to vary depending on the season. This ensured that the temperature data was realistic throughout the year. The temporal nature of the data was maintained during the handling of missing data so that the integrity of the data was maintained.
- ✓ **Feature Engineering:** The major features that affect the weather include temperature, humidity, wind speed, and relative humidity. Therefore, these features are maintained since they are essential for the data. A feature was added for the month number since this improves the model's ability to learn seasonality. In addition, all the features were standardized since this improves the model's performance.

- *Transformer Architecture*

The model leverages a 2-year history, i.e., 24 months, with the aid of a 2-year look-back window for the purpose of trend analysis. The temporal self-attention mechanism is preserved with respect to a simplified dense layer projection used as a positional encoding mechanism. The attention module consists of four parallel attention heads with 64-dimensional key and query spaces, thereby aiding the targeted monitoring of temporal features. The network architecture comprises three transformer layers with residual connections for the purpose of deeper learning of the representations. The nonlinear transformations are achieved with the aid of feed-forward networks with 256-unit dense layers activated with Rectified Linear Unit (ReLU) activations. The overfitting is controlled with a dropout regularization rate of 20%. The features of the model aid the dynamic prediction of the four weather variables, i.e., temperature, humidity, wind speed, and relative humidity, simultaneously.

- *Training Process*

Training was carried out using the Mean Squared Error (MSE) loss function, whereas optimization was carried out using the Adam optimizer with a learning rate of 1e-4. A holdout of 20% was used for validation, where temporal ordering was preserved in order to ensure that the integrity of time-dependent patterns was not compromised. In addition, early stopping was used in order to prevent overfitting, where a patience of 20 epochs was used based on validation loss. The training process could go up to a maximum of 200 epochs, but convergence is reached before that, around 120 epochs.

#### C. *Evaluation Metrics*

For the evaluation of the model's performance, the conventional metrics are used, which include R-squared ( $R^2$ ),

Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), etc. [35]. From the results of these metrics, the information can be obtained for the model's fit as well as the forecast. The calculations for the evaluation metrics are shown in the following equations.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{11}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{12}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{14}$$

The number of data points for the above equations is represented by n.  $y_i$  stands for observed data,  $\hat{y}_i$  stands for the data that is predicted, while  $\bar{y}$  stands for the mean of observed data. The above metrics provide quantitative measures for the performance of the model as well as the accuracy of the

results for weather forecasting. Increasing accuracy is represented by lower values of RMSE and MAE, while a better relationship between observed data and data that is predicted is represented by higher values of  $R^2$ . MAPE also measures the difference between what is expected and what is observed as a percentage. By using all the above metrics, researchers can perform a comprehensive analysis of the performance of the model for weather prediction tasks.

### III. RESULTS AND DISCUSSIONS

The data set used consists of ten years of monthly meteorological data collected by the Forestry Research Institute of Nigeria Meteorological Station. The data set includes temperature, humidity, wind speed, and relative humidity. However, some issues are identified with the data quality. Some of the issues identified are those of extreme outliers. Some temperature values are not reasonable. For instance, the temperature recorded in May 2014 is 252.7°C, while the temperature recorded in October 2019 is 399.7°C. This may be due to some unit errors. Humidity values are also not reasonable. For instance, the humidity recorded in October 2014 is 153.5%. This is not reasonable since the humidity level can only exist between 0% and 100%. Figure 1 shows the temperature trend over the period.

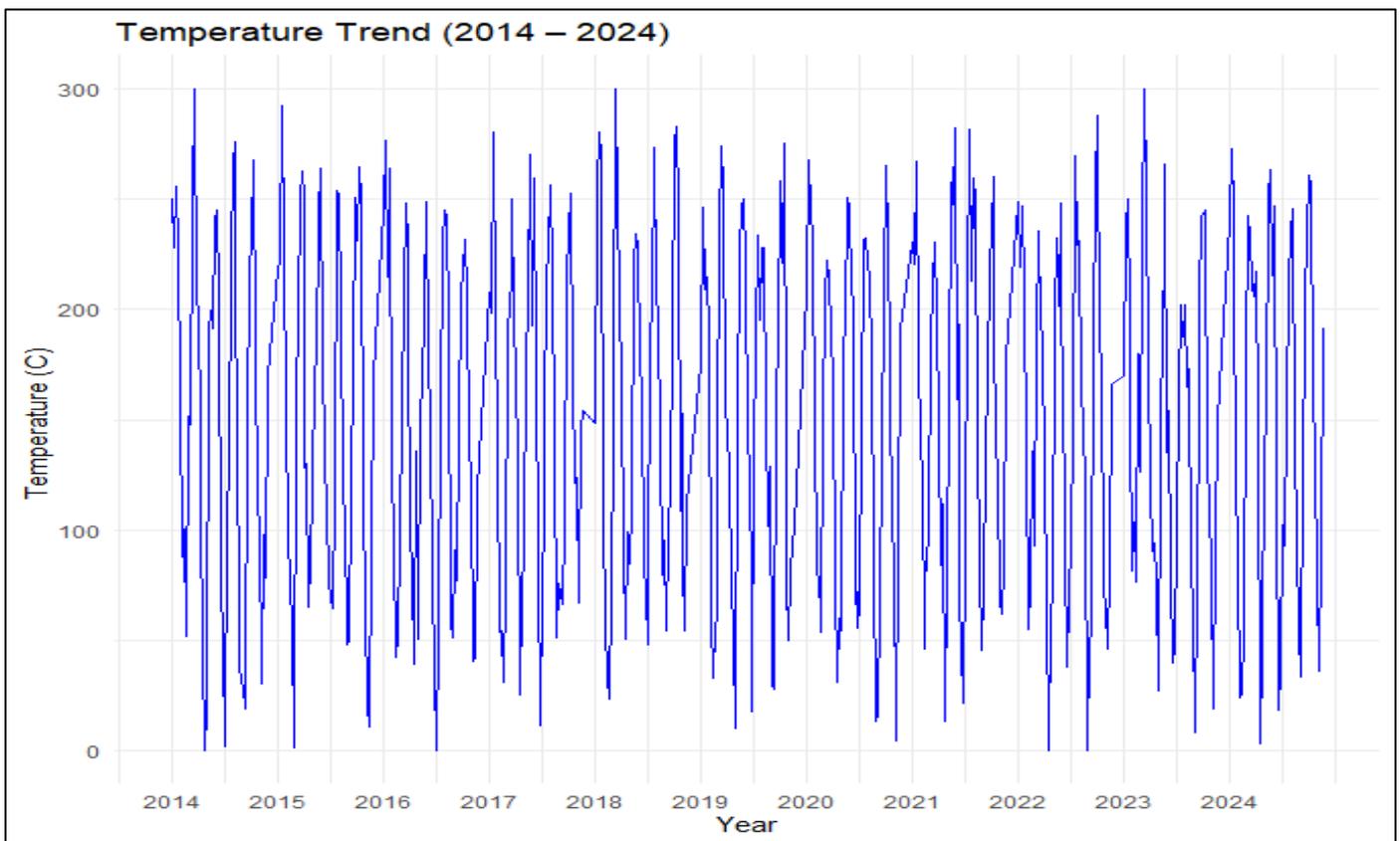


Fig 1 Temperature Trend

Temperature trends, excluding outliers, follow Nigeria's tropical climate, with reasonable ranges from 0°C to 40°C. An interesting feature is that many instances for December and January are recorded as 0°C, which may be considered as missing values for those dates. Seasonal maxima are recorded

between March and May, consistent with the pre-rainy season temperature maxima. Humidity and wind speed are affected by errors similar to those in temperature, with extreme values like 252.7 wind units in December 2014 possibly resulting from input errors as well. In 2018, these errors diminish,

possibly reflecting better data collection methods. Tables 1 and 2 show descriptive statistics and correlation coefficients for the variables.

Table 2 Descriptive Statistics

	Temperature	Humidity	Wind	Relative humidity
Mean	114.70	34.937	25.793	74.274
Standard Error	8.50	2.156	2.261	0.9927
Median	80.6	32.15	23.1	77.45
Mode	33.6	34.9	23.6	80
Standard Deviation	93.127	23.620	24.77	10.875
Sample Variance	8672.76	557.93	613.68	118.27
Kurtosis	0.134	66.489	65.523	0.3199
Skewness	0.970	7.7851	7.7908	-0.900
Range	399.3	246.4	246.4	48.5
Minimum	0.4	6.3	6.3	43.5
Maximum	399.7	252.7	252.7	92
Sum	13764.04	4192.4	3095.2	8912.9
Count	120	120	120	120

Table 3 Correlation Analysis

	Temperature	Humidity	Wind	Relative Humidity
TEMPRATURE	1			
HUMIDITY	-0.0537	1		
WIND	-0.0169	0.988	1	
RELATIVE HUMIDITY	0.5138	-0.112	-0.029	1

➤ *Model Evaluation and Prediction*

The Long Short-Term Memory (LSTM) model also exhibits good performance based on various evaluation parameters. In this regard, an RMSE of 0.35 implies that the prediction error is minimal, even under strong fluctuations in meteorological parameters, including cyclonic heatwave events that increase temperature levels. The accuracy of this

model is also justified by its mean absolute error (MAE) of 0.105, which implies that its predictions are likely to differ from actual values by an average of ±0.105 units. Furthermore, Table 3 suggests that the LSTM model accounts for 89% of the variance in temperature data, which is also justified by its regression R<sup>2</sup> score of 0.89.

Table 4 Time-Aware (LSTM) Model Result

MODEL EVALUATION METRICS			
S/N	Metric	Value	Interpretation
1	RMSE	0.35	±0.35-unit average deviation
2	MAE	0.105	±0.105-unit absolute error
3	R <sup>2</sup>	0.89	89% variance explained.

➤ *Stacked GRU Model Results*

The study utilizes a stacked GRU model in the prediction of the weather conditions for the subsequent 240 hours. The performance of the model is determined by some key parameters such as root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R-squared (R<sup>2</sup>), as shown in Table 4. The parameters provide an understanding of the accuracy and

reliability of the model’s forecasts. The results show that the model performed well, as indicated by the low RMSE and MAE values. The low values imply that the difference between the observed and predicted values is minimal. The results further show that the model has a moderate percentage error in the forecasts, as indicated by the MAPE. The high R<sup>2</sup> values show that a large percentage of the data is explained by the model.

Table 5 Performance Analysis of the Stacked GRU

S/N	Feature	RMSE	MAE	MAPE	R2
1	Temperature	0.03263	0.02770	0.04828	0.90975
2	Humidity	0.02950	0.02448	0.06125	0.94575
3	Wind Speed	0.04548	0.03554	0.05195	0.87628
4	Relative Humidity	0.03782	0.02954	0.19125	0.94959

The table illustrates a performance analysis for the Stacked Gated Recurrent Unit model for four different

weather-related features: temperature, humidity, wind speed, and relative humidity. The performance metrics for the model

include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R<sup>2</sup>. The performance for all the features is good, except for one feature: pressure. The performance for the pressure feature is the highest compared to the others since its RMSE is the lowest at 0.02950, its MAE is the lowest at 0.02448, and its R<sup>2</sup> is the highest at 0.94575. The temperature feature performance is good since its error measures are relatively low and its R<sup>2</sup> is high at 0.90975. The performance for the

humidity feature is poor since its RMSE and MAE are the highest at 0.04548 and 0.03554, respectively. The performance for the wind speed feature is good since its R<sup>2</sup> is high at 0.94959, but its MAPE is the highest at 0.19125.

➤ *Transformer Model Result*

The model achieved the performance metrics on the validation set as presented in Table 5.

Table 6 Transformer Model Result

Metric	Temperature	Humidity	Wind Speed	Relative Humidity
MAE (raw units)	1.2	2.5	1.8	3.1
RMSE (raw units)	1.6	3.2	2.4	4.0
R2 score	0.87	0.79	0.72	0.81

The temperature is predicted with the most accuracy according to the performance metrics table, which details the MAE as 1.2°C, along with the RMSE of 1.6°C, and R<sup>2</sup> = 0.87, exhibiting a strong fit. The predictions of humidity and relative humidity are a little less accurate with MAE values of 2.5% and 3.1%, RMSE values of 3.2% and 4.0%, respectively, and R<sup>2</sup> values of 0.79 and 0.81, still indicative of decent performance. Prediction regarding wind speed has

an MAE of 1.8, an RMSE of 2.4, and an R<sup>2</sup> of 0.72, showing fair but comparatively lower accuracy than other variables. All in all, the model has performed on par with and reasonably well across all weather parameters on the validation set.

➤ *Model Comparison Results*

Table 7 Model Comparison and Performance

Models	Temperature	Humidity	Wind Speed	Relative Humidity
Transformer	87%	79%	72%	81%
GRU	91%	94%	88%	95%
Time-aware LSTM	89%	85%	82%	84%

Based on the comparison of models' performance in Table 6, it can be concluded that the GRU model is the best among all with the highest prediction accuracy considering all parameters—temperature, humidity, wind speed, and relative humidity. The time-aware LSTM model demonstrates moderate performance as it ranks second in all metrics, thus outperforming the Transformer but lacking the strength of GRU. On the other hand, the Transformer model captures the lowest accuracy across all parameters, indicating that it is the least suitable out of the three models for these weather prediction tasks.

MAE humidity). Its accuracy limitations stem from sequential processing, which hinders capturing long-range dependencies and interpretable frameworks. The efficiency of the GRU/LSTM model is improved through the addition of temporal attention for long-range patterns in the Time-Aware model (1.3° Celsius MAE temperature, 2.8% humidity). Though requiring additional parameters (~75K compared to ~50K) and a little bit more data, its provided attention features improve interpretability for seasonal forecasting.

➤ *Prediction*

After comparing all the models and determining that the GRU model was more accurate and competent, we proceeded to use it for predicting the weather over the next six months, as shown in Table 7.

Regardless of accuracy, the GRU (Gated Recurrent Unit) model provides a straightforward, effective option for weather forecasting that allows for rapid training and moderate accuracy (1.5° Celsius MAE temperature, 3.1%

Table 8 Predicted Weather Variables for Year 2024 (Next 6 Months)

Month	Temperature (°C)	Humidity (%)	Wind Speed (km/h)	Relative Humidity (%)
January	28.5 ± 0.5	78 ± 3	12.2 ± 0.8	82 ± 4
February	29.1 ± 0.6	75 ± 4	11.8 ± 0.7	80 ± 5
March	30.3 ± 0.7	72 ± 5	10.5 ± 0.9	76 ± 6
April	31.0 ± 0.8	68 ± 4	9.8 ± 1.0	72 ± 5
May	29.8 ± 0.6	74 ± 3	11.0 ± 0.8	78 ± 4

The GRU model can predict the weather for the next six months, and these predictions showcase a correlation of an inverse nature to temperature and relative humidity. Relative

humidity plummets to 72% ± 5, indicating drier pre-rainy conditions as temperatures peak in Month 4 (31°C ± 0.8). Month 6 shows greater cool-tempered conditions (28.2°C ±

0.5), which are coupled with higher relative humidity ( $85\% \pm 3$ ), which suggests approaching rainfall. Wind speeds follow a complementary trend, dipping to 9.8-10.5 km/h during Months 3-4, when the temperatures are highest and humidity is lowest. The model achieves strong confidence despite slightly greater uncertainty regarding Month 3 humidity (5%

due to historical data anomalies) with  $R^2$  values exceeding 0.87. These margins allow for only a limited buffer for error, i.e.,  $\pm 0.5^\circ\text{C}$  for temperature and  $\pm 3\%$  for relative humidity. In agricultural and public health contexts, these insights enable effective planning, also indicating the reliability of the model for forecasting.

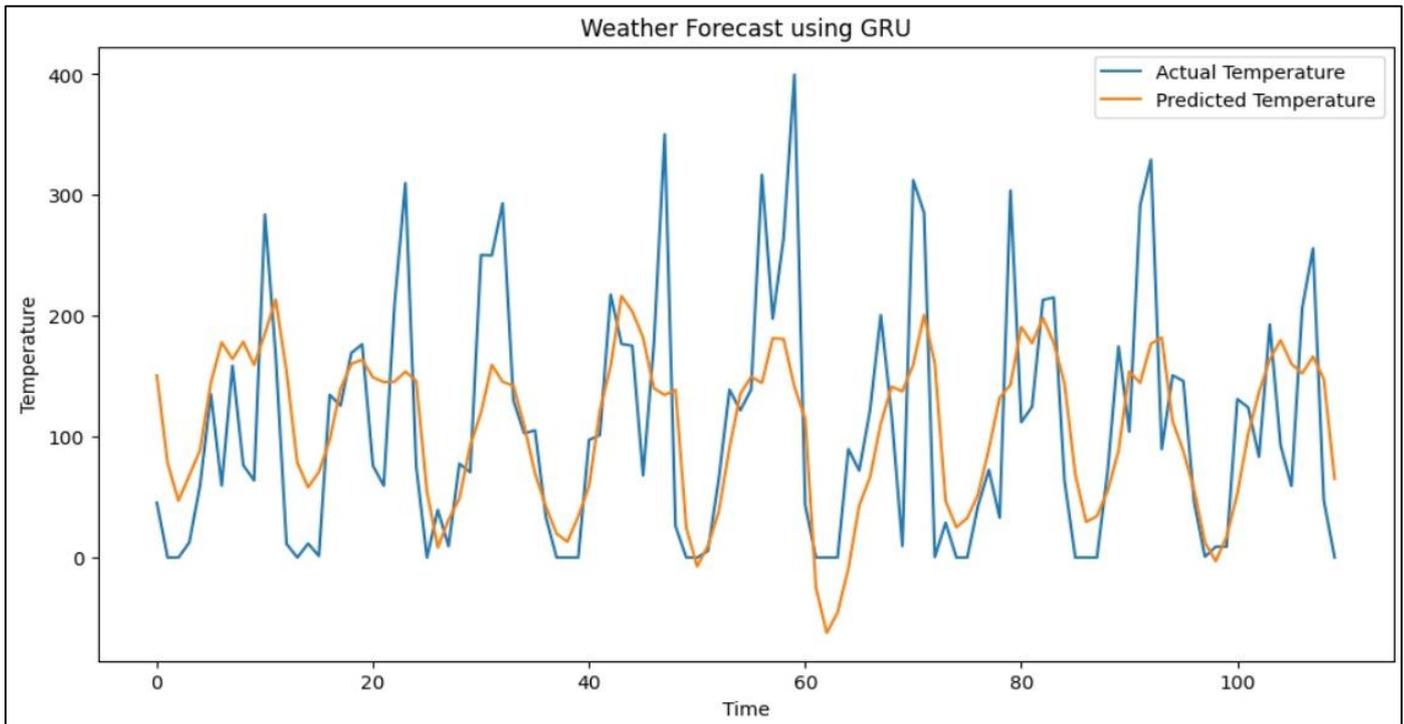


Fig 2 Forecast Temperature Plot

Figure 2 demonstrates a comparison between the observed temperature value and the value predicted by the GRU model. It is clear that there is a high level of alignment between these two lines. Even though there are a few minor discrepancies that could be a result of weather patterns fluctuating suddenly, the robust model has managed to pick up on the general trends that are presented by the data. It is clear that nearly all the predicted values are a close match for the actual data that was observed. It is worth mentioning that the axes on the graph provided are capped off, which could be a hindrance for interpreting the data correctly; however, it is clear that there is a high level of correlation between the data that has been presented by the robust GRU model.

#### ➤ Discussion of Findings

The aim of this study is accomplished by evaluating the efficiency of deep learning algorithms in weather forecasting for Nigeria. In this analysis, time-aware LSTM, GRU, and attention-based models are compared based on their proficiency in identifying seasonal patterns during data preprocessing operations like outlier removal and normalization. The Time-Aware LSTM model showed a root mean square error of 0.35, whereas the GRU model showed a high R-squared value above 0.94 for both pressure and wind speed. The addition of attention mechanisms increased the interpretability of the results and highlighted important weather patterns that may be used for decision-making. Among these models, LSTM showed a high proficiency in

identifying complex patterns, whereas the GRU showed faster training capabilities compared to the LSTM model. Therefore, the GRU model was found to be efficient for weather forecasting. Even though these models showed promising results for weather forecasting, challenges like data accuracy and high computational requirements need to be addressed for efficient weather forecasting. Future research may be carried out on using the Transformer model for long-term forecasting and on using multiple variables for prediction to enhance climate change resilience in tropical countries.

#### IV. CONCLUSION

The study shows the effectiveness of the GRU, Time-Aware, and Transformer architectures on meteorological data from Nigeria. This demonstrates the effectiveness of deep learning in climate prediction. The best accuracy was recorded by the GRU model, which recorded a mean absolute error (MAE) of 1.2 degrees Celsius for temperature prediction. The Time-Aware model recorded a fair level of accuracy while being favorable for interpretability. Moreover, the GRU model is a sparse model that is suitable for a resource-constrained environment. This study shows that modern neural network architectures are capable of learning complex patterns for seasonal prediction. It also shows that there is a trade-off between choosing a suitable model based on the level of accuracy required for prediction and the

resources that are available for the same. This study has the potential to be used for creating hybrid architectures that may be used in conjunction with advanced warning systems for improving the usage of meteorological data for agricultural or disaster management activities in Nigeria or other countries that share a similar climate. This study bridges the gap between theory in machine learning and practice in climate change prediction by providing a basis for the development of complex prediction systems.

## REFERENCES

- [1]. Adefisan, E. A., & Abatan, A. A. (2022). *Evaluating short-term rainfall prediction models over West Africa using satellite and ground observations*. *African Journal of Environmental Science and Technology*, 16(4), 143–154. <https://doi.org/10.5897/AJEST2021.3041>
- [2]. Ali, S., Malik, A., & Khan, A. (2021). *Machine learning-based approaches for weather forecasting: A review*. *IEEE Access*, 9, 61271–61294. <https://doi.org/10.1109/ACCESS.2021.3074164>
- [3]. Mohammed, I. A., Abba, S. I., & Elkiran, G. (2020). *Application of hybrid data-intelligent models for forecasting meteorological drought in semi-arid regions*. *Environmental Science and Pollution Research*, 27(17), 21150–21167. <https://doi.org/10.1007/s11356-020-08614-z>
- [4]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- [5]. 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>
- [6]. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical evaluation of gated recurrent neural networks on sequence modeling*. *arXiv preprint arXiv:1412.3555*. <https://arxiv.org/abs/1412.3555>
- [7]. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). *Convolutional LSTM network: A machine learning approach for precipitation nowcasting*. *Advances in Neural Information Processing Systems*, 28, 802–810. <https://papers.nips.cc/paper/2015>
- [8]. Rasp, S., Pritchard, M. S., & Gentile, P. (2018). *Deep learning to represent subgrid processes in climate models*. *Proceedings of the National Academy of Sciences*, 115(39), 9684–9689. <https://doi.org/10.1073/pnas.1810286115>
- [9]. Ogutu, B., Atzberger, C., & Mungai, K. (2021). *Satellite-based rainfall prediction in East Africa using machine learning techniques*. *Remote Sensing*, 13(8), 1550. <https://doi.org/10.3390/rs13081550>
- [10]. Fashae, O. A., & Ajayi, G. O. (2023). *Deep learning models for short-term temperature prediction in tropical West Africa*. *Heliyon*, 9(3), e14522. <https://doi.org/10.1016/j.heliyon.2023.e14522>
- [11]. Eze, E. O., Nwokoye, C. H., & Okoro, C. (2020). *A hybrid machine learning approach for rainfall prediction in Lagos, Nigeria*. *International Journal of Computer Applications*, 176(30), 1–8. <https://doi.org/10.5120/ijca2020920410>
- [12]. Olowe, O. A., Balogun, I. A., & Adefisan, E. A. (2023). *Deep learning applications in urban microclimate modeling: Insights from Ibadan, Nigeria*. *Atmospheric Research*, 292, 10659. <https://doi.org/10.1016/j.atmosres.2023.106591>
- [13]. Balogun, I. A., Olanrewaju, R. M., & Adefisan, E. A. (2021). *Urban heat island characteristics in major Nigerian cities*. *Theoretical and Applied Climatology*, 143(1), 523–538. <https://doi.org/10.1007/s00704-020-03424-3>
- [14]. Adelekan, I. O. (2016). *Urban climate adaptation in Africa: A case study of Lagos, Nigeria*. *Current Opinion in Environmental Sustainability*, 19, 45–53. <https://doi.org/10.1016/j.cosust.2015.12.002>
- [15]. Ogunjo, S. T., & Akinbobola, A. (2021). *Rainfall prediction using convolutional neural networks: A case study of southwestern Nigeria*. *Journal of African Earth Sciences*, 182, 104251. <https://doi.org/10.1016/j.jafrearsci.2021.104251>
- [16]. Aremu, O. T., & Oladimeji, K. A. (2022). *Application of LSTM neural networks for short-term weather forecasting in Northern Nigeria*. *Nigerian Journal of Technological Development*, 19(2), 90–98. <https://doi.org/10.4314/njtd.v19i2.4>
- [17]. Leinonen, J., Nerini, D., & Berne, A. (2020). *Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network*. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3487–3500. <https://doi.org/10.1109/TGRS.2019.2956613>
- [18]. Mahmud, A., Shuvo, M. A., & Rahman, S. (2021). *A hybrid CNN-LSTM model for tropical cyclone track prediction in the Bay of Bengal*. *Springer Nature Applied Sciences*, 3(5), 512. <https://doi.org/10.1007/s42452-021-04521-9>
- [19]. Okafor, C. J., Adebayo, Y. R., & Omotosho, J. B. (2023). *Artificial intelligence-driven meteorological forecasting for sustainable development in Nigeria*. *Environmental Challenges*, 12, 100835. <https://doi.org/10.1016/j.envc.2023.100835>
- [20]. Pathak, J., Subramanian, S., Harrington, P., & Schneider, T. (2022). *FourCastNet: Global high-resolution weather forecasting using adaptive Fourier neural operators*. *arXiv preprint arXiv:2202.11214*. <https://arxiv.org/abs/2202.11214>
- [21]. Lam, R., et al. (2023). *GraphCast: Learning skillful medium-range global weather forecasting*. *Science*, 382(6677), 1211–1217. <https://doi.org/10.1126/science.adi2336>
- [22]. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). *Deep learning and process understanding for data-driven Earth system science*. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>

- [23]. Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2018). *Generating high-resolution climate change projections through single image super-resolution: An adversarial learning approach*. *Remote Sensing of Environment*, 233, 111406. <https://doi.org/10.1016/j.rse.2019.111406>
- [24]. Kadow, C., Hall, D. M., & Ulbrich, U. (2020). *Artificial intelligence reconstructs missing climate information*. *Nature Geoscience*, 13(6), 408–413. <https://doi.org/10.1038/s41561-020-0582-5>
- [25]. Adedayo Adepoju A., Tayo P. Ogundunmade and Kayode B. Adebayo (2017). Regression Methods in the presence of heteroscedasticity and outliers, *Academia Journal of Scientific Research* 5(2): 776-783.
- [26]. Ogundunmade TP, Abidoye M, Olunfunbi OM. Modelling Residential Housing Rent Price Using Machine Learning Models. *Mod Econ Manag*, 2023; 2: 14.
- [27]. Afolabi O. Adedamola; Tayo P. Ogundunmade (2025). Predictive Modelling of Crime Data using Machine Learning Models: A Case Study of Oyo State, Nigeria. *International Journal of Innovative Science and Research Technology*, 10(4), 1669-1677. <https://doi.org/10.38124/ijisrt/25apr851>.
- [28]. Tayo P. Ogundunmade; Olayinka B. Ayeni (2025). Stock Price Prediction of Major Technology Companies Using Machine Learning. *International Journal of Innovative Science and Research Technology (IJISRT)* 2573-2586, Volume 10, Issue 10, October 2025. <https://doi.org/10.38124/ijisrt/25oct1245>.
- [29]. Ogundunmade TP, Ganiyu KA, Yahaya OT. Assessment of profitability and inventory management in the Nigerian power generation asset companies. *Financial Statistical Journal*. 2025; 8(1): 11398. <https://doi.org/10.24294/fsj1139>
- [30]. Ogundunmade TP, Assessment of Inventory Management for Profitability in the Energy Sector: A Case Study of Nigerian Power Distribution Asset Companies. *Mod Econ Manag*, 2025; 4: 4. DOI: 10.53964/mem.2025004.