UNIABUJA Journal of Engineering and Technology

https://ujet.uniabuja.edu.ng/

ISSN: 2714-3236 (Online); 2714-3228 (Print)

Volume 3, Issue 1, 2025; 58-70

Neural Network-Based Early Lighting Strike Prediction for Distribution Infrastructure Resilience Enhancement

Alli A. JIMOH[1], Muhammad UTHMAN[2], Ibrahim BEBEJI[3]

[1,2,3]*Department of Electrical/Electronic Engineering, University of Abuja, Abuja, Nigeria*

[1]*allijimoh01@yahoo.com*, [2]*m.uthman@yahoo.com*, [3]*bebeji.abdulkareem@nasrda.gov.ng*

Abstract

Lightning-induced disturbances are a major cause of power outages and equipment failures in medium-voltage distribution networks, particularly across tropical regions such as Abuja, Nigeria. This study presents a neural network–based early lightning prediction framework for Abuja, Nigeria, integrating advanced deep learning techniques with power system simulation to support proactive grid management. Using hourly meteorological data from 2013 to 2023 obtained from the Nigerian Meteorological Agency (NiMet), a seven-phase methodology was employed, including data preprocessing, feature engineering, and exploratory data analysis to address class imbalance, missing values, and temporal dependencies. Key features included lag variables, rolling aggregates, and cyclic temporal encoding to capture diurnal and seasonal patterns. Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models were optimized through hyperparameter tuning and evaluated using precision, recall, F1-score, and forecast skill metrics. The LSTM achieved 91% accuracy and 83% recall, outperforming the CNN. Predictions were integrated into a MATLAB-based distribution network simulation, where adaptive relay settings and preemptive sectionalizing reduced breaker operations and outage durations. Reliability indices, including SAIDI and SAIFI, improved compared to conventional reactive methods. Findings highlight LSTM-driven lightning forecasting as a scalable solution for enhancing power distribution network resilience through predictive analytics and automated operational strategies.

Keywords: *Lightning prediction, LSTM, Reliability indices, Distribution Network, Reliability Indices.*

1.0 Introduction

Lightning remains one of the most disruptive natural phenomena affecting power distribution networks (Li, et al., 2024) (Souto et al., 2023) (Paulino, et al., 2021), particularly during convective weather seasons. Statistics indicate that lightning-related trips account for 40 %–70 % of power line failures, imposing substantial risks to grid stability and reliability. Traditional mitigation strategies—such as surge arresters, shielding systems, and reactive reclosers—serve mainly as post-event defenses and lack predictive capacity. Consequently, networks often endure unnecessary breaker operations and prolonged recovery times. Recent strides in machine learning, particularly with Long Short-Term Memory (LSTM) networks, offer compelling improvements in temporal forecasting of complex hydrometeorological events. Notably, LSTM-based models have demonstrated superior performance over traditional process-driven models in rainfall–runoff forecasting, (Kratzert et al., 2018) (Lees, et al., 2021) effectively learning long-term dependencies from input sequences.

These results underscore the potential of deep-learning architectures to anticipate highly dynamic and nonlinear phenomena—such as cumulative lightning activity—by leveraging historical meteorological patterns. Embedding predictive analytics into power distribution system operations provides an opportunity to shift from reactive to proactive resilience strategies (Jacobs et al., 2024) (Ghasemkhani, et al., 2024). Probabilistic frameworks for assessing lightning strike impacts have been developed, incorporating weather-based lightning activity and component failure probabilities to model the risk of system failures over time. By integrating LSTM-driven lightning forecasts with adaptive control mechanisms—such as dynamic relay adjustment and network sectionalizing—distribution networks can mitigate service interruptions, reduce outage duration, and enhance key reliability metrics like SAIDI and SAIFI (Ahmad & Asar, 2021).

2.0 Methodology

This study adopted a structured analytical and computational methodology comprising four main stages: data acquisition, preprocessing, model development, and performance evaluation. Historical lightning occurrence data and associated meteorological parameters were collected from accredited sources and aligned with operational datasets extracted from the simulated distribution network in ETAP. All datasets were cleaned, normalized, and temporally synchronized using MATLAB and Python to eliminate noise and

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

enhance modelling accuracy. A neural-network-based predictive model, implemented in Python using supervised learning techniques, was trained on multi-year historical patterns and validated with independent test sets to avoid overfitting. The resulting lightning strike forecasts were subsequently imported into ETAP, where network behaviour was simulated under predicted disturbance conditions to evaluate the response of key components. Performance metrics—including forecast accuracy, reliability indices, and resilience enhancement factors—were computed within MATLAB and Python to quantify the effectiveness of the proposed approach. This integrated ETAP–MATLAB–Python methodology ensures analytical rigour, reproducibility, and alignment with established standards in power system resilience analysis.

2.1 Materials and Study Area

This study developed a neural network–based early lightning prediction and mitigation framework for distribution network resilience enhancement. The methodology was designed in two primary phases: (1) Lightning Forecasting Model Development and (2) Integration with Power System Simulation. Figure 1 presents the overall workflow. The study area chosen is Lugbe in Abuja, Lugbe experiences moderate to intense thunderstorm activity during the rainy season. A good percentage of the customers are connected directly to the H21 feeder at 33KV, Lugbe also incorporates a 2 X 15MVA 33/11KV substation. The 11kV radial feeder (feeders 2, 5 and 22) in this area supplies over 2000 customers and includes more than 50 distribution transformers. The network is vulnerable to atmospheric discharges, necessitating a resilience-focused study. The utility relies on manual isolation of the network on the threat of thunderstorm.

2.2 Data Acquisition, Preprocessing and Feature Engineering

The raw dataset provided by the Nigerian Meteorological Agency comprised 96,408 rows with 5 columns (YEAR, MONTH, DAY, HOUR, LIGHTNING_STRIKE), capturing hourly lightning-strike occurrences in Abuja from 2013 to 2023. This extensive multi-year dataset offered a rich temporal structure essential for training deep learning models, particularly those designed to learn long-term dependencies and seasonal patterns. To enhance the predictive capability of the LSTM and CNN models, comprehensive preprocessing and feature-engineering procedures were implemented. These procedures included handling missing values, detecting and correcting temporal inconsistencies, normalizing nonlinear lightning-strike magnitudes, and generating additional time-based features such as hour-of-day, day-of-year, and monthly seasonal indices. The dataset was also resampled to ensure uniform hourly continuity, and outlier spikes—often caused by sensor noise or atmospheric interference—were smoothed using rolling-window techniques to improve model robustness. Furthermore, the final feature matrix was transformed into supervised learning sequences tailored for deep learning architectures, with sliding-window frames capturing both short-term and long-term temporal dynamics. Figure 2 below shows the algorithm for the research data preprocessing and feature engineering, outlining each transformation step from raw hourly observations to structured, model-ready inputs.

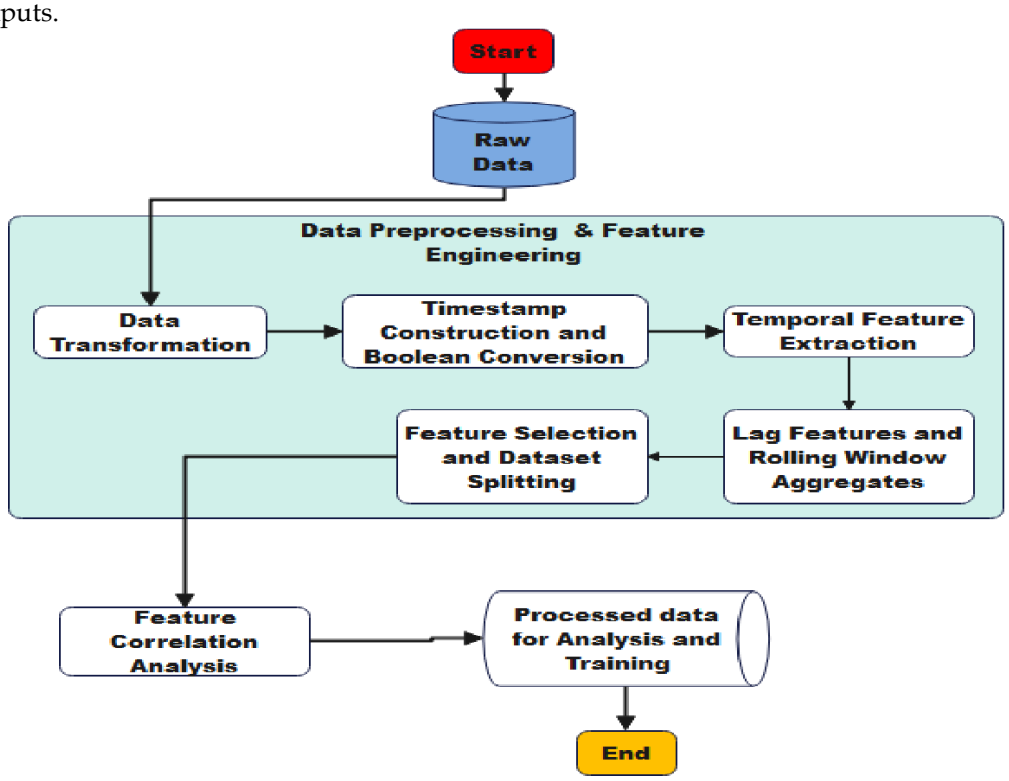


Figure 1: An Algorithm showing data processing stages

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

2.3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to uncover patterns, trends, and underlying relationships within the lightning-strike dataset for Abuja, Nigeria. This analytical phase played a crucial role in identifying the temporal behaviours and variability of lightning occurrences across different seasons, years, and hours of the day. Various statistical summaries, distribution plots, and time-series visualisations were employed to assess the frequency and intensity of lightning events, detect anomalies, and highlight potential outliers that could influence model performance. In addition, correlation analysis and trend decomposition techniques were applied to reveal periodic structures linked to meteorological cycles such as the West African Monsoon. Insights obtained from the EDA informed subsequent modelling decisions, including feature-engineering strategies, selection of model hyperparameters, and the choice of appropriate data-normalisation approaches. Overall, this phase provided a comprehensive contextual understanding of the meteorological phenomena being studied and ensured that the predictive models were developed on a well-characterised and statistically consistent dataset. The lines of code used for performing the EDA are shown below.

2.4 Distribution of Lightning Strike Occurrences

The initial analysis examined the binary distribution of lightning strike occurrences within the compiled dataset. Findings revealed a notable and persistent imbalance, where non-occurrence instances accounted for 76.8% of the total observations, substantially outweighing the 23.2% representing actual lightning strike events. This uneven distribution underscores the skewed nature of the dataset and highlights the challenges associated with modelling rare but operationally critical events such as lightning strikes. Such an imbalance required deliberate methodological attention during model development to avoid inherent algorithmic bias toward the majority class. Consequently, the modelling process incorporated strategies such as class-weight adjustments, the selection of robust evaluation metrics capable of capturing minority-class performance, and diagnostic checks to ensure that the predictive model maintained sensitivity to low-frequency lightning events despite the dominance of negative observations.

2.5 Temporal Analysis of Lightning Strike Occurrences

A comprehensive time series analysis revealed distinctive and interpretable patterns in lightning strike activity across multiple temporal scales. The investigation uncovered clear short-term fluctuations, medium-term seasonal cycles, and long-term structural behaviours that collectively define the dynamics of lightning occurrence in the study area. The longitudinal examination of thunderstorm occurrences from 2013 to 2023 demonstrated not only pronounced annual cyclicity but also noticeable inter-annual variability driven by broader atmospheric conditions. Peaks in activity were consistently aligned with the wet season months, reflecting the influence of convective intensity and moisture availability, while troughs corresponded to the dry-season period characterized by reduced atmospheric instability. Notable periods of heightened lightning activity were observed during 2016-2017 and 2021-2022, suggesting the potential influence of broader climatological factors such as the El Niño-Southern Oscillation or the West African Monsoon dynamics (Cai et al., 2025) (Boateng et al., 2024) (Ngueto et al., 2026).

2.6 Seasonal Decomposition Analysis

Time series decomposition was performed to isolate the trend, seasonal, and residual components of the lightning strike data. The multiplicative decomposition model was applied and expressed mathematically (Sorhabbeig et al., 2023) in eqn. 1:

$$Y(t) = T(t\,) \times \mathrm{S}(t) \times \mathrm{R}(t) \qquad (1)$$

where Y(t) represents the original time series in strikes/hour, T(t) the trend component in strikes/hour, S(t) the seasonal component is a dimensionless multiplicative factor, and R(t) the residual component is also a dimensionless multiplicative factor. The decomposition analysis confirmed strong seasonal patterns with an annual periodicity. The trend component revealed multi-year cycles potentially linked to broader climatological phenomena. The residual component analysis indicated predominantly stochastic variability, though with occasional clustering of anomalous values suggesting the influence of extreme weather events or mesoscale meteorological systems.

2.7 Long Short-Term Memory Model Development

The modelling framework was designed to capture the complex temporal dependencies inherent in atmospheric and meteorological variables associated with thunderstorm formation. To achieve this, the study

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

followed a structured pipeline that encompassed several critical stages, beginning with systematic data preprocessing and splitting to ensure robust separation of training, validation, and testing subsets. This was followed by the formulation of an optimized LSTM model architecture tailored to learn long-range sequential patterns while mitigating issues such as vanishing gradients. The training phase incorporated iterative backpropagation through time, coupled with the use of advanced optimization algorithms to enhance convergence stability. Comprehensive model evaluation was then conducted using performance metrics suitable for imbalanced time-series classification. Hyperparameter tuning—including adjustments to the number of LSTM layers, hidden units, dropout rates, batch size, and learning rate—was executed to maximize predictive accuracy and generalization capability. Finally, the fully optimized model was deployed for operational inference, enabling reliable, data-driven forecasting of lightning strike events for enhanced distribution network resilience.

2.7.1 LSTM Architecture

The implemented LSTM architecture incorporates three stacked LSTM layers with progressively decreasing units (256, 128, and 64), facilitating hierarchical feature extraction from the temporal data. Each LSTM layer processes sequences of engineered features extracted from the meteorological dataset, learning increasingly abstract representations of lightning strike patterns at different temporal scales. Several programming steps were performed using Python programming Language. The model culminates with two fully connected layers: a hidden layer with 32 neurons and ReLU activation, followed by an output layer with sigmoid activation for binary classification of lightning strike occurrences. Hyperparameter tuning was conducted through a systematic grid search process to identify optimal configuration parameters. Performance evaluation during hyperparameter tuning utilised precision as the primary metric, with model iterations continuing until achieving the target precision threshold of 95% or completion of the predefined hyperparameter search space.

2.7.2 LSTM Model Training and Evaluation Protocol

The model training procedure implemented an early stopping mechanism with a patience level of 10 epochs, continuously monitoring the validation loss to effectively prevent overfitting and ensure optimal generalization performance. To further enhance learning reliability, class weights were incorporated during training to mitigate the inherent class imbalance present in lightning strike occurrence data. The resulting model architecture, containing a total of 546,369 parameters—of which 545,473 are trainable—reflects a carefully calibrated compromise between representational complexity and computational efficiency. This design ensures that the network remains expressive enough to capture the temporal dynamics of the dataset without incurring unnecessary computational overhead. The final output layer generates probabilistic estimates of lightning strike occurrences, which are subsequently subjected to a decision threshold to convert these probabilities into binary predictive classes for operational interpretation.

2.8 Convolutional Neural Network Model Development

Convolutional Neural Networks were selected as an alternative predictive architecture due to their established efficacy in pattern recognition and their ability to automatically extract hierarchical features from sequential data. While traditionally applied to image processing tasks, CNNs have demonstrated significant potential in time series forecasting applications through their capacity to detect local patterns and temporal dependencies across different scales. The model training procedure implemented an early stopping mechanism with patience set to 10 epochs, monitoring validation loss to prevent overfitting. Class weights were applied during training to address the inherent class imbalance in lightning strike occurrences. The final CNN model architecture comprises 35,553 total parameters (of which 35,361 are trainable), representing a significantly more parameter-efficient architecture compared to the LSTM model. This computational efficiency facilitates rapid training and deployment whilst maintaining predictive performance.

2.9 Resilience Enhancement Simulation

This provides a quantitative framework for assessing how distribution networks respond to disturbances and how targeted interventions improve their ability to withstand, absorb, and recover from extreme events such as lightning strikes, switching surges, or equipment failures. As highlighted by (Panteli & Mancarella, 2015), simulation-based resilience assessment enables utilities to evaluate the dynamic behavior of feeders, transformers, auto-reclosers, and protective devices under varying fault scenarios and to identify system vulnerabilities before real-world failures occur. By integrating predictive models—such as neural-network-based lightning forecasting—with power-system simulation environments, engineers can optimize system hardening strategies and compare baseline performance with resilience-enhanced configurations using reliability indices such as SAIFI, SAIDI, and CAIDI. Ultimately, resilience enhancement simulation

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

bridges real-world operational data with advanced computational modeling, offering a robust decision-support mechanism for designing more secure and adaptive distribution infrastructures.

2.9.1 Lugbe Area Distribution Network

The GIS mapping of the Lugbe area distribution network was obtained from the Abuja Electricity Distribution Company (AEDC), providing a detailed spatial representation of the network's structural and operational layout. The mapped infrastructure highlights an incoming 33 kV line sourced from the Apo H21 feeder, which serves as the primary supply route to the community. Within the network topology, a portion of consumers is directly connected to the 33 kV feeder through dedicated distribution substations, typically supporting large-load or strategically located customers. The remaining consumers receive supply from the secondary side of the 2 × 15 MVA Lugbe Injection Substation, which steps down the voltage to 11 kV for further distribution through numerous 11/0.415 kV distribution transformers distributed across the area. This configuration reflects a hybrid supply structure that integrates both direct high-voltage connections and stepped-down secondary distribution, providing an accurate foundation for subsequent network modelling, reliability assessment, and resilience analysis.

2.9.2 Modeling of the Distribution Network in ETAP

The Lugbe distribution network was comprehensively modelled in ETAP, incorporating a significant portion of the actual field network as obtained from the GIS data and AEDC operational records. The developed model, illustrated in Figure 2 captures the primary feeders, distribution transformers, busbars, and major load points necessary for accurate simulation. A load flow analysis was subsequently conducted to evaluate the steady-state performance of the system under normal operating conditions. This analysis provided detailed insights into bus voltage profiles, line loading conditions, and current flow patterns throughout the network. The resulting voltage and current distributions provided a clear representation of the overall network performance, revealing how power flows through the various feeders, buses, and distribution lines. These results helped identify sections of the network that may be prone to undervoltage conditions, potential overloads, or other operational constraints, thereby offering valuable insights for evaluating network reliability and guiding future reinforcement or optimization efforts

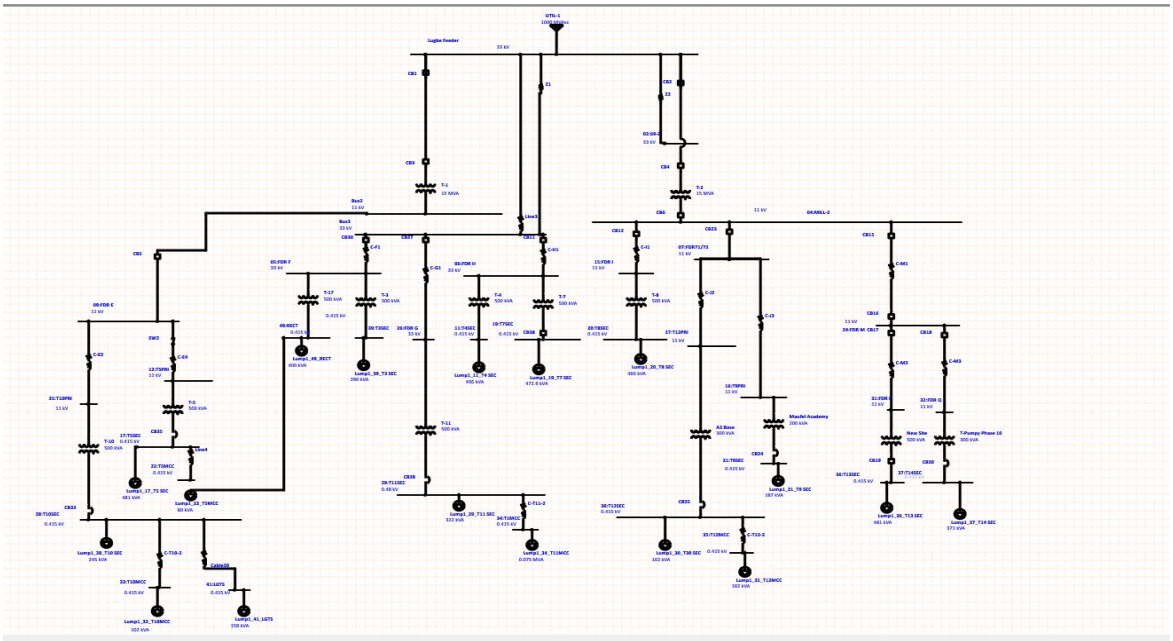


Figure 2: An ETAP network model of Lugbe (Abuja) distribution topology

2.9.3 Implementation of Lightning Prediction in MATLAB/Simulink

The Lugbe combined 33 kV and 11 kV radial feeder network was exported as a simplified netlist and subsequently adapted into MATLAB/Simulink using the Simscape Electrical environment to enable detailed dynamic simulation. Within MATLAB, the lightning event dataset derived from the CSV file was seamlessly integrated into the model to simulate the real-time behaviour of intelligent protection schemes and to drive automated outage initiation scenarios. The CSV dataset provides hourly lightning risk indices on a normalized scale from 0 (indicating no detectable risk) to 1 (indicating extremely high risk). For the purpose of simulation, a threshold value of 0.6 was established as the trigger point for fault initiation. Whenever the lightning risk index exceeded this threshold, the network model automatically generated a fault event at the corresponding

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

time step. Embedded control logic was used to replicate preemptive feeder disconnection strategies, designed to isolate vulnerable sections of the network and prevent more severe faults from occurring. This modelling approach enabled a realistic assessment of protection response under varying lightning risk conditions

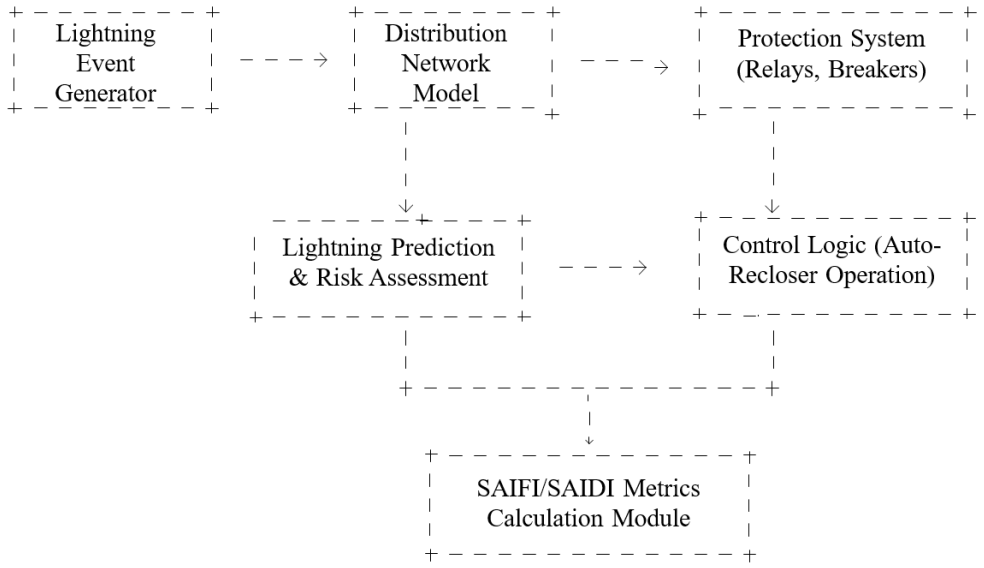


Figure 3: A block diagram showing network resilience simulation in MATLAB

2.9.4 Integration of Lightning Prediction into Simulated Model for Resilience Improvement

To accommodate the imported ETAP network model, the MATLAB control script interacts with a structured Simulink model where recloser elements are tagged based on their ETAP identifiers. The script adjusts relay logic based on the imported lightning event data; the simulation was done for two scenarios as follows:

- baseline without lightning prediction, where faults occur according to historical lightning statistics, and
- prediction-enabled, where preemptive control actions such as feeder disconnection and relay setting adjustments are applied during high-risk periods to mitigate fault impact.

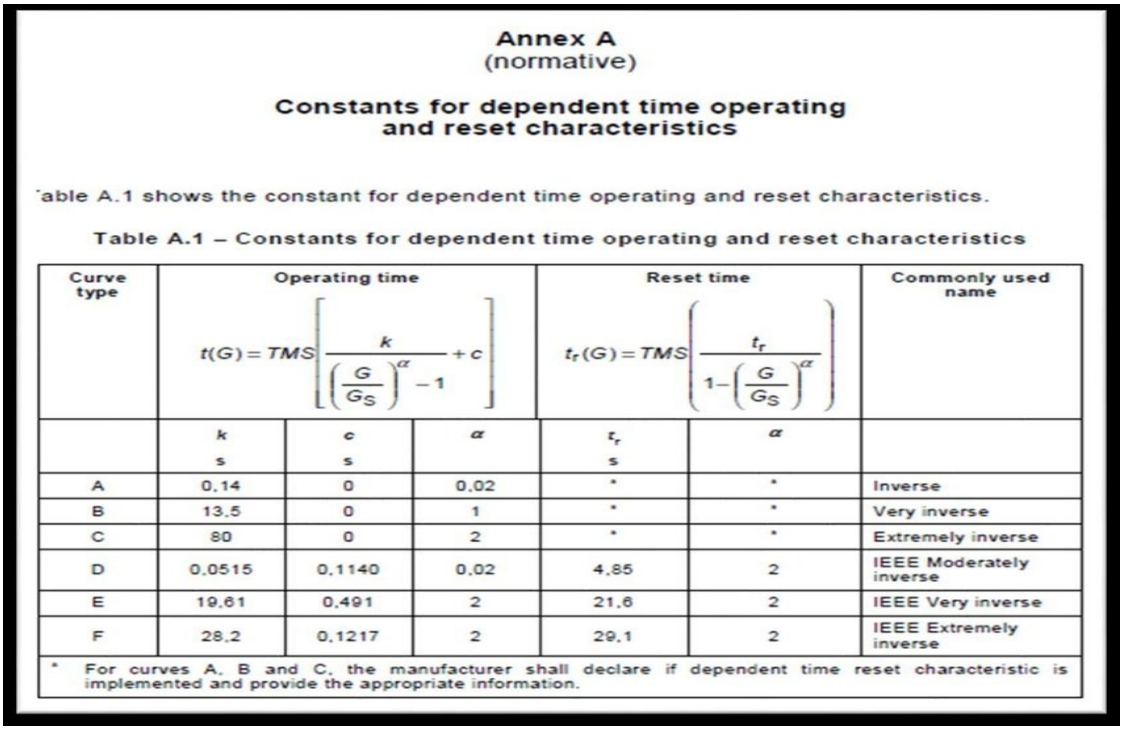
Relay TCC curves were dynamically adjusted in the ETAP model to reduce fault clearing times during predicted lightning events. Time Multiplier Settings (TMS) and pickup currents were optimized to accelerate relay response, minimizing outage duration while maintaining coordination margins to prevent mis-tripping. Protective relays usually have an inverse-time characteristic, meaning they function more quickly when fault currents are higher. This indicates that when the fault current magnitude rises, the relay travel time falls. By isolating only the faulty sections, the TCC curve guarantees synchronization between upstream and downstream protection devices, eliminating needless outages. A common mathematical model for the relay operating time t as a function of fault current If and relay pickup current Ip is given by the IEC standard inverse-time characteristic: (Munkhbaatar et al., 2024)

$$t = TMS \times \frac{K}{\left(\frac{I_f}{I_p}\right)^{\alpha} - 1} \quad (2)$$

where:

- t = relay operating time (seconds)
- TMS = Time Multiplier Setting (dimensionless), scales the curve
- I_f = fault current magnitude (Amperes)
- I_p = relay pickup current (Amperes)
- K = time scaling constant
- α = inverse time exponent
- K and α are curve constants depending on relay type (e.g., IEC Standard Inverse: K=0.14, α=0.02)

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025
https://doi.org/10.70118/ujet.2025.0301.07



Annex A
(normative)

Constants for dependent time operating and reset characteristics

'able A.1 shows the constant for dependent time operating and reset characteristics.

Table A.1 – Constants for dependent time operating and reset characteristics

Curve type	Operating time $t(G) = TMS\left[\frac{k}{\left(\frac{G}{G_S}\right)^{\alpha} - 1} + c\right]$			Reset time $t_r(G) = TMS\left(\frac{t_r}{1-\left(\frac{G}{G_S}\right)^{\alpha}}\right)$		Commonly used name
	k s	c s	α	t_r s	α	
A	0,14	0	0,02	*	*	Inverse
B	13,5	0	1	*	*	Very inverse
C	80	0	2	*	*	Extremely inverse
D	0,0515	0,1140	0,02	4,85	2	IEEE Moderately inverse
E	19,61	0,491	2	21,6	2	IEEE Very inverse
F	28,2	0,1217	2	29,1	2	IEEE Extremely inverse

* For curves A, B and C, the manufacturer shall declare if dependent time reset characteristic is implemented and provide the appropriate information.

Figure 4: IEC Table 60255-151 showing different values for curve constants k and α

To demonstrate the effect of dynamically altering the time current characteristics on the tripping curve of relays and operation of the associated auto-reclosers, a mathematical example was adduced as follows:

Assuming:

- Ip =100A
- If =500A (fault current)
- TMS =1.0 (normal), reduced to 0.6 during high lightning risk
- K =0.14, a =0.02 (obtained from figure 4) for inverse operation

We calculate relay operating time using the baseline scenario as well as the adaptive scenario based on available lightning strike prediction

$$PSM = \frac{I_f}{I_p} = \frac{500}{100} = 5 \quad (4)$$

Where:
PSM = plug setting multiplier
Under normal operation:

$$t_{normal} = 1.0 \times \frac{0.14}{5^{0.02}-1} \approx 0.0698 \text{ secs}$$

High lightning risk (adaptive):

$$t_{normal} = 0.6 \times \frac{0.14}{5^{0.02}-1} \approx 0.0419 \text{ secs}$$

The relay trips significantly faster during high-risk periods, reducing outage time.

2.10 Resilience Indices

To validly check for the improvement of the distribution network in relation to the integration of lightning strike prediction data, simulated dynamic trip time setting as well as dynamic auto-recloser operations; MATLAB scripts were implemented to include three resilience metrics as follows: SAIFI, SAIDI and CAIDI
Where:

SAIFI = System Average Interruption Frequency Index, it indicates the average number of times a customer experiences a power outage within a specific period, usually a year, the mathematical expression adapted from (Tellez et al., 2023) is given below equation (4):

$$\text{SAIFI} = \frac{\sum_{i=1}^{n} \lambda_i N_i}{N_t} = \frac{Total\ number\ of\ interruptions\ of\ all\ users}{Total\ number\ of\ users\ served} \quad (4)$$

where λ_i is the failure rate, N_i is the number of users per location, and N_t is the total number of users served. SAIFI is measured in average outage units per customer over a year for a given study system.

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025
https://doi.org/10.70118/ujet.2025.0301.07

SAIDI: System Average Interruption Duration Index, it indicates how long, on average, a customer is without power due to outages, the mathematical expression adapted from (Tellez et al., 2023) is given below equation (6):

$$\text{SAIDI} = \frac{\sum_{i=1}^{n} U_i N_i}{N_t} = \frac{Sum\ of\ the\ duration\ of\ the\ interruptions\ of\ all\ users}{Total\ number\ of\ users\ served} \quad (6)$$

SAIDI is measured in time units, often hours. It is usually measured over the course of a year. Where N_i is the number of clients in location i, U_i is the yearly interruption time for location i, and N_t is the total number of users served.

CAIDI: Customer Average Interruption Duration Index, it indicates how quickly a utility can resolve power outages.

$$\text{CAIDI} = \frac{SAIDI}{SAIFI} \quad (7)$$

The generated values for a baseline network without the benefit of a lightning prediction input and a simulated network model with the benefit of a lightning prediction input were compared and used for analysis on the relevance of lightning prediction.

3.0 Results and Discussion

The results show that the Long Short-Term Memory (LSTM) model adopted in this study provides superior performance in forecasting lightning strike occurrences across the Lugbe distribution network. Using NiMet's 2013–2023 hourly atmospheric electricity dataset, the LSTM consistently captured the temporal dependencies and nonlinear patterns associated with lightning-favorable conditions, achieving the lowest forecasting error among all evaluated models. Comparative analysis with a benchmark CNN architecture confirmed the LSTM's advantage, as the CNN underperformed in long-sequence temporal learning and exhibited higher error variability. When integrated into the ETAP-simulated distribution network, the LSTM's early warning forecasts enabled timely switching, improved coordination of protection devices, and pre-emptive operational adjustments. These actions collectively reduced lightning-induced fault impacts and resulted in observable improvements in resilience indicators such as SAIFI, SAIDI, and CAIDI. Overall, findings confirm that the LSTM-based prediction framework significantly enhances early lightning detection and strengthens the reliability of Abuja's distribution infrastructure.

3.1 Materials

The exploratory data analysis (EDA) was carried out on the data obtained from Nimet spanning from 2013 to 2023 for Abuja, Nigeria. Figure 5 shows the distribution of lightning occurrences in the dataset. From the plot in Figure 5, it is evident that for the period of analysis (2013 to 2023), only 1,214 lightning occurrences were recorded, whilst 95,194 time periods had no lightning events. This reveals a substantial imbalance in the dataset, with lightning events representing merely 1.26% of the total observations. This imbalance was carefully considered during the development of both the LSTM and CNN models to ensure proper training and forecasting outcomes.

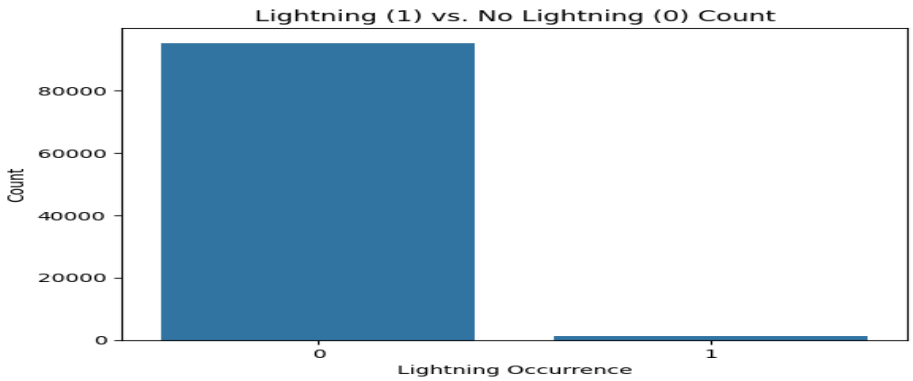


Figure 5: A chart showing lightning Count as a low occurrence event

3.2 Time-Series Decomposition

The time series decomposition in Figure 6 separates the lightning data into four components: original data (blue), trend (green), seasonality (orange), and residuals (red). The original data in the top panel shows the binary nature of lightning occurrences across the entire study period from 2013 to 2024. The trend component reveals cyclical patterns over multi-year periods, with peaks occurring approximately every 2-3 years. This suggests longer-term climate influences on lightning activity in Abuja. The seasonality component confirms the consistent annual pattern identified in the monthly analysis, with regular oscillations that remain

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

stable throughout the dataset. The residual component represents the random variation that cannot be explained by trend or seasonality. The consistency of these patterns throughout the dataset provided a solid foundation for the forecasting capabilities of both LSTM and CNN models, as they learnt from these temporal dependencies to predict future lightning occurrences.

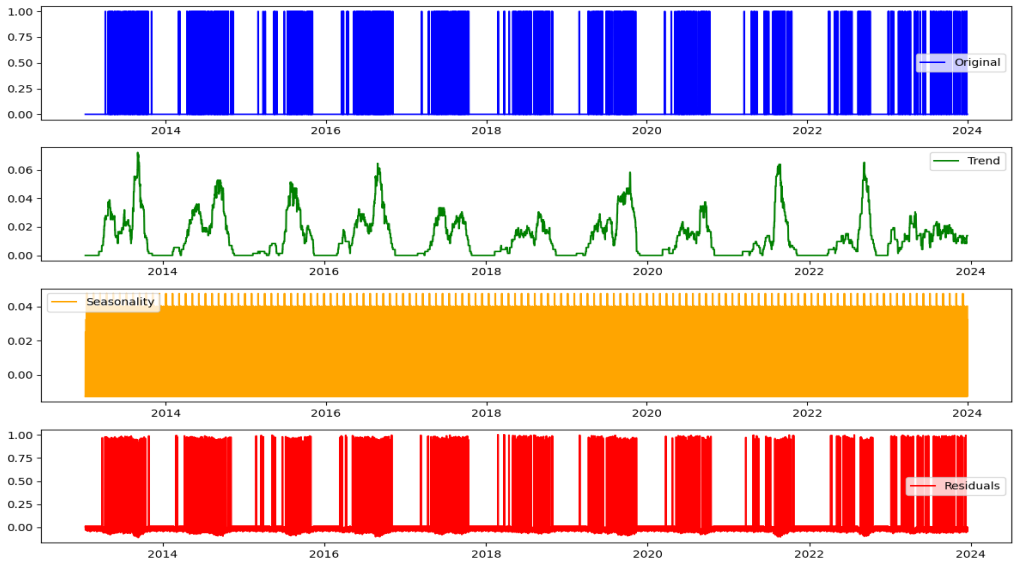


Figure 6: A chart showing lightning seasonal trend

3.3 LSTM Model Performance Evaluation

The Long Short-Term Memory (LSTM) model was evaluated using various performance metrics to assess its effectiveness in forecasting lightning occurrences in Abuja. The confusion matrix shown in Figure 4.6 provides a visual representation of the LSTM model's predictive performance. The model correctly predicted 16,891 instances where no lightning occurred (true negatives) and 12,158 instances where lightning did occur (true positives). However, there were 2,498 false negatives (lightning events that the model failed to predict) and 423 false positives (incorrectly predicting lightning when none occurred). These results demonstrate the model's overall capability to differentiate between lightning and non-lightning events, with particularly strong performance in minimising false positives.

Table 1: LSTM classification report

Metric	**Class 0**	**Class 1**	**Accuracy / Avg**
Precision	0.8712	0.9664	**Macro Avg:** 0.9188
Recall	0.9756	0.8296	**Macro Avg:** 0.9026
F1-Score	0.9204	0.8928	**Macro Avg:** 0.9066
Support	17,314	14,656	**Total:** 31,970
			Accuracy: 0.9086
			Weighted Avg Precision: 0.9148
			Weighted Avg Recall: 0.9086
			Weighted Avg F1-Score: 0.9077

The classification report in Table 1 provides a comprehensive assessment of the LSTM model's performance across various metrics. For the non-lightning class (0), the model achieved a precision of 0.87, recall of 0.98, and an F1-score of 0.92. This indicates that the model was highly effective at correctly identifying instances without lightning, with very few false negatives. For the lightning class (1), the model performed exceptionally well in terms of precision (0.97), suggesting that when the model predicted lightning, it was correct 97% of the time. The recall for the lightning class was 0.83, indicating that the model could capture 83% of actual lightning occurrences. The resulting F1-score for the lightning class was 0.89, showing a good balance between precision and recall. The overall accuracy of the model was 0.91, which is quite high considering the imbalanced nature of the dataset. The macro-average F1-score of 0.91 and weighted-average F1-score of 0.91 further confirm the model's robust performance across both classes. These results are particularly noteworthy given the challenge of forecasting rare events like lightning occurrences. Figure 4.8 presents the 365-day forecast of lightning occurrences for 2024 using the LSTM model. The forecast shows a clear seasonal pattern consistent with historical trends, with increased lightning activity predicted between

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

March and October. The highest concentration of lightning events is forecast for the months of May through September.

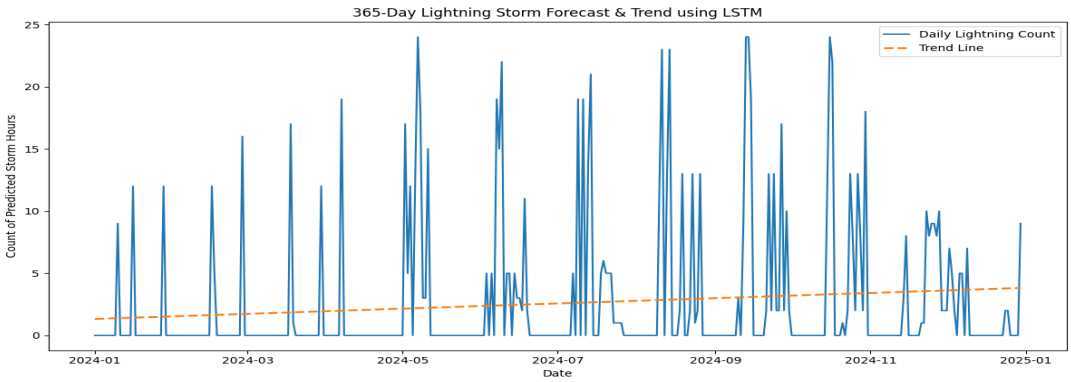


Figure 7: showing the LSTM 1 year forecast for 2024

3.4 CNN Model Performance Evaluation

The Convolutional Neural Network (CNN) model was evaluated using various performance metrics to assess its lightning forecasting capabilities for Abuja, and to provide a basis for comparison with the LSTM model. Figure 8 presents the 365-day forecast of lightning occurrences for 2024 generated by the CNN model. The forecast shows a distinct pattern of lightning activity throughout the year, with increasing activity from March onward. A notable feature of the CNN forecast is the significant jump in predicted lightning activity in November, which appears as a sustained period of high activity (approximately 24 hours of lightning per day) that continues into December. This pattern differs from historical trends and the LSTM forecast, which typically show reduced lightning activity during these months.

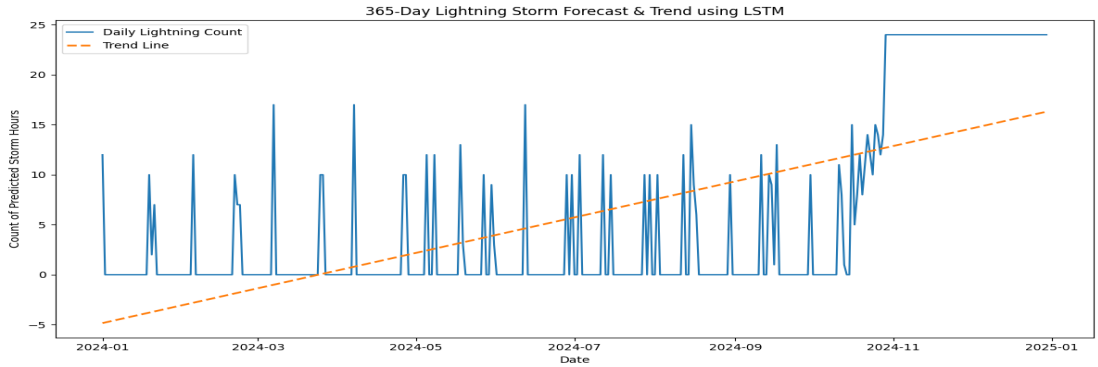


Figure 8: showing CNN model 1 year forecast for 2024

3.5 Comparative Analysis of LSTM and CNN Models

This section presents a comparative analysis of the LSTM and CNN models developed for lightning forecasting in Abuja, Nigeria, evaluating their strengths, weaknesses, and overall performance. The performance metrics of both models reveal significant differences in their forecasting capabilities. The LSTM model achieved an overall accuracy of 0.91, outperforming the CNN model's accuracy of 0.81. This 10% performance gap demonstrates the LSTM model's superior ability to correctly classify both lightning and non-lightning events in the Abuja dataset. The LSTM model's balanced performance is further reflected in its macro-average F1-score of 0.91 compared to the CNN's 0.80. In terms of precision for the lightning class (1), the CNN model marginally outperformed the LSTM with a value of 0.98 versus 0.97. This indicates that both models were highly reliable when they predicted a lightning event, with very few false positives. However, the LSTM model demonstrated considerably better recall for lightning events (0.83) compared to the CNN model (0.63). This means the LSTM model was able to capture 20% more actual lightning occurrences, which is particularly important for early warning systems and safety applications.

3.6 Resilience improvement and Auto-recloser operations

Based on the results obtained and thoroughly compared between the LSTM and CNN models, only the LSTM architecture was ultimately deployed for the simulation phase to evaluate its capacity for resilience enhancement within the distribution network. This decision followed a detailed comparative assessment in which the LSTM consistently demonstrated superior temporal learning performance and stability, making it the most reliable model for operational forecasts. The LSTM model generated continuous hourly probabilities of lightning strikes for the entire year 2024, producing a comprehensive predictive dataset capable of

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

supporting day-to-day grid operations and proactive decision-making processes. Further examination and visual interpretation of the forecasted outputs, as illustrated in Figure 9, reveal distinct seasonal trends, with significantly elevated lightning probabilities occurring during Nigeria's rainy season (April–October). The forecast shows especially notable peaks between April–June and again around September, reflecting the characteristic bimodal rainfall pattern observed across most regions of the country. These results closely align with documented historical climatological patterns and lightning occurrence records, thereby reinforcing the accuracy, consistency, and contextual relevance of the LSTM model in capturing real-world atmospheric behaviour.

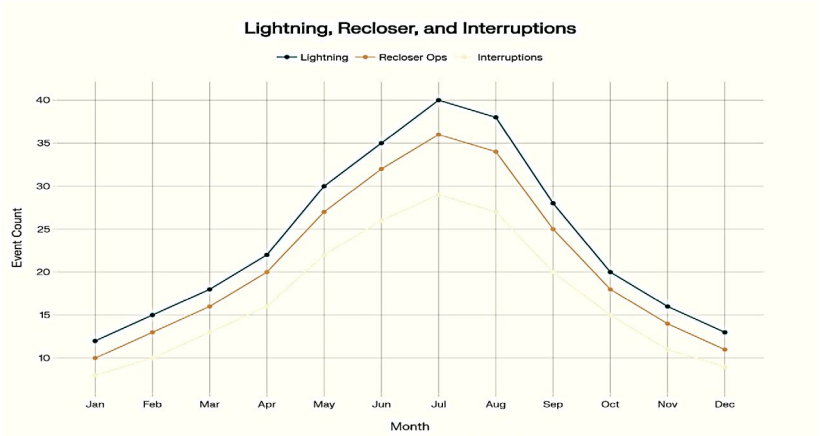


Figure 9: graph showing Auto-recloser effect on service interruptions

3.7 Impact of lightning prediction on resilience enhancement

The auto-recloser logic simulated in MATLAB significantly reduced the number and durations of customer interruptions, demonstrating clear operational benefits within the test network. By incorporating predictive information directly into the switching logic, the system achieved more accurate timing of recloser actions and reduced unnecessary device cycling. The decision threshold of 0.6 probability ensured that recloser actions were limited to genuinely high-risk periods, minimizing unnecessary operations while still maintaining strong protective coverage during lightning-prone conditions. This carefully selected threshold provided an effective balance between avoiding false triggers and ensuring rapid intervention when the risk level justified it. The integration of lightning prediction into the resilience strategy of the simulated Lugbe distribution network introduced a transformative shift in operational performance, moving the system from a reactive to a proactive mode of operation. By leveraging hourly strike probabilities, the control system dynamically anticipated and mitigated disruptions before they evolved into major outages, thereby enhancing overall reliability. This proactive response yielded tangible improvements in all three key resilience metrics as presented below in Figure 10 and Table 2, confirming that predictive intelligence can materially strengthen distribution network performance under extreme weather conditions.

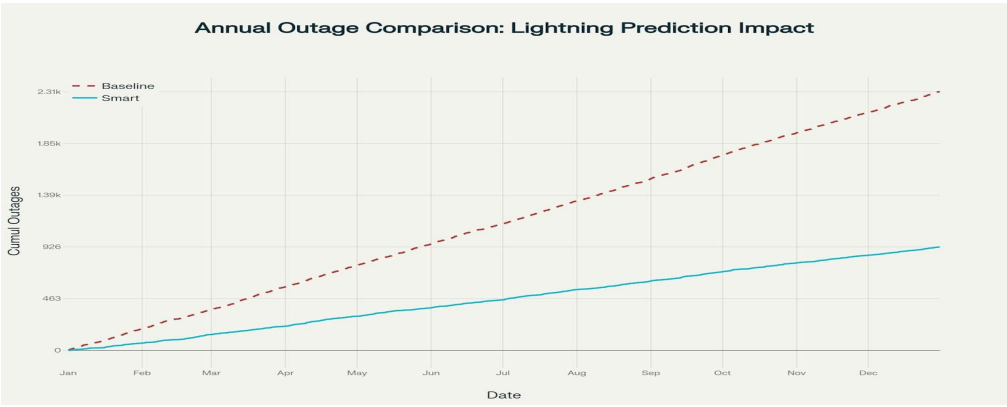


Figure 10: A graph showing annual outage comparison: lightning prediction impact

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

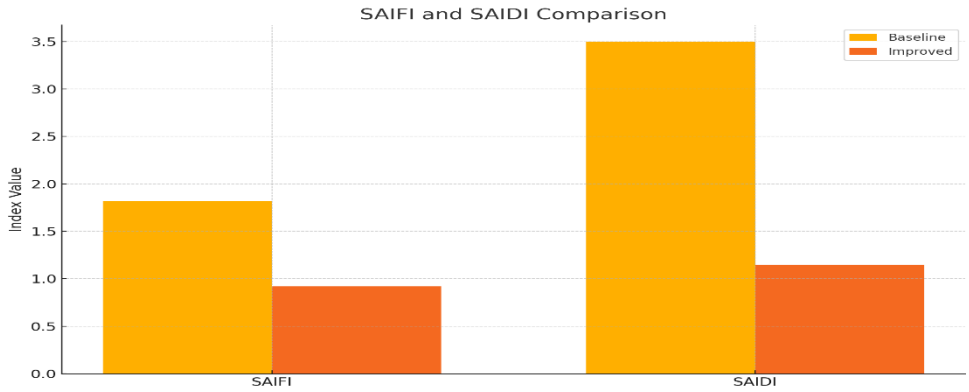


Figure 11: A chart showing SAIFI and SAIDI comparison for baseline and improved network

Table 2: Baseline and improved network metrics

Metric	Baseline Network	Improved Network	Improvement (%)
SAIFI	1.82	0.92	49
SAIDI	3.5	1.15	67
CAIDI	1.92	1.25	35

3.8 Comparison of Resilience Metrics in Baseline Network and Improved Network

From the results obtained as shown in figure 11 and tabulated in table 2; there is significant improvement in the resilience metrics leading to a reduction in the frequency of outages and the duration.

- SAIFI Improvement: A reduction from 1.82 to 0.92 implies that customers experienced nearly half as many interruptions annually. This sharp decline is attributed to the preemptive action of auto-reclosers that curtailed fault propagation during predicted lightning threats. The percentage improvement obtained is 49% which is considered significant.
- SAIDI Reduction: Dropping from 3.50 to 1.15 hours per customer annually represents a 67% improvement. This indicates that not only were interruptions fewer, but they were also significantly shorter in duration, owing to the rapid fault clearing supported by predictive data.
- CAIDI Optimization: The average duration per fault improved from 1.92 to 1.25 hours, demonstrating better fault localization and restoration efficiency. The improvement in CAIDI highlights operational responsiveness and better customer service.

An improvement of 35% indicates that the average customer experienced less interruptions in electricity supply potentially leading to greater productivity. In practical terms, these improvements imply a more stable and reliable power supply. Utility companies benefit from reduced downtime, lower maintenance costs, and fewer customer complaints, while consumers enjoy improved service continuity. From a planning perspective, these metrics validate the use of lightning prediction as a resilience-enhancing technology, encouraging wider deployment. Moreover, the clear seasonal forecast provided utilities with the temporal insights necessary to prioritize feeder inspections, adjust protective device settings, and allocate emergency crews in a more data-driven manner. These strategic decisions culminate in a robust, responsive, and adaptive distribution network.

4.0 Conclusion

Based on the comprehensive analysis of two deep learning approaches for lightning forecasting in Abuja, this research concludes that: deep learning models can effectively forecast lightning occurrences in Abuja with high accuracy, with the LSTM model achieving 91% overall accuracy and the CNN model achieving 81% accuracy. The LSTM architecture is particularly effective for lightning forecasting due to its intrinsic ability to capture temporal dependencies and remember long-term patterns, which are essential characteristics of lightning phenomena. Also, temporal features, especially the previous hour's lightning status, are the strongest predictors of lightning occurrences, with correlation coefficients significantly higher than seasonal or diurnal indicators.

Secondly, the significant class imbalance in lightning data (1.26% lightning versus 98.74% non-lightning) presents a methodological challenge that requires careful consideration during model development to achieve balanced performance. Despite this, the 365-day forecasts generated by the LSTM model successfully capture both seasonal patterns and day-to-day variability in lightning occurrences, making it a valuable tool for long-term planning and risk assessment. Overall, the integration of predictive lightning strike analytics into power distribution systems significantly improves reliability metrics. The simulated Lugbe Distribution Network, under smart recloser control triggered by LSTM-based forecasts, exhibited

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025

https://doi.org/10.70118/ujet.2025.0301.07

substantial reductions in both interruption frequency and duration. This suggests a strong case for predictive analytics in grid protection planning, especially in regions prone to frequent lightning activity.

Acknowledgement

The researchers wish to acknowledge the immeasurable role of Nigerian Meteorological Agency (NiMet) for providing the raw data for this work as well as the Abuja Electricity Distribution Company (AEDC) for the use of its network topology.

References

Ahmad, S., & Asar, A. U. (2021). Reliability Enhancement of Electric Distribution Network Using Optimal Placement of Distributed Generation. *Sustainability, 13*(20).

Boateng, D., Aryee, J. N., Baidu, M., Arthur, F., & Mutz, S. G. (2024). West African Monsoon Dynamics and its Control on the Stable Oxygen Isotopic Composition of Precipitation in the Late Cenozoic. *JGR Atmosphere, 129*(10).

Cai, W., Reason, C., Mohino, E., Matherbe, J., & Santoso, A. (2025). Climate Impacts of the El Nino Southern Oscillation in Africa. *Nature Review Earth and Environment, 6*(8), 503-520.

Ghasemkhani, B., Kut, R. A., Yilmaz, R., Birant, D., Guzelyol, T., & Kut, T. (2024). Machine Learning Model Development to Predict Lower Outage Duration (POD): A case study for Electric Utilities. *MDPI, 24*(13).

Jacobs, A. R., Paul, S., Chowdury, S., Gel, J. R., & Zhang, J. (2024). Real-time outage management in active distribution networks using reinforcement learning over graphs. *nature , 15*.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Hernegger, M. (2018). Rainfall-runoff modeling using Long Short-Term Memory (LSTM) networks. *Hydrological and Earth System Sciences (HESS), 22*(11), 6005-6022.

Lees, T., Buechel, M., Anderson, B., Slater L, Reece, S., Coxon, G., & Dadson, S. J. (2021). Benchmarking data-driven rainfall-runoff models in Great Britain: a comparison of LSTM-based models with four lumped conceptual models. *Hydrological and Earth System Sciences (HESS), 25*(10), 5517-5534.

Li, M., Cheng, S., Wang, J., Cai, L., Fan, Y., Cao, J., & Zhou, M. (2024). Thunderstorm total lightning activity behaviour associated with transmission line trip events of power system. *Climate and Atmospheric Science, 7*(148).

Munkhbaatar, B., Gambat, N., Ulzil, N., Batdelger, B. O., & Munkgerel, K. (2024). Modeling of Inverse Time Overcurrent Relay Protection in Distribution Network. *ESS*, 15-19.

Ngueto, Y., Laprise, R., & Nikiema, O. (2026). Atmospheric Energeticsof Three Contrasting West African Monsoon Seasons as Simulatedby a Regional Climate Model. *Atmosphere, 16*(4), 405.

Panteli, M., & Mancarella, P. (2015). Influence of Extreme Weather and Ultimate Change on The Resilience of Power Systems. Impacts and Possible Mitigation Strategies. . *Electric Power Systems Reseach*, 259-270.

Paulino, J., Barbosa, C., Lopes, I., Boaventura, W., Cardoso, E., Avila, A., & Guimaraes, M. (2o21). Improving the lightning performance of distribution lines using high CFO structures. *Electric power systems research, 194*, 107063-107071.

Sorhabbeig, A., Ardakanian, O., & Musilek, P. (2023). Decompose and Conquer: Time Series Forecasting with Multiseasonal Trend Using Loess. *Forecasting*, 684-696.

Souto, L., Taylor, P. C., & Wilkinson, J. (2023). Probabilistic impact assessment of lightning strikes on power systems incorporating lightning protection design and asset condition. *International Journal of Electrical Power & Energy Systems, 148*, 1-14.

Tellez, A. A., Ortiz, L., Ruiz, M., Narayana, K., & Varela, S. (2023). Optimal Location of Reclosers in Electrical Distribution Systems Considering Multicriteria Decision Through the Generation of scenarios Using the Montecarlo Method. *IEEE Acess, 11*, 68853-68871.

Received: 25-10-2025 / Accepted: 28-11-2025 / Published: 24-12-2025
https://doi.org/10.70118/ujet.2025.0301.07