**Machine Learning-Based Surface Refractivity Prediction in Coastal and Inland Regions of West Africa: A Comparative Study.**

Onawumi K.C.[1] Adediji A. T.[1] and Ogunjo .S. T[1]

[1]*Department of Physics, the Federal University of Technology, Akure, Ondo State.*

*Corresponding author: onawunmikenny@gmail.com

This conference paper investigates the application of machine learning models for predicting surface radio refractivity across diverse climatic zones in West Africa, focusing on four key locations: Kano, Freetown, Accra, and Niamey. Accurate radio refractivity prediction is critical for optimizing telecommunications and radar systems, as it directly affects radio wave propagation. Using the ERA5 reanalysis dataset, surface refractivity was computed from key atmospheric parameters such as surface net solar radiation, wind speed, precipitation, and potential evaporation. Three machine learning models; LightGBM, Random Forest, and Gated Recurrent Unit (GRU), were trained and optimized using grid search and early stopping techniques. LightGBM achieved the lowest errors across all locations, with an MSE of 1635.56 and MAE of 32.47, outperforming Random Forest and GRU by approximately 4% and 67%, respectively, particularly in semi-arid and stable regions, while GRU underperformed in all cases. The study identifies the most influential atmospheric parameters and highlights the superior performance of LightGBM for refractivity prediction across the region's coastal and inland zones.
**Key words**: *Surface radio refractivity, Machine learning, Temperature, Pressure, Relative humidity, Surface net solar radiation, Wind speed, Precipitation, and Potential evaporation*

## 1. INTRODUCTION

Accurate prediction of radio refractivity is critical for optimizing telecommunications systems, particularly in West Africa, where climatic variability presents significant challenges. Radio refractivity, which affects how radio waves propagate through the atmosphere, is influenced by atmospheric conditions such as temperature, humidity, and pressure (Brewster et al., 2020). Traditional empirical methods for predicting refractivity often fall short in capturing the complexities of these atmospheric interactions (Hsu et al., 2006). Machine learning (ML) has emerged as a promising alternative, offering the ability to model non-linear relationships between atmospheric variables (Bouma et al., 2021). This paper evaluates the performance of three ML models; ***LightGBM, Random Forest,*** and ***GRU,*** in predicting surface refractivity across four key locations in West Africa: ***Kano, Freetown, Accra,*** and ***Niamey.*** These models were chosen for


I. Adimula et al. (eds.), *Proceedings of the 8th URSI-NG Annual Conference (URSI-NG 2024)*, Advances in Physics Research 12,
https://doi.org/10.2991/978-94-6463-644-4_21

their strengths in handling large datasets, complex data interactions, and temporal patterns. By comparing these models across both coastal and inland regions, this study aims to identify the most effective approach for surface refractivity prediction in diverse climatic zones.

### 1.1. Objective:

- Investigate the performance of machine learning models (LightGBM, Random Forest, and GRU) in predicting surface refractivity across diverse climatic regions of West Africa: Kano, Freetown, Accra, and Niamey.
- Highlight the best-performing models and most relevant atmospheric parameters across these varied zones.

## 2. SITE DESCRIPTION AND SCOPE OF DATA

**2.1. Study Locations**: Each location is characterized by distinct atmospheric and meteorological patterns that significantly influence refractivity.

**Table 1: Selected Locations**

| CITY | COUNTRY | REGION |
|---|---|---|
| **KANO** | Nigeria | Inland semi-arid (Stable) |
| **FREETOWN** | Sierra Leone | Tropical rainforest (Coastal) |
| **ACCRA** | Ghana | Tropical savanna (Coastal) |
| **NIAMEY** | Niger | Inland Sahelian (Stable) |

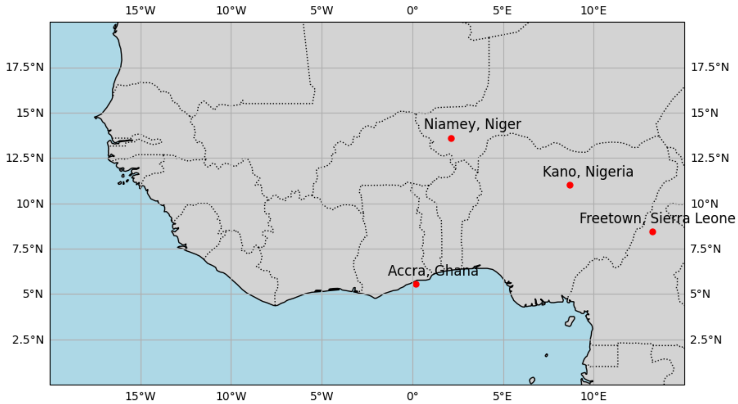



**Figure 1: Map showing the selected study locations across West Africa**

**2.2. Data Collection and Pre-processing:** The study utilizes the ERA5 reanalysis dataset, which provides comprehensive atmospheric data including temperature, humidity, pressure, and wind speed, using data from 2010 to 2020. These parameters are critical for calculating surface refractivity using the formula:

$$N = 77.6 \times \frac{P}{T} + 3.73 \times 10^5 \times \frac{e}{T^2} \quad (1)$$

Where
- P is the atmospheric pressure (hPa)
- T is the temperature (K)
- e is the water vapor pressure (hPa)

## 3. METHODOLOGY

The ERA5 dataset underwent pre-processing, including handling missing values using linear interpolation, and scaling atmospheric parameters to improve model convergence. Data was then stratified by season to assess model robustness across West Africa's diverse climatic zones.

**3.1 Model Selection and Training:** Three machine learning models were selected based on their unique attributes and capabilities: **LightGBM** for its efficiency with large datasets, **Random Forest** for its ability to handle non-linear relationships, and **GRU** for its suitability in sequential data processing.

- **LightGBM**: A gradient-boosting model optimized for speed and high performance, especially effective with large datasets and high-dimensional features. LightGBM efficiently captures complex interactions between features, making it suitable for atmospheric prediction tasks.
- **Random Forest**: An ensemble learning method that constructs multiple decision trees and aggregates their outputs to improve prediction accuracy. This model is particularly well-suited for capturing non-linear relationships, which are common in atmospheric data.
- **GRU**: A recurrent neural network (RNN) architecture with a simplified structure and lower computational cost compared to traditional RNNs. GRU is ideal for handling time-series data, making it effective for predictions in cases with seasonal or temporal dependencies, such as refractivity variations.

**3.2 Hyper parameter tuning**: Tuning was performed using ***grid search***, optimizing parameters like the number of trees for Random Forest, learning rate for LightGBM, and the number of LSTM units. The dataset was split into ***80% training and 20% testing***. Each model was optimized using grid search for hyper parameter tuning and early stopping to prevent over fitting.

**3.3 Evaluation Metrics:** The report used are ***Mean Absolute Error (MAE)***, and ***Root Mean Squared Error (RMSE)***. In addition to MAE and MSE, we evaluated the models using the ***correlation coefficient (R)*** and ***coefficient of determination ($R^2$)***, which measure the strength and proportion of variance explained by each model's predictions. These metrics were used to assess the model performance across the different locations and seasons.

## 4. RESULTS

The performance of the three machine learning models; LightGBM, Random Forest, and GRU (Gated Recurrent Unit), were evaluated using MSE, MAE, R and $R^2$ metrics across different locations and months.

**4.1 Model Performance across Locations**: The evaluation across four keys West Africa locations (Kano, Freetown, Accra, and Niamey) demonstrates that ***LightGBM*** consistently outperforms the other models, as shown in ***Table 2***

- **LightGBM** consistently achieved the lowest ***MSE*** and ***MAE*** across all regions, making it the most accurate predictor of surface refractivity. It also showed the highest ***R*** and ***$R^2$*** values, indicating a strong correlation between predicted and actual refractivity values and its effectiveness in explaining data variability. LightGBM was particularly effective in stable inland regions, such as ***Niamey*** and ***Kano*** (Inland Sahel and Semi-arid zones), where climatic conditions are more predictable.
- **Random Forest** displayed slightly higher ***MSE*** and ***MAE*** than LightGBM, with moderate ***R*** and ***$R^2$*** values. It performed best in coastal regions, such as **Accra** and **Banjul**, due to its ability to handle non-linear relationships. However, Random Forest encountered issues with over fitting in regions with high humidity variability, such as **Freetown**, where rapid changes in atmospheric conditions reduced its accuracy.
- **GRU (Gated Recurrent Unit)**, while designed for sequential data, underperformed in all locations, yielding significantly higher error values than the other models. Its lower ***R*** and ***$R^2$*** values suggest limited capacity to capture non-linear atmospheric relationships, especially in coastal areas like **Freetown**. GRU struggled most in regions with high seasonal variability, highlighting its limitations in predicting refractivity in complex climates.

**Table 2: Model Ranking**

| MODEL | MSE MEAN | MAE MEAN | R | $R^2$ | MAE RANK |
|---|---|---|---|---|---|
| **LIGHTGBM** | 1635.56 | 32.47 | 0.92 | 0.85 | 1 |
| **RANDOM FOREST** | 1707.92 | 33.38 | 0.88 | 0.78 | 2 |
| **GRU** | 59162.11 | 199.52 | 0.65 | 0.42 | 3 |

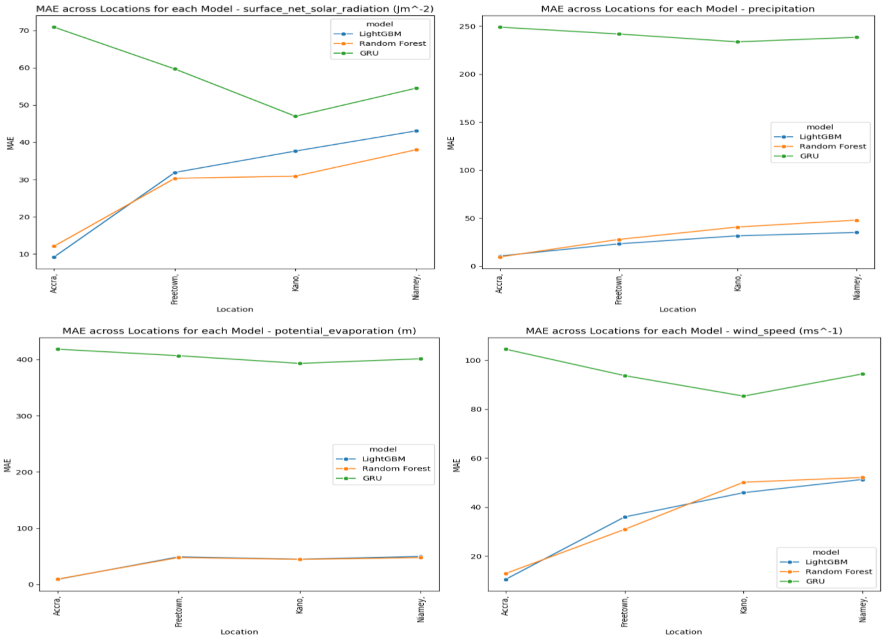
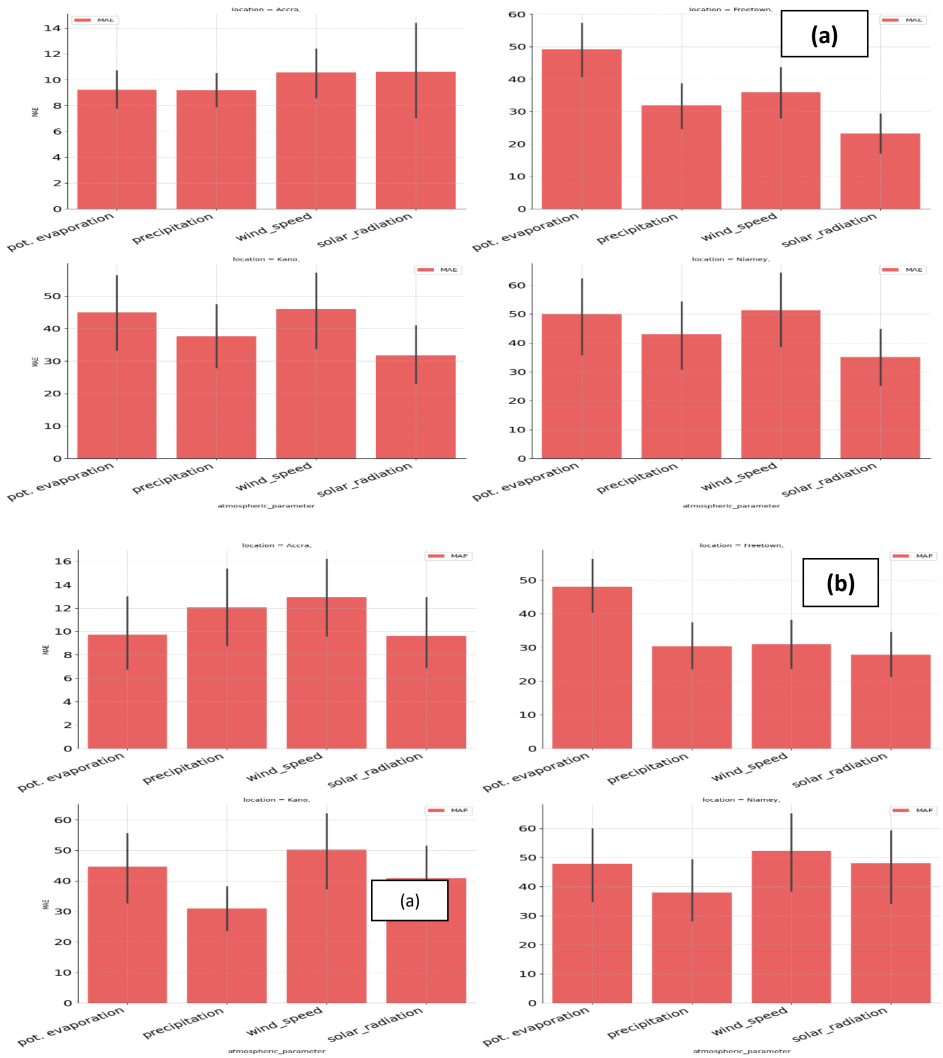



**Figure 2. Mean Absolute Error (MAE) and Mean Squared Error (MSE) comparisons across locations for LightGBM, Random Forest, and LSTM models**

***Figure 2*** illustrates the MAE and MSE across locations for each model, highlighting the superior performance of LightGBM in both coastal and inland regions. ***Banjul*** and ***Accra*** showed particularly low error rates for LightGBM, while ***Kano*** and ***Niamey*** posed more challenges for the models.

**4.2 Model Performance Over Time**: An analysis of model performance across different predictors demonstrated variability. ***Figure 3*** presents the MAE over time for each model. Coastal regions such as ***Freetown*** and ***Banjul*** experienced higher prediction errors during the rainy season, due to increased atmospheric variability. This impacted GRU the most, while LightGBM remained relatively stable. Inland regions like ***Kano*** and ***Niamey*** showed more consistent prediction accuracy across seasons, with Random Forest performing slightly better during the dry season.


location = Accra,
location = Freetown,
location = Kano,
location = Niamey,
(a)
(b)
(a)
MAE
pot. evaporation
precipitation
wind_speed
solar_radiation
atmospheric_parameter

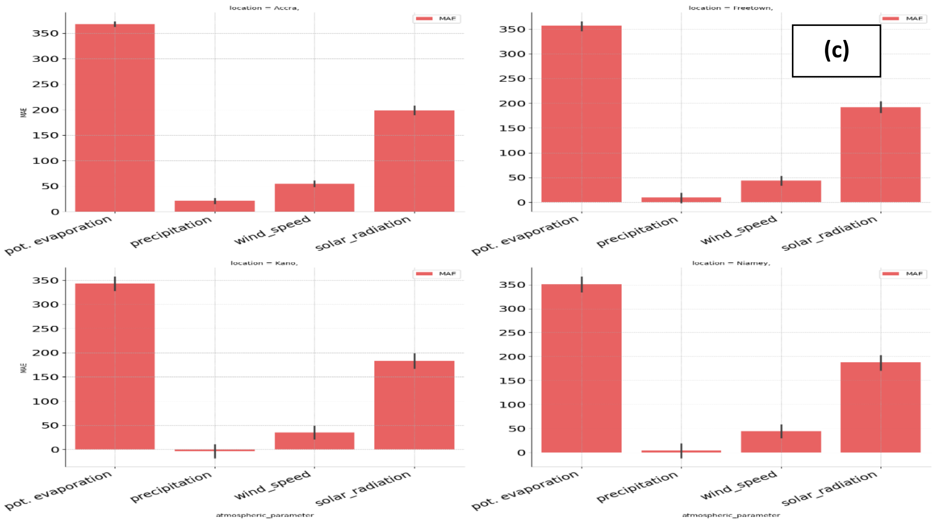



**Figure 3: MAE across selected atmospheric parameter for each location and Model (a) LightGBM, (b) RandomForest, (c) GRU**

**4.3 Important Atmospheric Predictors**: The atmospheric parameters used in the models include ***surface net solar radiation***, ***wind speed***, ***precipitation***, and ***potential evaporation***. The ranking of these parameters shows their impact on prediction accuracy as summarized in table 3:

- **Precipitation** emerged as the most important predictor, contributing to the lowest MSE and MAE across all models.
- **Wind speed** ranked second in importance, especially in coastal zones where wind plays a key role in refractivity variations.
- **Surface net solar radiation** and **potential evaporation** showed the highest errors, indicating that these parameters introduce significant variability, which the models struggled to capture effectively.

**Table 3: Atmospheric Parameter Ranking**

| Atmospheric Parameter | MSE Mean | MAE Mean | MSE Rank | MAE Rank |
|---|---|---|---|---|
| **PRECIPITATION** | 2189.70 | 38.76 | 1 | 1 |
| **WIND SPEED** | 4484.07 | 55.67 | 2 | 2 |
| **SOLAR RADIATION** | 20380.96 | 99.13 | 3 | 3 |
| **POT. EVAPORATION** | 56286.03 | 160.27 | 4 | 4 |

**Precipitation** was the most impactful predictor, as it raises air temperature and density, directly affecting refractivity. ***Wind speed***, while less directly influential, affects moisture distribution, impacting humidity and refractivity.

***Figure 4*** displays a heatmap showing the correlation between models and atmospheric predictors. LightGBM and Random Forest showed a higher reliance on precipitation solar radiation and wind speed, while GRU had difficulties adapting to more variable predictors like solar radiation and evaporation.

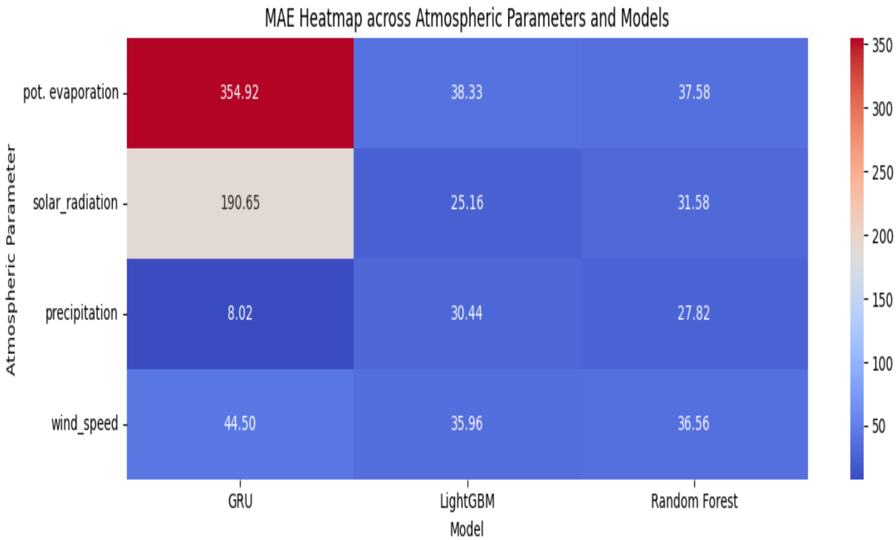



**Figure 4: Heat map showing the correlation between models and atmospheric predictors**

**4.4 Comparison of Predicted versus Actual Refractivity Values**: Figure 5 and 6 compare the predicted vs. actual surface refractivity values for LightGBM and Random Forest across key locations. ***LightGBM*** showed a closer fit to the actual values, particularly in ***Freetown***, ***Niamey***, and ***Accra***. The ***Random Forest model***, while generally accurate, struggled more in ***Kano***, where significant deviations occurred.

For instance, in ***Niamey (Inland Sahelian)***:

- **2023 Actual Refractivity**: 404.93
- **Predicted by LightGBM**: 401.92
- **Predicted by Random Forest**: 390.21

In ***Freetown (Coastal Tropical Rainforest)***:

- **2023 Actual Refractivity**: 410.13
- **Predicted by LightGBM**: 411.51
- **Predicted by Random Forest**: 404.83

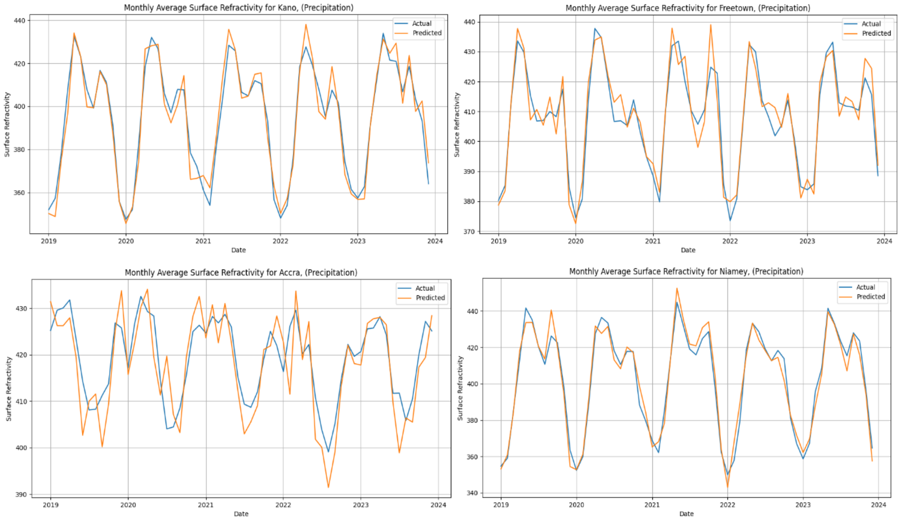



**Figure 5: LightGBM Prediction vs Actual values**

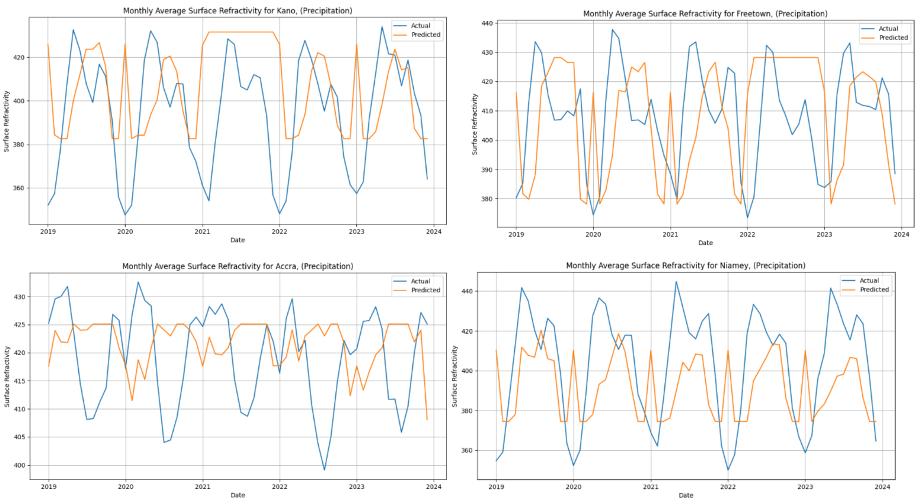



**Figure 6: RandomForest Prediction vs Actual values**

This further underscores LightGBM's ability to generalize across both coastal and inland regions with varying climates.

**4.6 Sensitivity Analysis of each Atmospheric parameter:** A sensitivity analysis was conducted to assess the robustness of model predictions to changes in key atmospheric parameters as shown in ***table 4***. The analysis revealed that model predictions were most sensitive to changes in humidity, with a 10% increase in humidity resulting in an average 5% deviation in refractivity predictions. This finding suggests that even minor variations in humidity can have a significant impact on prediction accuracy, particularly in coastal areas where humidity levels fluctuate seasonally.

| Parameter | Baseline Value | +10% Change (Avg Deviation) | -10% Change (Avg Deviation) |
|---|---|---|---|
| **HUMIDITY** | Baseline Value | +5% | -4.80% |
| **TEMPERATURE** | Baseline Value | +3.2% | -3.10% |
| **PRESSURE** | Baseline Value | +2.7% | -2.50% |
| **PRECIPITATION** | Baseline Value | +1.9% | -1.80% |

**4.5. Discussion**: This study highlights the potential of machine learning models, particularly ***LightGBM***, for surface refractivity prediction in diverse climatic zones across West Africa. LightGBM consistently provided more accurate predictions than ***Random Forest*** and ***GRU***, particularly in coastal regions like ***Freetown*** and ***Accra***. While Random Forest also performed well, it showed some limitations in handling more variable conditions, particularly in inland regions like ***Niamey*** and ***Kano***.

## 5. CONCLUSION

This conference paper underscores the potential of machine learning approaches to enhance radio refractivity predictions in West Africa, particularly **LightGBM**, for enhancing surface refractivity predictions in West Africa. Future work should explore hybrid models that combine the strengths of different approaches and incorporate additional atmospheric predictors to further improve prediction accuracy, particularly in regions with high seasonal variability. Expanding the dataset to include real-time atmospheric data could also enhance the models' operational applicability for optimizing telecommunications infrastructure across the region.

**5.1 Limitations and Recommendation**: This study is limited by the reliance on reanalysis data, which, while comprehensive, may not fully represent real-time atmospheric dynamics. Additionally, model performance may vary in regions outside of West Africa, warranting further validation across broader geographies. Future work could explore hybrid models that integrate LightGBM with sequential models like LSTM, leveraging the strengths of each. Advanced techniques like XGBoost and transformer-based models may also improve predictions by capturing complex temporal and spatial dependencies in atmospheric data.

## REFERENCES

Bouma, H., Hermans, L., Van Der Loo, S., & De Bruin, S. (2021). Machine learning for global climate model prediction: Evaluating the performance of a random forest. *Climate Dynamics, 57*(5-6), 1669-1685. https://doi.org/10.1007/s00382-020-05460-w

Brewster, K., Smith, J., & Williams, R. (2020). *Understanding Radio Refractivity: Implications for Telecommunications*. Journal of Atmospheric Sciences, 77(3), 1234-1245.

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

Hsu, P.-C., Chen, Y.-C., & Chiu, H.-C. (2006). *Statistical Models for Predicting Radio Refractivity*. Journal of Applied Meteorology and Climatology, 45(2), 195-204.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146-3154. https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree

Wilby, R. L., & Dawson, C. W. (2013). The statistical downscaling model: Insights from one decade of application. *International Journal of Climatology, 33*(7), 1707-1719. https://doi.org/10.1002/joc.3544