THE OPEN JOURNAL OF MATHEMATICAL SCIENCES AND APPLICATIONS
ISSN oooo-oooo (2026) #PP.2
(https://openjournal.utar.edu.my/index.php/cmsojmsa)

A Multistage XGBoost Architecture with Class Weighting and SMOTE for Imbalance-Aware Rain Rate Forecasting

Nurhayati Hasan[1], Hafiz Basarudin[*1], Boon Kuan Chung[1], Lloyd Ling[1], Aizat Faiz Ramli[1], and Noor Hidayah Mohd Yunus[2]

[1]*Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Malaysia*
[2]*Communication Technology Section, Universiti Kuala Lumpur, Malaysia*

Received: 20 Feb 2026, accepted: 11 Apr 2026, published online: 2 Oct 2026.

Abstract

Rainfall-induced signal attenuation, or rain fade, remains a major challenge for satellite and 5G communications in tropical climates. This study proposes a multistage XGBoost framework for hourly rain rate forecasting that explicitly addresses the issue of class imbalance in tropical rainfall data. Using 14 years of Hydro-Estimator satellite data from Peninsular Malaysia (2009–2022), the framework decomposes the forecasting task into three stages: (i) binary classification (rain/no rain), (ii) multi-class rain intensity classification, and (iii) regression to predict the actual rain rate. To mitigate the dominance of 'no rain' and 'light rain' cases, class weighting and SMOTE oversampling were applied during training. The results demonstrate that imbalance handling significantly improves detection of minority rain events. Binary classification recall improved from 0.36 to 0.64, while intensity classification macro-F1 increased from 0.32 to 0.40. In the regression stage, the combined architecture achieved near-zero MAE (0.0000), RMSE (0.0019), and R^2 = 1.0, highlighting the effectiveness of incorporating classification outputs into the regression model. Forecasting for 2023 showed that the SMOTE-enhanced model better captured extreme rainfall peaks and seasonal variability compared to baseline models. These findings confirm the potential of structured, imbalance-aware learning frameworks for accurate rainfall forecasting. The proposed method provides valuable input for adaptive communication strategies, including link margin adjustment and frequency switching, thereby enhancing the resilience of wireless networks in tropical regions.

Keywords: Rainfall forecasting, Multistage XGBoost, SMOTE oversampling, Rain fade, 5G communication

Math. Subj. Class. (2020): 05C15, 05C10

1 Introduction

Rainfall strongly affects wireless communication at millimeter-wave and satellite frequencies. Accurate hourly forecasting is critical for adaptive network planning in tropical climates. However, rainfall data are highly imbalanced, dominated by dry hours, leading to poor model sensitivity to extreme rain events.

2 Method

A multistage XGBoost pipeline was developed with three stages: binary classification (rain/no rain), multi-class intensity classification, and regression of actual rain rate. Class imbalance was mitigated using class weighting and SMOTE oversampling, improving the representation of rare high-intensity rainfall events.

3 Results

The framework was trained on 2009–2022 data and evaluated on 2023 forecasts. Key findings:

- Rain/no rain recall improved from 0.36 $\rightarrow$ 0.64 with SMOTE.
- Rain intensity macro-F1 improved from 0.32 $\rightarrow$ 0.40.
- Regression achieved MAE 0.0000, RMSE 0.0019, $R^2 = 1.0$.
- 2023 forecasts captured seasonal peaks and improved tail distribution fidelity compared to baseline.

4 Conclusions

The proposed multistage XGBoost with imbalance handling enhances hourly rain rate prediction and captures rare, high-intensity events vital for 5G and satellite communication systems. Future work will extend the approach with sequence-aware models such as LSTM and Transformer for improved temporal forecasting.

References

Fernández, A. et al. (2019). Smote for learning from imbalanced data: progress and challenges. *Journal of Artificial Intelligence Research*, 61:863–905.

Hasan, N. et al. (2024). Rain height and satellite interference over malaysia from 1992 to 2022. *Engineering, Technology and Applied Science Research*, 14(5):16874–16880.

Jaiswal, A. K. and Mahto, R. P. (2020). Rainfall prediction using machine learning techniques: A survey. *IEEE Access*, 8:134770–134792.

Liu, Y. et al. (2020). Short-term rainfall forecasting using lstm recurrent neural network. *Water*, 12(6):1681.

* Corresponding author: hafizba@utar.edu.my

Wang, J. et al. (2022). Improved rainfall forecasting using xgboost and smote. *Computers and Electronics in Agriculture*, 197:106957.