

Machine Learning–Based Nowcasting of Extreme Rainfall over Aceh Using ERA5 Reanalysis

Muhammad Raza Adzani

Department of Informatics

Syiah Kuala University

Banda Aceh, Indonesia

email: raza.a22@mhs.usk.ac.id

Ahmad Siddiq

Department of Informatics

Syiah Kuala University

Banda Aceh, Indonesia

email: m.siddiq2@mhs.usk.ac.id

Abstract—Extreme rainfall is a major trigger of floods and landslides in many regions of Indonesia, including Aceh Province, where short-lived but intense precipitation events can rapidly escalate into disasters. Existing early-warning systems often rely on numerical weather prediction models that may exhibit limited skill for very short lead times at local scales. This study proposes a machine learning–based framework for three-hour-ahead extreme rainfall nowcasting over Aceh using ERA5 single-level reanalysis data. The problem is formulated as a binary classification task on a $0.25^\circ \times 0.25^\circ$ latitude–longitude grid, where extreme events are defined as three-hour accumulated precipitation exceeding a high percentile threshold. A set of meteorological and temporal predictors, including near-surface atmospheric variables and short-term precipitation history, is planned to be used as model input. Logistic regression and random forest classifiers are proposed and evaluated using a chronological data-splitting strategy to emulate an operational setting. This study aims to assess the feasibility of reanalysis-driven machine learning approaches for short-range extreme rainfall early warning in data-scarce regions such as Aceh.

Index Terms—extreme rainfall, nowcasting, machine learning, random forest, ERA5 reanalysis, Aceh

I. INTRODUCTION

Extreme rainfall is one of the primary drivers of floods and landslides across Indonesia, and its impacts are particularly severe in regions with complex topography and limited monitoring infrastructure. Aceh Province, located at the northern tip of Sumatra, frequently experiences intense convective rainfall associated with monsoonal circulation and local atmospheric instability. These short-lived but high-intensity events often result in flash floods that cause infrastructure damage and threaten human safety.

Operational rainfall early-warning systems typically depend on numerical weather prediction (NWP) models, rain-gauge observations, and, where available, weather radar. However, in many parts of Indonesia, observational networks remain sparse and radar coverage is limited. Moreover, NWP models may exhibit biases and reduced skill when predicting extreme precipitation at short lead times and fine spatial scales. These limitations motivate the exploration of complementary, data-driven approaches that can exploit existing atmospheric datasets.

Reanalysis products such as ERA5 provide physically consistent atmospheric variables at high temporal resolution over

long periods. Although reanalysis data are not forecasts, they can be leveraged to learn empirical relationships between atmospheric conditions and subsequent extreme rainfall. Recent advances in machine learning have demonstrated strong potential for extracting such relationships from large environmental datasets. This study therefore proposes a machine learning–based framework for three-hour-ahead extreme rainfall nowcasting over Aceh using ERA5 reanalysis variables.

II. RELATED WORK

Machine learning techniques have been increasingly applied to precipitation forecasting and extreme rainfall analysis as complements to traditional statistical and numerical approaches. Random forest classifiers have been shown to improve exceedance forecasts for heavy precipitation and to outperform linear models such as logistic regression in distinguishing extreme from non-extreme events [1], [8]. These tree-based methods are particularly effective at capturing non-linear interactions among meteorological predictors.

Several studies have focused on correcting or post-processing reanalysis precipitation using machine learning, demonstrating improved hydrological simulations and enhanced representation of extreme rainfall [2]. Other work has applied machine learning models to rainfall occurrence and intensity prediction at various temporal scales, highlighting the potential of relatively simple models to capture complex atmospheric relationships [3], [4], [10].

Deep learning approaches, including convolutional neural networks and recurrent architectures, have also been explored for precipitation nowcasting using radar and satellite data [5]–[7], [9]. While these methods often achieve high skill, they typically require dense observational coverage and large volumes of high-resolution data, which may not be available in many tropical regions. In contrast, fewer studies have explicitly investigated short-lead extreme rainfall classification using reanalysis variables alone, particularly for regions such as Aceh. This motivates the proposed reanalysis-driven machine learning framework.

III. METHODOLOGY

A. Data Source and Study Area

The proposed study will use ERA5 single-level reanalysis data obtained from the Copernicus Climate Data Store. Hourly data covering the period 2020–2024 will be extracted for a domain surrounding Banda Aceh, Indonesia (5.0° – 6.0° N, 95.0° – 96.0° E). The spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$ corresponds to a 5×5 grid that includes both coastal and inland areas exposed to intense rainfall.

B. Problem Formulation

The prediction task is formulated as a binary classification problem at the grid-cell level. For each grid point and time step, the objective is to predict whether the accumulated precipitation over the subsequent three hours will exceed an extreme threshold. Extreme rainfall events are planned to be defined using a high percentile (e.g., the 95th percentile) of the three-hour accumulated precipitation distribution computed over the study period.

C. Feature Engineering

Input features are planned to include near-surface meteorological variables, short-term precipitation history, and temporal encodings. Lagged precipitation features and rolling means will be used to capture rainfall persistence, while diurnal and seasonal cycles will be represented using sine and cosine transformations of time variables.

D. Proposed Models and Evaluation

Two machine learning models are proposed for evaluation: logistic regression as a baseline and random forest as a non-linear ensemble model. A chronological data-splitting strategy will be used to emulate an operational nowcasting scenario and avoid temporal leakage. Model performance will be assessed using classification metrics such as accuracy, precision, recall, and the area under the receiver operating characteristic curve, with particular emphasis on recall for extreme events.

REFERENCES

- [1] G. R. Herman and R. S. Schumacher, “Money Doesn’t Grow on Trees, but Forecasts Do: Forecasting Extreme Precipitation with Random Forests,” *Monthly Weather Review*, vol. 146, no. 5, pp. 1571–1600, 2018.
- [2] H. Sun *et al.*, “Corrected ERA5 precipitation by machine learning significantly improved flow simulations,” *Journal of Hydrometeorology*, vol. 23, no. 10, pp. 1663–1679, 2022.
- [3] B. Yang *et al.*, “A method for monthly extreme precipitation forecasting,” *Water*, vol. 15, no. 8, 2023.
- [4] S. Chkeir *et al.*, “Nowcasting extreme rain with machine learning techniques,” *Atmospheric Research*, vol. 282, 2023.
- [5] S. Ravuri *et al.*, “Skillful precipitation nowcasting using deep generative models,” *Nature*, vol. 597, pp. 672–677, 2021.
- [6] G. Ayzel *et al.*, “RainNet: A CNN for radar-based precipitation nowcasting,” *Geoscientific Model Development*, vol. 13, pp. 2631–2644, 2020.
- [7] F. Gamboa-Villafruela, “Immediate precipitation forecasting using deep learning,” *Proceedings*, vol. 8, no. 1, 2021.
- [8] D. Wolfensberger and A. Feinberg, “RainForest: A random forest algorithm for precipitation estimation,” *Atmospheric Measurement Techniques*, vol. 14, pp. 3169–3193, 2021.
- [9] X. Shi *et al.*, “Convolutional LSTM network for precipitation nowcasting,” *NeurIPS*, 2015.
- [10] X. Shi, “Smart dynamical downscaling of extreme precipitation,” *Geophysical Research Letters*, vol. 47, 2020.