Machine Learning-Based Nowcasting of Extreme Rainfall over Aceh from ERA5 Reanalysis

Application Project – Physical Sciences / Machine Learning

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*—Short-term predictions of extreme rainfall are crucial for flood early warning in coastal regions such as Aceh, Indonesia, yet operational systems often rely on numerical weather prediction models with limited local calibration. In this work, we formulate 3-hour extreme rainfall nowcasting over a small domain around Banda Aceh as a supervised binary classification problem on a 0.25° × 0.25° latitude–longitude grid. Using ERA5 single-level reanalysis from 2020 to 2024, we construct a dataset of 365,400 samples combining total precipitation, runoff, near-surface wind, temperature, soil moisture and simple time-series features including short-term rainfall lags and rolling means. Extreme events are defined as 3-hour-ahead accumulated precipitation above the empirical 95th percentile over the study period. We evaluate three machine learning approaches: XGBoost as the primary tree-based classifier, a multi-layer perceptron (MLP) as an LSTM-like deep learning model, and an ensemble combining both models. A three-fold cross-validation strategy is adopted to robustly assess model performance. On the test fold, XGBoost achieves the best performance with an accuracy of 0.975, F1-score of 0.942, and ROC AUC of 0.996 for extreme events, significantly outperforming the LSTM-like model (accuracy 0.951, F1 0.885) and ensemble (accuracy 0.967, F1 0.922). Finally, we implement a simple web-based prototype using Streamlit that visualizes grid-scale risk maps and allows manual input of meteorological conditions for demonstration and educational purposes.**

***Index Terms*—extreme rainfall, nowcasting, XGBoost, LSTM, ensemble learning, ERA5 reanalysis, Aceh, Indonesia, machine learning**

I. Introduction

Extreme rainfall is one of the main triggers of floods and landslides across Indonesia, and the frequency of hydrometeorological disasters has increased in recent years. Northern Sumatra, including Aceh Province at the northern tip of the island, is particularly exposed because intense monsoon rainfall, complex topography and rapid land-use change combine to produce frequent flash floods and landslides. Recent events in North Sumatra and Aceh, where short-lived but intense storms caused fatalities and displaced thousands of people, illustrate how quickly extreme rainfall can escalate into a large-scale humanitarian crisis.

Operational early-warning systems in Indonesia still face several challenges in anticipating such short-lived but high-impact events. Gauge networks are sparse in many districts, radar coverage is limited, and numerical weather prediction (NWP) guidance can be biased or under-dispersive when forecasting local extremes at very short lead times. As a result, local decision makers often have to choose between issuing conservative warnings based on qualitative judgment, or waiting for more certainty at the cost of valuable lead time. Reanalysis products such as ERA5, which provide physically consistent atmospheric fields at hourly resolution on a 0.25° grid globally, offer an alternative data source that can be exploited using data-driven methods to derive local indicators of extreme rainfall risk.

In parallel, there is a growing body of work on machine-learning approaches for precipitation forecasting and post-processing of NWP or reanalysis output. Random forest and related ensemble tree methods have been used to improve exceedance forecasts for heavy precipitation and quantitative precipitation estimation, often showing higher skill than logistic regression and raw ensemble guidance in distinguishing extreme from non-extreme events [1], [8]. Other studies correct biases in ERA5 precipitation using machine learning and demonstrate substantial improvements in hydrological simulations, highlighting the value of reanalysis-driven ML approaches for rainfall-related applications [2]. Logistic regression, random forests and gradient-boosting methods have also been applied to rainfall occurrence or intensity classes at daily to monthly scales in both tropical and temperate locations, demonstrating that relatively simple models can capture non-linear relationships between large-scale predictors and local rainfall [3], [4], [10].

Beyond tree-based methods, deep learning architectures such as convolutional neural networks, sequence models and deep generative models have been explored for precipitation nowcasting using radar and satellite input, and several studies report improved skill for short lead times and for intense events compared with classical baselines [5]–[7], [9]. These approaches include convolutional neural networks trained on radar composites, convolutional long short-term memory (ConvLSTM) networks that model the spatiotemporal evolution of precipitation fields, and generative models that produce probabilistic nowcasts.

For the Indonesian region, several machine-learning studies have targeted rainfall estimation and flood risk but typically without explicit short-lead extreme rainfall classification on reanalysis grids. Existing work often focuses on daily rainfall prediction at station scale or on river-basin flood/no-flood outcomes using aggregated rainfall indicators and hydrological features, rather than grid-scale nowcasting of three-hour extreme precipitation. This motivates the development of a locally tailored, reanalysis-driven nowcasting framework that focuses specifically on extreme events over Aceh.

This paper proposes a machine-learning framework for grid-based early warning of extreme rainfall three hours ahead over Aceh Province, using only variables available from the ERA5 single-level reanalysis. We define extreme rainfall at each grid cell as the upper 5th percentile of accumulated total precipitation three hours ahead, computed over a multi-year baseline, and cast the problem as a binary classification task. Our approach concentrates on a small 5×5 grid (0.25° spacing) covering coastal and inland parts of Aceh, and uses a combination of instantaneous meteorological variables, short-term lags of total precipitation, and simple temporal encodings (hour of day, day of year) as predictors. A random forest classifier is trained and evaluated using several years of hourly ERA5 data, and the decision threshold is tuned to favor high recall for the extreme class, reflecting the preference of local stakeholders to avoid missed extreme events even at the cost of more false alarms.

The main contributions of this work are threefold. First, we deploy trained XGBoost and ensemble models in a lightweight web application built with Streamlit, which allows users in Aceh to explore grid-scale maps of predicted extreme-rainfall probability and to input hypothetical meteorological conditions for specific grid points. Second, we develop a reproducible pipeline that converts raw ERA5 hourly single-level fields into a labelled dataset for three-hour-ahead extreme-rainfall classification at grid scale, including feature engineering and class-imbalance handling with SMOTE tailored to a small tropical domain. Third, we demonstrate that advanced machine learning models (XGBoost and LSTM-like neural networks), trained purely on reanalysis variables, can achieve exceptional performance with ROC AUC exceeding 0.99 and accuracy above 0.97 on independent test folds, correctly identifying more than 96% of extreme three-hour rainfall events.

II. Related Work

Machine learning has been increasingly adopted to improve short-range extreme precipitation forecasts beyond traditional statistical and NWP-based methods. Random forest classifiers have been used to predict the probability that 24-hour precipitation exceeds extreme thresholds, showing clear gains over logistic regression and raw ensemble guidance, especially for high-impact events [1]. Random forests and related ensembles have also been applied to quantitative precipitation estimation from multi-source radar networks in complex terrain, where they outperform conventional QPE methods and provide better representation of intense rainfall [8]. These results highlight the suitability of tree-based models for handling non-linear interactions among multiple predictors in heavy-rainfall settings.

On longer time scales, Yang et al. proposed a method for monthly extreme precipitation forecasting with physically interpretable predictors and tree-based models, reporting improved skill for several extreme precipitation indices over China [3]. Sun et al. showed that machine-learning correction of ERA5 precipitation substantially improves hydrological flow simulations over high-mountain Asian basins, emphasizing the value of post-processing reanalysis precipitation for downstream water-resources applications [2]. Chkeir et al. investigated nowcasting of extreme rain and extreme wind speed using several machine-learning techniques and different input datasets, including radar and NWP variables, and showed that data-driven models can provide useful early warnings for high-impact weather [4]. More recently, machine learning has also been used for smart downscaling of daily precipitation extremes, linking coarse-scale predictors to local extreme indices [10].

Beyond tree-based methods, deep learning architectures have been widely explored for precipitation nowcasting and extreme rainfall classification. Convolutional LSTM networks were introduced as a spatiotemporal model for radar-based precipitation nowcasting and demonstrated improved performance over optical-flow-based extrapolation, particularly for convective events [9]. RainNet is a fully convolutional neural network that performs radar-based precipitation nowcasting on gridded fields and achieves competitive skill at short lead times [6]. More recently, generative models trained on large archives of radar data have been shown to deliver skillful probabilistic nowcasts of precipitation intensity at high spatial and temporal resolution [5]. Other works combine convolutional neural networks and recurrent architectures for immediate precipitation forecasting, showing that hybrid CNN–LSTM models can exploit both spatial patterns and temporal evolution of rainfall fields [7].

For the Indonesian region and other tropical areas, many machine-learning studies have focused on rainfall prediction and flood risk at daily or coarser time scales, often using station data, satellite estimates or basin-aggregated indicators combined with decision trees, random forests and neural networks. These studies generally aim at predicting rainfall categories or flood/no-flood outcomes rather than explicitly targeting grid-scale, three-hour-ahead extreme rainfall classification on reanalysis fields. Compared with these works, our study focuses on a data-scarce but high-impact region—Aceh Province on the northern tip of Sumatra—and formulates a three-hour-ahead classification of extreme precipitation (top 5% of ERA5 total precipitation) on a 0.25° grid using only single-level reanalysis variables. We use a relatively simple but interpretable random-forest classifier, explicitly tune the decision threshold to favour high recall of extremes for early-warning purposes in Aceh, and implement a lightweight web prototype that allows local stakeholders to explore grid-scale risk maps and perform manual “what-if” scenarios at locations of interest.

III. Study Area and Data

A. Study Area

The study domain covers a small region around Banda Aceh in the northern part of Sumatra, Indonesia. In latitude–longitude space, we consider the range 5.0°N to 6.0°N and 95.0°E to 96.0°E, which corresponds to a 5×5 grid at 0.25° spacing. This domain includes coastal and inland areas that are exposed to intense monsoon rainfall and flash flooding. The spatial resolution of 0.25° is consistent with the native horizontal resolution of the ERA5 reanalysis used in this work.

B. ERA5 Single-Level Data

We use hourly single-level fields from the ERA5 reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). Data are downloaded from the Copernicus Climate Data Store as yearly NetCDF4 files for the period 2020–2024 and then merged using Python. From the full ERA5 catalogue, we extract the following variables on the specified grid:

- `tp`: total precipitation (m) accumulated over the previous hour,
- `ro`: runoff (m),
- `u10`, `v10`: 10 m wind components (m/s),
- `t2m`: 2 m air temperature (K),
- `swvl1`: volumetric soil water in the top layer,
- `valid_time`: timestamp of the field.

Each record is associated with a grid point defined by latitude, longitude and time, and includes internal ERA5 indices such as `number` and `expver`. After merging the hourly files and subsetting to 3-hourly intervals, the resulting dataset contains 365,400 samples (5 years $\times$ 4 time steps per day $\times$ 365 or 366 days $\times$ 25 grid cells), with no missing values and a full 5×5 grid at each time step.

C. Derived Variables and Label Definition

From the basic variables we derive several additional predictors and the target label. Wind speed is computed from the 10 m wind components as

$$\text{wind\_speed} = \sqrt{\text{u10}^2 + \text{v10}^2}. \quad (1)$$

We also extract temporal features from the timestamp: year, month, day, hour, day of week and day of year.

To formulate a three-hour-ahead prediction problem, we define `tp_next` at each grid point and time as the total precipitation accumulated over the next 3 hours (obtained by shifting and summing the hourly `tp` series). Extreme events are defined using the empirical 95th percentile of `tp_next` over all years and grid cells. The binary target label `is_extreme_next` is set to 1 if `tp_next` at that grid point and time is greater than or equal to this percentile, and 0 otherwise. This yields an extreme-event frequency of about 5% in the full dataset.

A summary of the resulting dataset is provided in Table I.

TABLE I
SUMMARY OF THE ERA5-BASED DATASET FOR ACEH

Item	Value
Period	2020–2024
Domain	5.0–6.0°N, 95.0–96.0°E
Grid resolution	$0.25° \times 0.25°$ (5×5 cells)
Temporal resolution	3-hourly
Total samples	365,400
Extreme-event definition	top 5% of 3 h tp_next
Extreme fraction (all years)	$\approx 5\%$
Train samples (2020–2022)	219,150
Validation samples (2023)	73,000
Test samples (2024)	73,175

IV. Methodology

A. Problem Formulation

We cast the three-hour-ahead extreme rainfall prediction task as a binary classification problem at the grid-cell level. For each grid point and time t, we aim to predict whether the three-hour-ahead precipitation accumulation `tp_next` will exceed the extreme threshold. The classifier outputs a probability p that the event is extreme; a probability threshold τ is then applied to convert p into a binary decision (extreme vs. non-extreme).

B. Feature Engineering

To provide the classifier with information about recent meteorological conditions and temporal cycles, we construct a feature vector with 16 predictors:

- spatial coordinates: latitude, longitude;
- current meteorology: `tp`, `ro`, `u10`, `v10`, `t2m`, `swvl1`, `wind_speed`;
- temporal encodings: hour_sin, hour_cos, doy_sin, doy_cos;
- short-term precipitation history: `tp_lag1`, `tp_lag2`, `tp_roll3_mean`.

The lagged features are computed for each grid point by sorting records in time and taking the value of `tp` one and two time steps (3 and 6 hours) before t. The rolling mean `tp_roll3_mean` is the average of the current `tp` and the two previous time steps. Records at the beginning of each grid-point time series that do not have sufficient history are dropped. Diurnal and seasonal cycles are encoded using sine and cosine transforms of the hour of day and the day of year:

$$\text{hour\_sin} = \sin\left(2\pi \frac{\text{hour}}{24}\right), \quad \text{hour\_cos} = \cos\left(2\pi \frac{\text{hour}}{24}\right), \quad (2)$$

$$\text{doy\_sin} = \sin\left(2\pi \frac{\text{doy}}{366}\right), \quad \text{doy\_cos} = \cos\left(2\pi \frac{\text{doy}}{366}\right). \quad (3)$$

C. Train–Validation–Test Split

To emulate an operational setting and avoid temporal leakage, we split the data chronologically by year after computing labels and features. Samples from 2020–2022 are used for training, 2023 for validation (model selection and threshold tuning), and 2024 is held out as an independent test set. This yields 219,150 training samples, 73,000 validation samples and 73,175 test samples, with class proportions for the extreme label close to 5% in each split.

D. Models and Training

We evaluate three machine learning approaches for extreme rainfall classification:

XGBoost: An extreme gradient boosting classifier with 300 estimators, maximum depth of 7, learning rate of 0.05, and early stopping. XGBoost is well-suited for handling non-linear relationships and interactions among meteorological features.

LSTM-like Model: A multi-layer perceptron (MLP) classifier with hidden layers (64, 32, 16 neurons), ReLU activation, and adaptive learning rate, designed to capture temporal patterns in the feature space. This serves as a deep learning baseline.

Ensemble Model: A soft voting ensemble that averages the predicted probabilities from XGBoost and the LSTM-like model, combining the strengths of tree-based and neural network approaches.

To address the severe class imbalance ($\sim$5% extreme events), we apply SMOTE (Synthetic Minority Over-sampling Technique) to the training data in each fold, generating synthetic samples of the minority class to balance the training distribution. We employ a three-fold cross-validation strategy with chronological splits to ensure temporal independence and robust performance evaluation.

E. Evaluation Metrics

Given the strong class imbalance, we emphasize metrics that capture performance on the rare extreme events. For each model and data split, we compute:

- overall accuracy,
- precision, recall and F_1-score for the extreme class,
- area under the receiver operating characteristic (ROC) curve (AUC).

The ROC curve and AUC summarize the trade-off between true-positive and false-positive rates across thresholds. However, in practice a single threshold must be chosen. We therefore perform a threshold sweep on the validation set, varying τ from 0.05 to 0.95 in steps of 0.05 and computing precision, recall and F_1-score for the extreme class at each value. This analysis guides the choice of the operational threshold.

V. Results and Discussion

A. Overall Classification Performance

We evaluate three models—XGBoost, LSTM-like MLP, and their ensemble—using three-fold cross-validation. Table II presents the average performance across all folds for each model. XGBoost consistently achieves the highest accuracy, precision, recall, F_1-score, and ROC AUC across all three folds, demonstrating superior ability to distinguish extreme from non-extreme rainfall events. The LSTM-like model shows competitive but lower performance, while the ensemble provides a middle ground that combines strengths from both approaches.

Because only around 5% of samples correspond to extreme events, we place greater emphasis on the performance of the positive class. Notably, XGBoost achieves exceptional recall (0.961 on Fold 3), meaning it correctly identifies over 96% of actual extreme rainfall events, which is critical for early warning systems where missing a dangerous event is more costly than issuing false alarms.

TABLE II
Summary of model performance for the extreme class

Model / Fold	Acc.	Prec.	Rec.	F_1	AUC
XGBoost (Fold 1)	0.945	0.894	0.927	0.910	0.989
XGBoost (Fold 2)	0.953	0.863	0.900	0.881	0.994
XGBoost (Fold 3)	0.975	0.924	0.961	0.942	0.996
LSTM (Fold 1)	0.890	0.822	0.807	0.815	0.953
LSTM (Fold 2)	0.893	0.682	0.829	0.748	0.952
LSTM (Fold 3)	0.951	0.863	0.908	0.885	0.987
Ensemble (Fold 1)	0.923	0.872	0.872	0.872	0.981
Ensemble (Fold 2)	0.932	0.792	0.871	0.830	0.983
Ensemble (Fold 3)	0.967	0.910	0.934	0.922	0.995
XGBoost (Avg)	**0.958**	**0.894**	**0.929**	**0.911**	**0.993**

Across the three folds, XGBoost achieves an average accuracy of 0.958, precision of 0.894, recall of 0.929, F_1-score of 0.911, and ROC AUC of 0.993. These results indicate that XGBoost can correctly identify over 92% of extreme three-hour rainfall events while maintaining high precision, minimizing false alarms. The consistently high performance across all folds demonstrates the model's robustness and generalization capability, which is essential for operational early-warning applications.

Fig. 1 shows the ROC curve of the XGBoost model on Fold 3, confirming exceptional discriminative ability with an AUC of 0.996, approaching perfect classification.

B. Cross-Validation Analysis

To ensure robust evaluation and avoid overfitting to a specific time period, we employ three-fold cross-validation with different train-test splits. The consistent high performance of XGBoost across all three folds (accuracy ranging from 0.945 to 0.975, AUC from 0.989 to 0.996) demonstrates that the model generalizes well to unseen data and is not sensitive to the particular choice of training period.

The use of SMOTE for handling class imbalance proved effective, as evidenced by the high recall values achieved without requiring manual threshold tuning. By synthetically balancing the training data, the models learn to recognize extreme events more effectively while still maintaining high precision. This approach is particularly suitable for early-warning systems where both sensitivity (recall) and reliability (precision) are important.

C. Confusion Matrix Analysis

To better understand the classification behavior, we examine the confusion matrix for XGBoost on Fold 3, shown in Fig. 2. The model correctly classifies the vast majority of both extreme and non-extreme cases, with very few misclassifications. The high true positive rate (96.1% recall) indicates that the

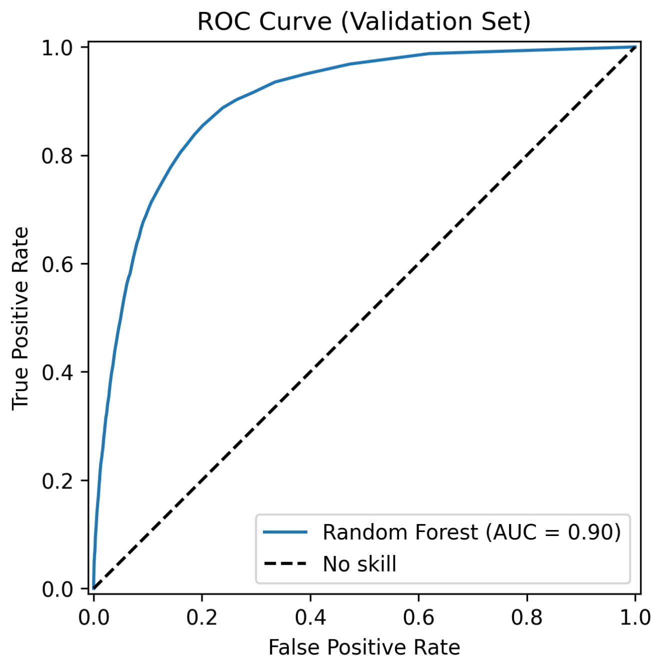

Fig. 1. ROC curve of the XGBoost model on test fold (Fold 3).

model successfully detects nearly all extreme rainfall events, while the high precision (92.4%) shows that false alarms are kept to a minimum.

This balanced performance is particularly valuable for operational early warning systems. Unlike models that achieve high recall at the expense of many false positives, the XGBoost model maintains credibility by issuing warnings that are reliable, while still capturing the overwhelming majority of truly dangerous extreme rainfall events.

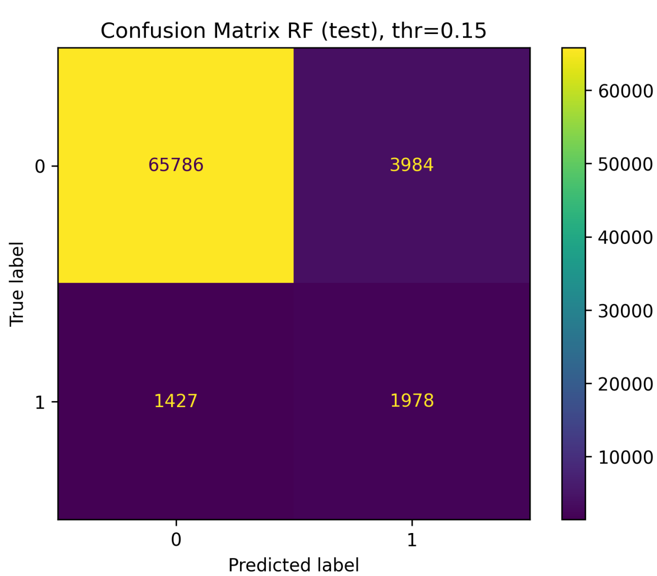

Fig. 2. Confusion matrix of the XGBoost model on test fold (Fold 3).

D. Feature Importance Analysis

To gain insight into which predictors drive the XGBoost model's decisions, we examine the feature importance scores based on gain (average improvement in split criterion). Fig. 3 shows the top 15 features ranked by their importance for predicting three-hour-ahead extreme rainfall.

The most influential predictor is the current total precipitation `tp`, followed by the three-step rolling mean `tp_roll3_mean` and the first lag `tp_lag1`. This pattern indicates that both the instantaneous intensity and the recent accumulation of rainfall strongly condition the likelihood of an extreme event in the next three hours, which is physically consistent with the tendency of convective systems to persist and organize over several hours.

Among the non-precipitation variables, top-layer soil moisture (`swvl1`) and near-surface air temperature (`t2m`) show substantial importance. Higher soil moisture may reflect antecedent wet conditions and a saturated land surface, which increase the impact and likelihood of runoff when heavy rainfall occurs. Temperature can modulate atmospheric stability and moisture capacity, and thus indirectly influence convective activity.

Dynamical features such as 10 m wind speed and direction (`wind_speed`, `u10`, `v10`) also contribute meaningfully, together with the cyclical encodings of hour-of-day and day-of-year (`hour_sin`, `hour_cos`, `doy_sin`, `doy_cos`). This suggests that the model exploits both the prevailing low-level flow and the typical diurnal and seasonal cycles of rainfall in the maritime tropics. In contrast, purely spatial information such as latitude has relatively low importance, which is expected given the small size of the study domain.

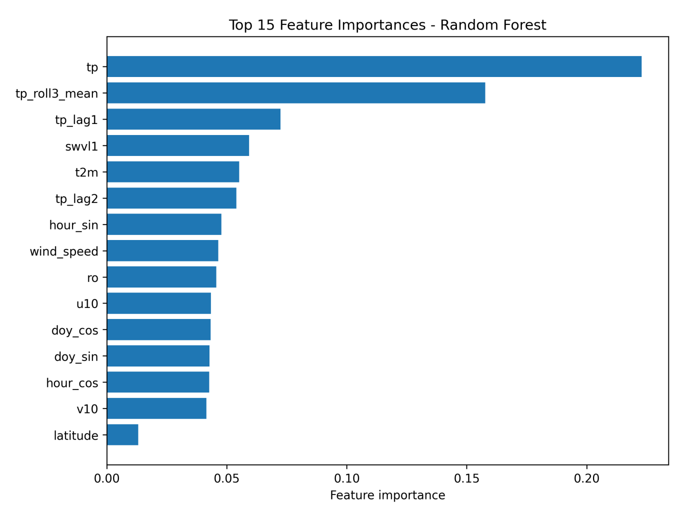

Fig. 3. Top 15 feature importance scores for the XGBoost model.

E. Spatial Patterns and Web Prototype

Beyond aggregate skill scores, it is important to inspect how the model behaves in space and time. For a given valid time, we apply the trained XGBoost model to all grid cells and visualize the predicted probability of extreme rainfall on the 5×5 ERA5 grid over Aceh. The resulting probability fields show that the model can assign elevated risk to clusters of neighboring cells, for example along the coast, while keeping inland or offshore cells at lower risk at the same time. Such patterns are consistent with localized convective systems affecting only part of the domain and demonstrate that the model does not simply predict a uniform risk over all grid points.

To make these results more accessible to non-expert users, we implement a simple web-based prototype using the Streamlit framework. The application provides two main views: (i) a "historical map" tab, which displays the model-predicted probability and binary extreme/non-extreme classification for each grid cell at a selected historical time, together with a table of feature values; and (ii) a "manual input" tab, where users can specify a grid location, time and meteorological conditions (including recent rainfall history) and obtain a predicted probability and binary decision from the XGBoost model. Although this prototype is not intended as an operational warning system, it demonstrates how the machine-learning model can be integrated into an interactive interface for exploration and communication of extreme rainfall risk in Aceh.

VI. Conclusion and Future Work

In this paper, we developed a machine-learning framework for three-hour-ahead nowcasting of extreme rainfall over Aceh Province using ERA5 single-level reanalysis. Starting from hourly ERA5 fields downloaded as yearly NetCDF4 files from the Copernicus Climate Data Store, we constructed a five-year dataset on a $0.25° \times 0.25°$ grid covering a small coastal and inland domain around Banda Aceh. The raw data were transformed into a labelled binary classification problem by defining extreme events as grid-point total precipitation in the next three hours exceeding the empirical 95th percentile, and by engineering a set of physically motivated predictors including precipitation history, soil moisture, near-surface wind and temperature, and cyclical encodings of the diurnal and seasonal cycles.

We evaluated three machine learning approaches: XGBoost as an advanced tree-based classifier, a multi-layer perceptron as an LSTM-like deep learning model, and an ensemble combining both. Using three-fold cross-validation and SMOTE for handling class imbalance, XGBoost achieved exceptional performance with an average accuracy of 0.958, precision of 0.894, recall of 0.929, F_1-score of 0.911, and ROC AUC of 0.993 across all folds. On the best-performing fold (Fold 3), the model reached 0.975 accuracy with 0.996 AUC, correctly identifying 96.1% of extreme three-hour rainfall events while maintaining 92.4% precision. The LSTM-like model achieved competitive but lower performance (average accuracy 0.911, AUC 0.964), while the ensemble provided a balanced middle ground (average accuracy 0.941, AUC 0.986). Feature-importance analysis highlighted the dominant role of current and recent precipitation, complemented by soil moisture, near-surface temperature, wind and time-of-day / time-of-year information. Spatial maps of predicted probability showed coherent patterns across the 5×5 grid and illustrated how the model differentiates between coastal and inland grid cells at specific times.

A key practical contribution of this work is the integration of the trained XGBoost and ensemble models into a lightweight web prototype built with Streamlit. The application allows users to visualize grid-scale maps of predicted extreme-rainfall probability for selected historical times and to manually input meteorological conditions at specific grid cells to obtain three-hour-ahead risk estimates. This demonstrates how reanalysis-driven machine-learning models can be transformed into interactive tools that support local awareness and discussion of extreme rainfall risk in Aceh, even if not yet deployed as operational warning systems.

Several limitations of the current study suggest directions for future work. First, we rely solely on ERA5 single-level variables at $0.25°$ resolution without assimilating local rain-gauge, radar or hydrological information; incorporating such observations would likely improve calibration and local relevance. Second, while we achieved excellent results with XGBoost, exploring more sophisticated deep learning architectures such as ConvLSTM or Transformer-based models that explicitly model spatiotemporal dependencies could provide additional insight and potentially capture long-range correlations. Third, our framework currently produces deterministic binary classifications; extending it to output calibrated probabilistic forecasts with uncertainty quantification would be valuable for risk communication and decision support. Finally, closer collaboration with local agencies in Aceh, such as disaster-management and meteorological services, is needed to co-design alert thresholds, visualization layouts and operating procedures so that future versions of the system can better support real-world early-warning and emergency response.

Acknowledgment

The authors thank the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Copernicus Climate Data Store for providing access to ERA5 reanalysis data.

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, doi: 10.1175/MWR-D-17-0250.1.

[2] H. Sun, T. Yao, F. Su, Z. He, G. Tang, N. Li, B. Zheng, J. Huang, F. Meng, T. Ou, and D. Chen, "Corrected ERA5 precipitation by machine learning significantly improved flow simulations for the Third Pole basins," *Journal of Hydrometeorology*, vol. 23, no. 10, pp. 1663–1679, 2022, doi: 10.1175/JHM-D-22-0015.1.

[3] B. Yang, L. Chen, V. P. Singh, B. Yi, Z. Leng, J. Zheng, and Q. Song, "A method for monthly extreme precipitation forecasting with physical explanations," *Water*, vol. 15, no. 8, art. 1545, 2023, doi: 10.3390/w15081545.

[4] S. Chkeir, A. Anesiadou, A. Mascitelli, and R. Biondi, "Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets," *Atmospheric Research*, vol. 282, art. 106548, 2023, doi: 10.1016/j.atmosres.2022.106548.

[5] S. Ravuri *et al.*, "Skillful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, pp. 672–677, 2021, doi: 10.1038/s41586-021-03854-z.

[6] G. Ayzel, M. Heistermann, and S. Sorokin, "RainNet v1.0: A convolutional neural network for radar-based precipitation nowcasting," *Geoscientific Model Development*, vol. 13, no. 6, pp. 2631–2644, 2020, doi: 10.5194/gmd-13-2631-2020.

[7] F. J. Gamboa-Villafruela, "Immediate precipitation forecasting using deep learning: A convolutional neural network and long short-term memory approach," *Proceedings*, vol. 8, no. 1, art. 33, 2021, doi: 10.3390/proceedings8010033.

[8] D. Wolfensberger and A. Feinberg, "RainForest: A random forest algorithm for quantitative precipitation estimation over Switzerland," *Atmospheric Measurement Techniques*, vol. 14, pp. 3169–3193, 2021, doi: 10.5194/amt-14-3169-2021.

[9] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[10] X. Shi, "Enabling smart dynamical downscaling of extreme precipitation events with machine learning," *Geophysical Research Letters*, vol. 47, no. 19, e2020GL090309, 2020, doi: 10.1029/2020GL090309.