Edu Komputika 12 (1) (2025)

Edu Komputika Journal

https://journal.unnes.ac.id/journals/edukom/

Performance Evaluation of Machine Learning Methods for Real-Time Rainfall Classification

Rahmi Hidayati✉ and Kartika Sari

Department of Computer System, Faculty of Mathematics and Natural Sciences, Tanjungpura University, Pontianak, 78124, Indonesia

Article Info

Article History:
Received: 1 July 2025
Revised: 11 January 2026
Accepted: 20 January 2026

Keywords:
Classification, Evaluation, Machine Learning, Rainfall

Abstract

Reliable real-time rainfall intensity classification is essential for supporting early warning systems and disaster mitigation, particularly in regions vulnerable to hydrometeorological hazards. This study evaluates three machine learning algorithms SVM, Neural Network, and AdaBoost for multiclass rainfall intensity classification using real-time data collected from Internet of Things (IoT)-based sensors. Rainfall intensity is categorized into four classes: no rain, light rain, moderate rain, and heavy rain, based on threshold values defined by BMKG standards. The dataset is imbalanced and dominated by the no rain class, therefore, model performance is evaluated using imbalance aware metrics, including per-class precision and recall, macro F1-score, balanced accuracy, and overall accuracy. Experimental results show that SVM and Neural Network achieve very high overall accuracy of up to 99.46%, however, this performance is mainly influenced by accurate classification of the majority class, leading to low recall for minority rainfall classes. In contrast, AdaBoost provides a more balanced baseline performance, achieving an accuracy of 92.4% and a macro F1-score of 0.714 on the original dataset. To enhance minority class detection, the SMOTE is applied to the training data using an 80:20 train test split. After data balancing, AdaBoost demonstrates improved recall and macro F1-score for light and moderate rain classes, although overall accuracy decreases to 77.1%. These results are acceptable for early warning applications, where sensitivity to rainfall onset is prioritized over majority class dominance. Consequently, balanced AdaBoost, evaluated using time-based data partitioning and imbalance aware metrics, is considered an effective approach for real-time IoT-based rainfall classification.

✉ Correspondence Address:
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak,
Kalimantan Barat 78124
E-mail: rahmihidayati@siskom.untan.ac.id

ISSN 2252-6811
E-ISSN 2599-297X

INTRODUCTION

Rainfall is one of the key elements of the meteorological system and plays a significant role in influencing human activities and environmental balance (Abdullah et al., 2021; Samad et al., 2021). Variations in rainfall intensity and its spatial and temporal distribution directly affect agriculture, water resource management, transportation, and disaster risk reduction efforts, particularly in areas prone to floods and landslides. In tropical regions such as Indonesia, changes in rainfall patterns driven by climate variability further increase the challenges of maintaining food security and sustainable environmental management (Moharana et al., 2022). Moreover, the increasing intensity and frequency of extreme weather events necessitate rainfall monitoring and classification systems that are adaptive, accurate, and timely. Reliable real-time rainfall classification is therefore essential to support early warning systems, risk-based decision making, and rapid response across various sectors (Fu et al., 2021; Hatwell et al., 2020).

Recent advances in sensor technology and the Internet of Things (IoT) have enabled real-time and continuous rainfall data collection with high temporal resolution (Ghada et al., 2022; Luqman Hakim & Prawito, 2020). However, the availability of large volumes of data is only beneficial when supported by automated classification methods capable of processing streaming data efficiently and accurately. Real-time rainfall classification differs fundamentally from offline data analysis, as it must handle dynamically changing data patterns, limited computational time, and continuously evolving environmental conditions. Consequently, the development and evaluation of machine learning models specifically designed for real-time IoT-based rainfall classification remain a major research challenge (Liden et al., 2023).

Numerous studies have applied machine learning approaches to rainfall prediction and classification, however, several research gaps remain (Hakim & Dewi, 2021; Rumapea et al., 2022a). Most existing studies rely on historical or offline datasets that do not fully represent the characteristics of real-time IoT data streams. These datasets are typically subjected to extensive preprocessing and data balancing, which limits their ability to reflect actual operational conditions in early warning systems. In addition, comparative analyses of algorithms such as AdaBoost (Lakshminarayanan et al., 2021; Shanmugasundar et al., 2021), Support Vector Machine (SVM) (Benmahamed et al., 2021; Moharana et al., 2022), and Artificial Neural Networks (Poedjiastoeti et al., 2024; Torcasio et al., 2024) using real-time IoT-based rainfall data remain limited. As a result, there is still no clear reference for selecting appropriate algorithms for adaptive real-time rainfall classification in early warning systems (Ho et al., 2023). Another issue that has received limited attention is the high class imbalance commonly observed in rainfall datasets. In practical field monitoring, data are generally dominated by no rain conditions, while light, moderate, and heavy rainfall events occur far less frequently. This imbalance can lead to misleading performance evaluations when relying solely on accuracy, as models may appear to perform well overall but fail to detect rare yet critical rainfall events that are crucial for disaster mitigation. Furthermore, imbalanced data conditions influence algorithm behavior, with some methods favoring the majority class, whereas ensemble approaches may provide more stable performance. Unfortunately, many previous studies have not explicitly addressed the impact of class imbalance on model evaluation (Marrocu & Massidda, 2022), nor have they employed more representative metrics such as per-class precision, recall, or macro-averaged F1-score to ensure fair assessment.

Several recent studies have reported promising results in rainfall classification using machine learning techniques. For example, the FCN-LSTM model achieved accuracy levels exceeding 96% (Rumapea et al., 2022). Ensemble methods such as AdaBoost and XGBoost have also demonstrated strong performance in real-time environmental monitoring (Hatwell et al., 2020; Wang et al., 2024). High F1-scores have been reported for SVM, LSTM, and LightGBM models applied to regional weather data (Ho et al., 2023b; Torcasio et al., 2024). However, most of these studies employ relatively balanced datasets or focus on specific regions, providing limited analysis of real-time performance under highly imbalanced conditions. Consequently, their applicability to IoT-based early warning systems remains constrained (Wang et al., 2024).

Based on these challenges, this study aims to evaluate and compare the performance of three machine learning algorithms: AdaBoost, SVM, and Artificial Neural Networks for real-time multiclass rainfall classification using data collected from IoT sensors. Unlike previous studies that rely on static datasets, this research utilizes real-time rainfall data classified into four categories: no rain, light rain, moderate rain, and heavy rain. The analysis focuses on the impact of data imbalance on model performance and the suitability of each algorithm. Model evaluation is conducted using confusion matrices, per-class precision, recall, and F1-score, with the macro-averaged F1-score serving as the primary metric to ensure balanced evaluation across all classes.

The contributions of this study are threefold: the provision of a real-time IoT-based rainfall dataset that reflects actual field conditions with a high degree of class imbalance, a comparative performance analysis of AdaBoost, SVM, and Artificial Neural Networks under imbalanced real-time data scenarios, and recommendations for algorithm selection and appropriate evaluation methods for IoT-based rainfall classification systems. It is expected that the findings of this study will enhance the performance of real-time early warning systems and serve as a reference for future research and development in rainfall classification and disaster risk mitigation.

RESEARCH METHODS

This study employs a systematic workflow to support real-time rainfall classification using IoT-based sensor data. The research stages consist of data collection, labeling in accordance with BMKG standards, data preprocessing, dataset partitioning, class imbalance handling, model training, and performance evaluation. Each stage is described in detail to ensure methodological clarity and reproducibility

A. Data Description

The data used in this study consist of real-time rainfall measurements collected over a 14-day period using IoT-based weather sensors installed in Pontianak City, Indonesia. The sensors operate continuously and record environmental parameters at fixed time intervals, thereby producing time-series data that represent actual field conditions. A total of approximately 3,750 records were obtained, with an average of 215 records per day. Each record includes air temperature (°C), relative humidity (%), rainfall intensity (mm/h), and rainfall status. Rainfall intensity labeling was performed in accordance with the classification standards established by the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG). Hourly rainfall measurements were categorized into four classes: no rain (0 mm/h), light rain (> 0–5 mm/h), moderate rain (> 5–10 mm/h), and heavy rain (> 10 mm/h) (Muflihati et al., 2024).

Table 1. Rainfall Labeling based on BMKG Standards

No	Rainfall Intensity Range	Rainfall Class	Label Code
1.	0	No Rain	0
2.	Greater than 0 – 5	Light Rain	1
3.	Greater than 5 – 10	Moderate Rain	2
4.	Greater than 10	Heavy Rain	3

This threshold-based approach ensures that the labeling process is clear, consistent, and reproducible. All labels are derived directly from rainfall intensity measurements recorded by the sensors, thereby ensuring compliance with national meteorological standards.

The following table presents examples of environmental parameter measurements, including air temperature, humidity, and rainfall intensity, along with their corresponding classifications and labels. The observations indicate that a rainfall intensity of 0.0 mm/h is classified as no rain (label 0), which generally occurs under conditions of higher air temperature and lower humidity. When the rainfall intensity falls within the range of greater than 0 to 5 mm/h, the data are classified as light rain (label 1), typically accompanied by a decrease in air temperature and an increase in humidity. Furthermore, rainfall intensities ranging from greater than 5 to 10 mm/h are classified as moderate rain (label 2), characterized by further reductions in air temperature and increased humidity levels. Rainfall intensities exceeding 10 mm/h are classified as heavy rain (label 3) and are associated with the lowest air temperatures and the highest humidity levels in the sample data. Overall, the table demonstrates a consistent relationship between increasing rainfall intensity, decreasing air temperature, and increasing humidity, and illustrates the application of rainfall labeling based on predefined intensity thresholds.

Table 2. Example of Rainfall Measurements and Labels

Air Tempera-ture (^{0}C)	Humidity (%)	Rainfall Inten-sity (mm/h)	Rainfall Class	Label
29.4	78	0.0	No Rain	0
28.9	81	0.8	Light Rain	1
28.5	83	2.6	Light Rain	1
27.8	85	4.9	Light Rain	1
27.2	87	6.1	Moderate Rain	2
27.0	88	7.4	Moderate Rain	2
26.8	89	9.8	Moderate Rain	2
26.5	90	11.2	Heavy Rain	3
26.3	91	13.7	Heavy Rain	3
26.0	92	18.5	Heavy Rain	3
29.1	76	0.0	No Rain	0
28.7	80	1.3	Light Rain	1
27.6	84	5.4	Moderate Rain	2
26.9	88	10.6	Heavy Rain	3
29.6	74	0.0	No Rain	0

B. Data Preprocessing

The preprocessing stage was conducted to enhance data quality and support optimal model performance. This stage consists of handling missing data, label encoding, and feature normalization. Missing values caused by sensor or communication disruptions were addressed using the forward-fill method, in which missing entries are replaced with the most recent valid observations. This approach is well suited for time-series environmental data, as changes in environmental conditions tend to occur gradually.

Subsequently, categorical rainfall labels were transformed into sequential numerical values, namely 0 for no rain, 1 for light rain, 2 for moderate rain, and 3 for heavy rain. This encoding preserves the ordinal relationships among rainfall intensity levels and supports algorithms that exploit ordinal class information.

Finally, numerical features including air temperature, humidity, and rainfall intensity were normalized using the min–max scaling method to map values into the range of 0 to 1. This normalization was applied to prevent any single feature from dominating the learning process and to improve training stability, particularly for Artificial Neural Network and SVM models.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

The methodological workflow of this research begins with the real-time acquisition of rainfall data through IoT-based sensors. This is followed by a data preprocessing stage, where the raw sensor inputs are cleaned, labeled, and normalized. The dataset is then partitioned into training and testing subsets using an 80:20 split. Three classification algorithms AdaBoost, Neural Network, and SVM, are implemented to model the data. Each algorithm is assessed based on multiple performance metrics, including accuracy, precision, recall, and F1-score. The comparative analysis aims to identify the most effective classifier for real-time rainfall classification. The complete sequence of the research diagram is illustrated in Figure 1.

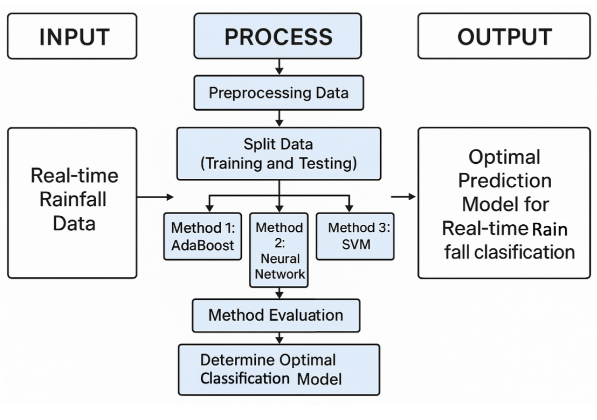

Figure 1. Research diagram

C. Handling Class Imbalance

The dataset used in this study exhibits an imbalanced class distribution, with the no rain class dominating the light, moderate, and heavy rain classes. To address this significant class imbalance, an explicit resampling-based data-balancing procedure was applied, designed to prevent data leakage and ensure reliable model evaluation. The method employed was the Synthetic Minority Over-sampling Technique (SMOTE), selected for its ability to increase the number of samples in minority classes by generating synthetic data through interpolation among nearest-neighbor samples in the feature space. This approach enhances minority class representation without directly duplicating existing data, thereby enabling the model to learn more informative patterns. In its implementation, SMOTE was configured with an auto-sampling strategy to equalize the number of samples in each class to that of the majority class, a nearest-neighbor parameter (k=5), and a fixed random state to ensure experimental reproducibility. The selection of these parameters aims to maintain a balance between improving minority class proportions and preserving feature distribution stability, thereby enabling more accurate and consistent evaluation of model performance.

Data splitting strategy, given that the data are time-series in nature and intended for real-time system deployment, the dataset was partitioned chronologically. The first 80% of the data were used for training, while the remaining 20% were reserved for testing. This strategy was employed to prevent temporal information leakage and to reflect real-world operational conditions, in which the model is trained on historical data and evaluated using more recent observations.

Table 3 presents the distribution of rainfall classes in the entire dataset, which consists of 3,750 samples. The no rain class dominates with 2,480 samples, accounting for 66.13% of the total data, followed by light rain with 850 samples (22.67%). The moderate rain class contains 285 samples (7.7%), while heavy rain has the fewest samples, totaling 135 samples or 3.6% of the dataset. This distribution indicates a significant class imbalance, with the No Rain class substantially outnumbering the other rainfall classes.

Table 3. Rainfall Class Distribution of the Entire Dataset

Rainfall Class	Number of Samples	Percentage
No Rain	2,480	66.13 %
Light Rain	850	22.67 %
Moderate Rain	285	7.7 %
Heavy Rain	135	3.6 %
Total	**3,750**	**100 %**

Table 4 presents the distribution of rainfall classes in the dataset, with the no rain class dominating with 1,984 samples, accounting for 66.13% of the total data. The light rain class is the second most prevalent, with 680 samples (22.67%), followed by moderate rain with 228 samples (7.7%). Meanwhile, the heavy rain class has the fewest samples, totaling 108 samples or 3.6% of the dataset. This distribution indicates a significant class imbalance, with the no rain class substantially outnumbering the other rainfall classes.

Table 4. Rainfall Class Distribution of the Training Data 80% Before Balancing

Rainfall Class	Number of Samples	Percentage
No Rain	1,984	66.13 %
Light Rain	680	22.67 %
Moderate Rain	228	7.7 %
Heavy Rain	108	3.6 %
Total	**3,000**	**100 %**

Table 5 presents the distribution of rainfall classes in the test dataset, in which the no rain class dominates with 496 samples, representing 74.67% of the total data. The light rain class ranks second with 170 samples (14.67%), followed by moderate rain with 57 samples (8%). Meanwhile, the heavy rain class has the fewest samples, totaling 27 samples or 2.66% of the dataset. This distribution indicates a relatively significant class imbalance in the test data, with the no rain class substantially outnumbering the other rainfall classes.

Table 5. Rainfall Class Distribution of the Test Data 20%, Before Balancing

Rainfall Class	Number of Samples	Percentage
No Rain	496	74.67 %
Light Rain	170	14.67 %
Moderate Rain	57	8 %
Heavy Rain	27	2.66 %
Total	**750**	**100 %**

The number of training samples after SMOTE balancing was obtained by oversampling the training dataset to address class imbalance. Following data partitioning, the no rain class contained the largest number of samples, totaling 1,984 instances, and was therefore used as the reference class. Using the SMOTE method, the minority classes, light rain, moderate rain, and heavy rain, were synthetically augmented until each reached the same number of samples as the majority class. As a result, each class contained 1,984 samples, and the total number of training samples after balancing increased to 7,936.

Data balancing using SMOTE was applied exclusively to the training data after the dataset was partitioned chronologically. This strategy was adopted to prevent data leakage, preserve the validity of the model evaluation process, and simulate real-world operational conditions in an early warning system.

D. Model Description

This study investigates three machine learning algorithms, namely AdaBoost, SVM, and Neural Networks. To ensure that the experiments can be independently replicated and meet scientific standards in the field of machine learning, all aspects of the implementation, model configurations, and training protocols are described in a detailed and transparent manner. All experiments were conducted using the Python programming language, with support from widely used machine learning and data analysis libraries, including NumPy, Pandas, Scikit-learn, Imbalanced Learn, Matplotlib, and Seaborn, all in stable and compatible versions. The experiments were executed on a Windows operating system with hardware specifications sufficient to support real-time data processing.

To ensure consistent results and facilitate reproducibility, the same random seed was used across all experimental stages, including data splitting, class balancing with SMOTE, and model initialization. In this study, the random seed was set to 42, ensuring that all processes were deterministic and reproducible within the same computational environment.

The configuration and hyperparameters of each model were determined based on common practices and preliminary experiments and were then kept fixed throughout the comparative evaluation process. The AdaBoost model employed a decision tree-based estimator with a maximum depth of three, 100 estimators, and a learning rate of 1.0. The SVM model was configured with a Radial Basis Function (RBF) kernel, a regularization parameter C=1.0C = 1.0C=1.0, a gamma value set to scale, and a one vs rest multiclass strategy. Meanwhile, the neural network model was implemented as a multilayer perceptron with two hidden layers consisting of 16 and 8 neurons, respectively, using the ReLU activation function, a Softmax output activation, the Adam optimizer, a learning rate of 0.001, a batch size of 32, and a maximum of 100 training epochs.

Model training was performed using the preprocessed training data and, in certain experiments, class balancing with SMOTE. The test data were not used during training or hyperparameter tuning. For the neural network model, an early stopping strategy was applied to prevent overfitting, whereby training was automatically terminated if the validation loss did not improve for 10 consecutive epochs. A total of 10% of the training data was allocated as internal validation data. In contrast, early stopping was not explicitly applied to the AdaBoost and SVM models; however, performance evaluation was consistently conducted using a test dataset that was completely independent of the training process.

All experiments were conducted using fixed random seeds and publicly available machine learning libraries. Detailed implementation settings, hyperparameters, and training protocols are provided to ensure full reproducibility of the reported results.

E. Model Evaluation

The evaluation of model performance in this study was designed to provide a fair and informative assessment of a rainfall dataset with significant class imbalance. Accordingly, a combination of class-based and aggregate evaluation metrics was employed, all of which were computed based on predictions generated from a test dataset that was intentionally kept imbalanced.

The confusion matrix illustrates the correspondence between the model's predictions and the actual conditions of the data. True Positive (TP) represents the number of instances that are predicted as positive and are indeed positive. False Positive (FP) refers to instances that are predicted as positive but are actually negative, and can therefore be considered prediction errors or false alarms. False Negative (FN) indicates instances that are predicted as negative but are actually positive, meaning that the model fails to detect an event. Meanwhile, True Negative (TN) represents instances that are predicted as negative and correctly correspond to the actual condition. The confusion matrix is used to provide a clearer understanding of model performance and serves as the basis for calculating evaluation metrics such as precision, recall, and F1-score.

Table 6. Confusion Matrix

	Actual Positive	**Actual Positive**
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Precision was used to measure the accuracy of positive predictions for each class and is therefore important for assessing the potential occurrence of false alarms, particularly in the context of early warning systems. Recall was used to evaluate the model's ability to detect all instances that truly belong to a given class and is considered a critical metric for minority classes, as missed detections may have significant consequences.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

To balance these two metrics, the per-class F1-score was employed, representing the harmonic mean of precision and recall. A high F1-score indicates that the model is capable of accurately detecting events while simultaneously minimizing prediction errors.

$$\text{F1-Score} = 2\,x\,\frac{\text{Precision x Recall}}{\text{Precision+ Recall}} \quad (5)$$

As the primary evaluation metric, this study emphasizes the use of the macro-averaged F1-score, which is calculated as the unweighted average of F1-scores across all classes. This approach assigns equal importance to each class, making it particularly suitable for evaluating model performance on imbalanced datasets and for assessing performance improvements in minority classes. In contrast, micro-averaged and weighted F1-scores were not adopted as primary metrics because they tend to be dominated by the majority class and are less sensitive to minority class performance. In addition, balanced accuracy was reported as a complementary metric, calculated as the average recall across all classes. This metric provides a more equitable representation of the model's ability to recognize all classes, regardless of the sample distribution.

RESULT AND DISCUSSION

The following are the research results that demonstrate the performance of each classification algorithm for rainfall data classification. Each model exhibited unique strengths and weaknesses in handling the imbalanced dataset. These findings provide important insights for selecting the appropriate algorithm for rainfall classification.

A. Comparison of Rainfall Classification Model Performance

Table 7 presents the accuracy and macro-average F1-score values for the three initial models trained on the original imbalanced dataset. The macro-average F1-score is calculated as the unweighted mean of the F1-scores across all classes, ensuring that both majority and minority rainfall classes contribute equally to the model's performance evaluation.

Table 7. Model Performance on the Imbalanced Dataset

No.	Model	Accuracy	Macro F1-Score
1.	SVM	0.99	0.97
2.	Neural Network	0.99	0.97
3.	AdaBoost	0.92	0.71

The results indicate that the SVM and Neural Network models achieve very high accuracy and macro-average F1 scores. However, a deeper examination using the confusion matrix shows that this performance is predominantly driven by the models' effectiveness in classifying the majority class, no rain. While both models correctly identify nearly all no rain instances, they demonstrate low recall for the minority classes, namely light rain, moderate rain, and heavy rain. This suggests that high accuracy alone does not adequately represent a model's capability to detect rainfall events that are crucial for early warning systems.

The confusion matrix of the SVM model, Figure 2, shows that nearly all of the 496 no rain instances are correctly classified, while only a small portion of the rainfall data is successfully identified, namely 170 light rain, 53 moderate rain, and 27 heavy rain instances.

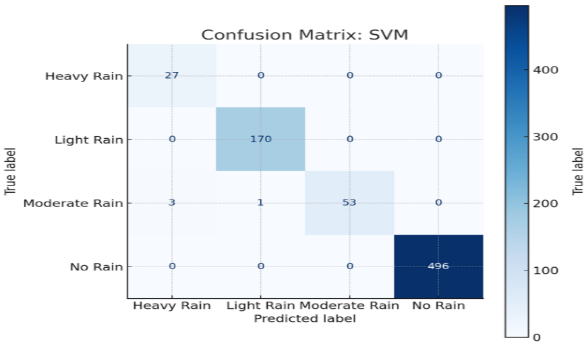

Figure 2. Confusion matrix SVM

The model exhibits a prediction pattern that heavily favors the dominant class, no rain, and shows poor recognition of minority classes. This indicates that data imbalance significantly affects the model's performance. Although the overall accuracy and F1-score may appear high, this performance does not reflect consistent classification across all rainfall categories. These results highlight the importance of implementing data balancing techniques or alternative approaches to improve the model's sensitivity to minority classes, especially in real-time early warning systems, where accurate classification of all rainfall levels is crucial.

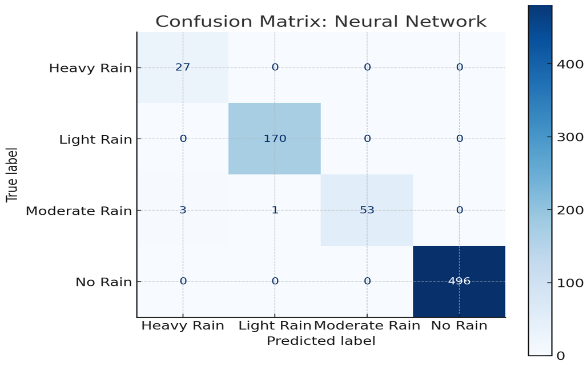

Figure 3. Confusion matrix neural network

The Neural Network model, implemented as a Multilayer Perceptron (MLP), reveals a performance pattern closely resembling that of the SVM model. As illustrated in the confusion matrix, the model excels in correctly classifying the majority class, no rain, with 496 accurate predictions. However, its recognition of minority classes, namely light rain (170 instances), moderate rain (53 instances), and heavy rain (27 instances), is significantly limited. This indicates that the model tends to overfit to the dominant class, which is a common challenge in imbalanced datasets. As a result, although the overall accuracy and F1-score appear high, they largely reflect performance on the majority class and may not provide a complete picture of the model's generalization capability.

In contrast, the AdaBoost model demonstrates a more balanced classification pattern. Its confusion matrix indicates that it is better able to detect variations across multiple rainfall classes, particularly light rain and moderate rain. Despite achieving a lower overall accuracy of 92.4% and an F1-score of 0.714 compared to SVM and Neural Network models, AdaBoost offers relatively better detection performance for minority classes. This more even distribution of predictions suggests that AdaBoost is inherently more resilient to class imbalance, as it iteratively focuses on hard-to-classify instances during training.

These findings underscore the critical importance of selecting appropriate evaluation metrics, particularly in applications with prevalent class imbalance. Relying solely on accuracy can be misleading, as it may mask poor performance on the minority classes, which are often the most crucial in early warning systems. Therefore, metrics such as precision, recall, and the macro-average F1-score should be prioritized when assessing models intended for real-time, risk-sensitive environments, such as rainfall-based disaster alerts.

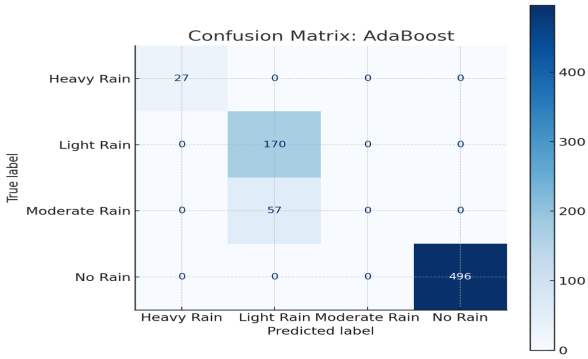

Figure 4. Confusion matrix Adaboost

B. Impact of Class Balancing on Minority Class Performance

To reduce the impact of class imbalance, a resampling-based data balancing technique was applied to the training data. Subsequently, the model performance was reevaluated using the original imbalanced test data. Table 8 presents a comparison of accuracy, macro-average precision, macro-average recall, and macro-average F1-score before and after data balancing.

After the balancing process, all models experienced a decrease in accuracy, falling within the range of 75–77%. This reduction is a natural consequence of diminishing the dominance of the majority class. However, more importantly, the macro average recall and F1-score values show improvements that reflect a fairer recognition of minority classes.

Among all balanced models, the balanced AdaBoost achieved the highest macro-average F1 score of 0.35 and a recall of 0.37. Although these absolute values may appear relatively low, their interpretation is highly significant. The F1-score represents the harmonic mean of precision and recall across all classes. In highly imbalanced, multiclass real-time environmental datasets, these values indicate a substantial improvement in sensitivity to minority classes compared to unbalanced models, which tend to overlook rainfall events.

Table 8. Metrics Comparison per Model

No.	Model	Accuracy	Precision	Recall	Macro F1-Score
1.	SVM	0.99	1.00	1.00	0.97
2.	Neural Network	0.99	0.95	0.94	0.97
3.	AdaBoost	0.92	0.73	0.70	0.71
4.	SVM (Balanced)	0.75	0.35	0.33	0.32
5.	Neural Network (Balanced)	0.76	0.37	0.36	0.34
6.	AdaBoost (Balanced)	0.77	0.38	0.37	0.35

Per-class analysis indicates that the data balancing process provides notable benefits for the light rain and moderate rain classes, as evidenced by consistent improvements in recall following the application of SMOTE. In contrast, the heavy rain class remains the most difficult to detect due to its extremely limited sample size, a common challenge in real-world rainfall monitoring. Consequently, the performance gains observed in the balanced AdaBoost model are not merely descriptive but are quantitatively substantiated by increases in macro average recall and F1-score values for minority classes, which are of primary importance in early warning system applications.

C. Evaluation Using ROC and Precision-Recall Curves

In each confusion matrix heatmap, the main diagonal represents the correctly classified instances for each class, while the off-diagonal cells illustrate the misclassified samples. Visual analysis reveals that both the SVM and Neural Network models exhibit a dominant tendency to classify data into the majority class, no rain, while performing poorly on minority classes such as light rain, moderate rain, and heavy rain. This observation aligns with earlier metric-based findings, where both models demonstrated high overall accuracy and F1-scores but lacked balance in recognizing less frequent rainfall categories.

Among all evaluated models, the Balanced AdaBoost algorithm provides the most consistent performance. It manages to recognize minority classes, particularly light and moderate rain, with greater stability and reduced bias toward the majority class. Despite minor classification errors still present, this model outperforms others in delivering a more balanced prediction output, making it a promising candidate for deployment in real-time rainfall classification systems that demand high sensitivity across diverse rainfall intensities.

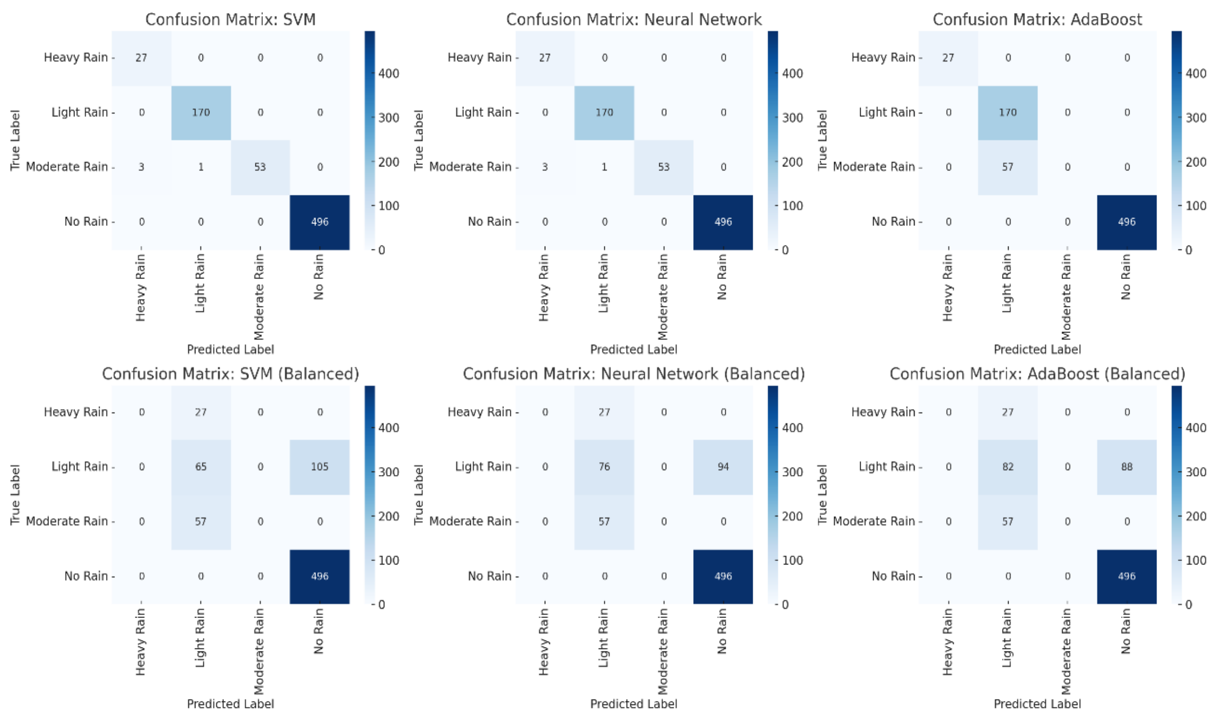

Figure 5. Heatmap confusion matrix per model

Further evaluation using the Receiver Operating Characteristic (ROC) Curve and Precision-Recall Curve reveals critical insights into the models' capabilities to distinguish between different rainfall classes, particularly under imbalanced data conditions. The ROC Curve for the

Balanced AdaBoost model demonstrates its ability to effectively discriminate between multiple classes. This is evidenced by high Area Under the Curve (AUC) values for the majority class and reasonably strong AUC scores for certain minority classes. These results indicate that, compared to the other models, Balanced AdaBoost achieves a more balanced sensitivity across class distributions, enhancing its suitability for real-time early warning systems that require reliable detection of less frequent rainfall events.

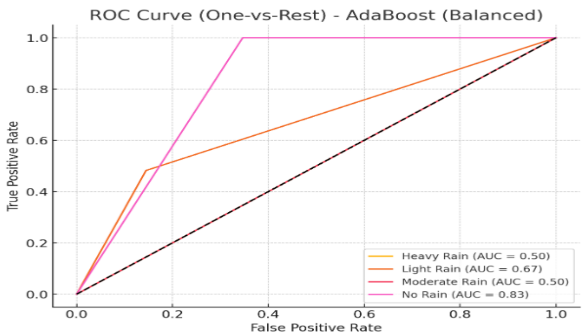

Figure 6. ROC curve Adaboost balanced

The precision-recall curve further confirms the model's ability to maintain both precision and recall for minority classes, which is especially critical in risk-based early warning system applications. This metric is more informative than accuracy in imbalanced datasets, as it highlights the model's capability to correctly identify rare but significant rainfall events, such as light and moderate rain, without being overwhelmed by the dominance of the no rain class.

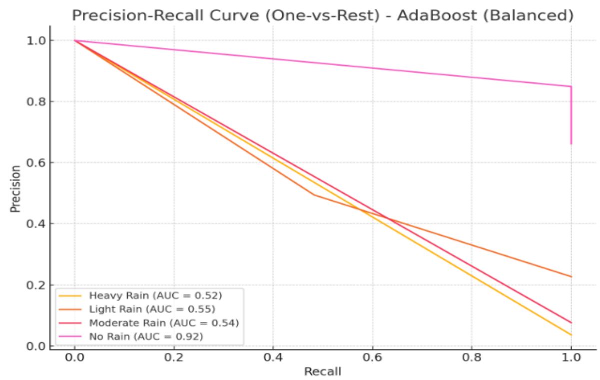

Figure 7. Precision-Recall Curve Adaboost Balanced

Meanwhile, the ROC Curve and Precision Recall Curve for the SVM and Neural Network models indicate that both models perform well in classifying the majority class; however, their performance drops significantly when it comes to minority classes. This observation underscores the importance of applying data balancing techniques and selecting appropriate evaluation metrics when dealing with rainfall datasets, which are often inherently imbalanced.

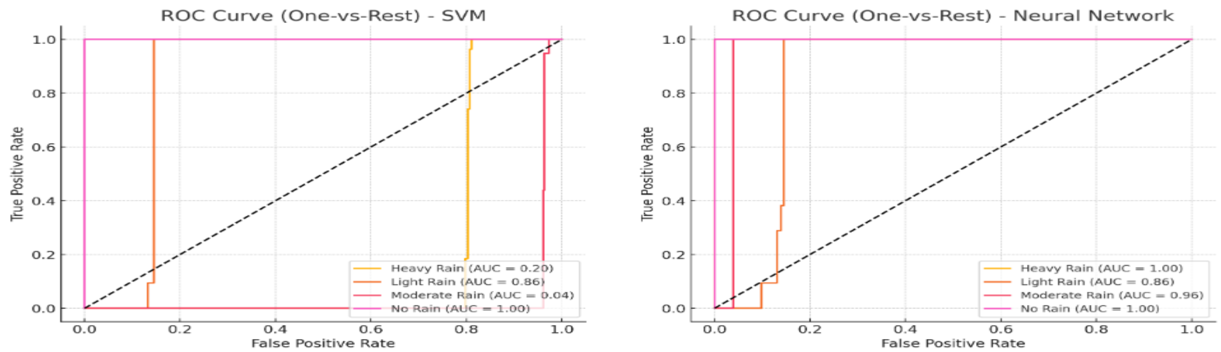

Figure 8. ROC curve SVM and neural network

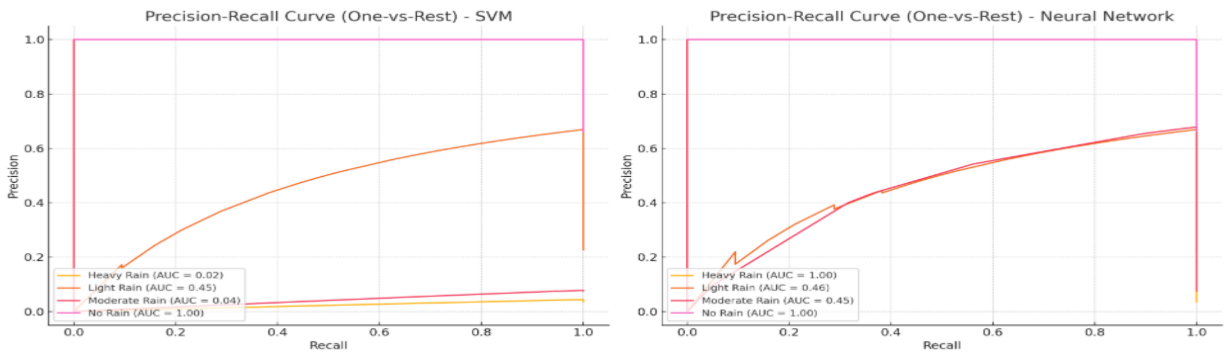

Figure 9. Precision-recall curve SVM and neural network

The SVM and Neural Network models demonstrate strong performance in classifying the dominant rainfall class, namely No Rain. The high accuracy and F1-score values achieved by both models are primarily attributed to their consistent success in predicting the majority class. However, their performance declines significantly when identifying minority classes, namely Light Rain, Moderate Rain, and Heavy Rain. This condition indicates the presence of class imbalance bias, where the models tend to overlook infrequent rainfall events, even though such events often carry greater importance in early warning systems.

Without the application of data balancing techniques or adjustments to the decision threshold, the SVM and Neural Network models are unable to generalize effectively across the full range of rainfall intensities. In contrast, the balanced AdaBoost model exhibits more uniform classification performance. It distributes predictions more proportionally across the four rainfall categories, thereby improving the detection rates of minority classes without substantially compromising overall accuracy.

The ability to achieve balanced classification is particularly critical in real-world applications, especially for rare but high-impact rainfall events, such as moderate to heavy rain, which play a vital role in disaster risk reduction and emergency preparedness. A comparative analysis of performance metrics, including precision, recall, and F1-score, demonstrates that the balanced AdaBoost model has superior capability for handling class imbalance. Consequently, this model is considered more suitable for deployment in real-time rainfall classification systems with imbalanced data distributions.

Unlike conventional accuracy-based evaluations that tend to emphasize the majority class, the use of the balanced AdaBoost model enables more equitable recognition across all rainfall categories, particularly the minority classes that are central to early warning system applications. These findings underscore the importance of adopting class-sensitive evaluation metrics and data-balancing strategies in developing machine learning models that are not only high-performing but also robust, fair, and reliable for supporting critical decision-making in environmental monitoring contexts.

D. Interpretation and Practical Implications

Overall, the findings of this study indicate that machine learning models trained on highly imbalanced rainfall datasets can produce deceptively high accuracy while failing to effectively detect minority rainfall events. This limitation makes such models unsuitable for early warning systems, where timely identification of rainfall, particularly light to moderate precipitation, is often more critical than accurately predicting non-rain conditions.

The balanced AdaBoost model emerges as the most suitable approach, not due to superior accuracy, but because it achieves the most effective trade-off between overall performance and sensitivity to minority classes. This advantage is quantitatively supported by improvements in per-class recall and macro average F1-score, metrics that closely align with the operational demands of early warning systems.

These results underscore the importance of adopting class distribution-sensitive evaluation metrics, such as per-class recall and macro average F1-score, alongside data balancing strategies in the development of rainfall classification models. By explicitly identifying which classes benefit from balancing and examining recall improvements among minority classes, this study offers a transparent and data-driven rationale for selecting the balanced AdaBoost model as the most appropriate solution for real-time IoT-based rainfall early warning systems.

CONCLUSION

This study evaluates the performance of AdaBoost, SVM, and Neural Network algorithms for real-time rainfall classification using imbalanced data collected from IoT sensors. The evaluation aligns with the requirements of an early warning system, particularly the ability to detect rainfall events early and identify rare events. Therefore, the performance assessment emphasizes per-class recall and the macro-averaged F1-score, which are more appropriate than overall accuracy for imbalanced datasets. The results show that SVM and Neural Network achieve high accuracy and macro-averaged F1-scores; however, their performance is largely dominated by predictions of the majority class (no rain). The low recall values for minority classes, especially light rain and moderate rain, indicate the limitations of these models for early warning systems, as they risk missing rainfall events.

In contrast, AdaBoost demonstrates a more balanced performance across rainfall classes. After applying data balancing techniques, the Balanced AdaBoost model achieves the highest recall and macro-averaged F1-score, indicating improved sensitivity to rainfall events in minority classes. Although this improvement is accompanied by a decrease in overall accuracy, it is considered acceptable because rainfall detection is prioritized in early warning systems. The superiority of the Balanced AdaBoost model is contextual and based on the study's objectives and evaluation criteria, rather than representing the best model for all conditions. This model provides the best trade-off between sensitivity and reliability for real-time rainfall classification under imbalanced data conditions. This study has several limitations, including a short data collection period of 2 weeks, extreme class imbalance, particularly in the heavy rain class, and an evaluation based on a single data split and a single balancing method. Future work is recommended to extend the data collection period and geographical coverage, explore alternative imbalance handling approaches, and apply online learning techniques to improve the reliability of early warning systems.

ACKNOWLEDGEMENTS

The authors declare that this study was conducted without any conflict of interest.

REFERENCES

Abdullah, A. S., Ruchjana, B. N., Jaya, I. G. N. M., & Soemartini. (2021). Comparison of SARIMA and SVM model for rainfall forecasting in Bogor city, Indonesia. *Journal of Physics: Conference Series*, *1722*(1). https://doi.org/10.1088/1742-6596/1722/1/012061

Benmahamed, Y., Kherif, O., Teguar, M., Boubakeur, A., & Ghoneim, S. S. M. (2021). Accuracy improvement of transformer faults diagnostic based on DGA data using SVM-BA classifier. *Energies*, *14*(10). https://doi.org/10.3390/en14102970

Fu, W., Yu, S., & Wang, X. (2021). A novel method to determine basic probability assignment based on adaboost and its application in classification. *Entropy*, *23*(7). https://doi.org/10.3390/e23070812

Ghada, W., Casellas, E., Herbinger, J., Garcia-Benadí, A., Bothmann, L., Estrella, N., Bech, J., & Menzel, A. (2022). Stratiform and Convective Rain Classification Using Machine Learning Models and Micro Rain Radar. *Remote Sensing*, *14*(18). https://doi.org/10.3390/rs14184563

Hakim, A. L., & Dewi, R. (2021). Automatic Rain Detection System Based on Digital Images of CCTV Cameras Using Convolutional Neural Network Method. *IOP Conference Series: Earth and Environmental Science*, *893*(1). https://doi.org/10.1088/1755-1315/893/1/012048

Hatwell, J., Gaber, M. M., & Atif Azad, R. M. (2020). Ada-WHIPS: Explaining AdaBoost classification with applications in the health sciences. *BMC Medical Informatics and Decision Making*, *20*(1). https://doi.org/10.1186/s12911-020-01201-2

Ho, J., Zhang, G., Bukovcic, P., Parsons, D. B., Xu, F., Gao, J., Carlin, J. T., & Snyder, J. C. (2023). Improving Polarimetric Radar-Based Drop Size Distribution Retrieval and Rain Estimation Using a Deep Neural Network. *Journal of Hydrometeorology*, *24*(11), 2057–2073. https://doi.org/10.1175/JHM-D-22-0166.1

Lakshminarayanan, K., Muthukumaran, N., Robinson, Y. H., Shanmuganathan, V., Kadry, S., & Nam, Y. (2021). Deep Learning-Based Hookworm Detection in Wireless Capsule Endoscopic Image Using AdaBoost Classifier. *Computers, Materials and Continua*, *67*(3), 3045–3055. https://doi.org/10.32604/cmc.2021.014370

Liden, P., Stenvall, B., & Thidevall, N. (2023). Proposed flood risk classification of buildings - A case study with a heavy rain event. *Journal of Physics: Conference Series*, *2654*(1). https://doi.org/10.1088/1742-6596/2654/1/012017

Luqman Hakim, A., & Prawito. (2020). Rain detection in image using convolutional neural network. *Journal of Physics: Conference Series*, *1528*(1). https://doi.org/10.1088/1742-6596/1528/1/012010

Marrocu, M., & Massidda, L. (2022). Coupling a Neural Network with a Spatial Downscaling Procedure to Improve Probabilistic Nowcast for Urban Rain Radars. *Forecasting*, *4*(4), 845–865. https://doi.org/10.3390/forecast4040046

Moharana, L., Sahoo, A., & Ghose, D. K. (2022). Prediction of Rainfall Using Hybrid SVM-HHO Model. *IOP Conference Series: Earth and Environmental Science*, *1084*(1). https://doi.org/10.1088/1755-1315/1084/1/012054

Muflihati, H., Nirmala, I., Hidayati, R., Rekayasa,), & Komputer, S. (2024). Pemantauan Cuaca Lokal Dengan Sistem Pendeteksi

Curah Hujan Berbasis IoT. *Jurnal Teknologi Informasi*, *5*(3). https://doi.org/10.46576/djtechno

Poedjiastoeti, H., Adi, H. P., & Wahyudi, R. S. (2024). Prediction of the Hydrological Drought Index (HDI) based on Rain Simulation Data using an Artificial Neural Network (ANN) in the Randugunting Watershed, Rembang Regency. *IOP Conference Series: Earth and Environmental Science*, *1321*(1). https://doi.org/10.1088/1755-1315/1321/1/012013

Rumapea, H., Zarlis, M., Sihombing, P., Efendi, S., Sinambela, M., & Saragih, I. J. A. (2022). Convective Cloud Classification Model for Reconstruction of Heavy Rain That Triggers the Flood and Landslide in Parapat, North Sumatera. *IOP Conference Series: Earth and Environmental Science*, *1083*(1). https://doi.org/10.1088/1755-1315/1083/1/012017

Samad, M. A., Diba, F. D., & Choi, D. Y. (2021). Rain attenuation scaling in south korea: Experimental results and artificial neural network. *Electronics (Switzerland)*, *10*(16). https://doi.org/10.3390/electronics10162030

Shanmugasundar, G., Vanitha, M., Čep, R., Kumar, V., Kalita, K., & Ramachandran, M. (2021). A comparative study of linear, random forest and adaboost regressions for modeling non-traditional machining. *Processes*, *9*(11). https://doi.org/10.3390/pr9112015

Torcasio, R. C., Papa, M., Del Frate, F., Mascitelli, A., Dietrich, S., Panegrossi, G., & Federico, S. (2024). Data Assimilation of Satellite-Derived Rain Rates Estimated by Neural Network in Convective Environments: A Study over Italy. *Remote Sensing*, *16*(10). https://doi.org/10.3390/rs16101769

Wang, X., Zhang, C., Qiang, Z., Xu, W., & Fan, J. (2024). A New Forest Growing Stock Volume Estimation Model Based on AdaBoost and Random Forest Model. *Forests*, *15*(2). https://doi.org/10.3390/f15020260