



Preliminary Analysis of Machine Learning Performance and the Effect of Outliers in Daily Rainfall Classification in Jambi City

Muhammad Risyad Naufal*, Erni, and Marathur Rodhiyah

Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Sriwijaya, South Sumatera, 30662, Indonesia

*Corresponding Author: nuafal@unsri.ac.id

ARTICLE INFO

Article history:

Received 05 February 2026

Revised 17 February 2026

Accepted 03 March 2026

Available online 10 March 2026

E-ISSN: 2656-0755

P-ISSN: 2656-0747

How to cite:

M. R. Naufal, Erni, and M. Rodhiyah, "Preliminary Analysis of Machine Learning Performance and the Effect of Outliers in Daily Rainfall Classification in Jambi City," *Journal of Technomaterial Physics*, vol. 08, no. 01, pp. 8-19, Feb. 2026, doi: 10.32734/jotpv8i1.24702.

ABSTRACT

Rainfall is a crucial meteorological parameter that significantly affects various sectors, particularly in tropical regions such as Jambi City. However, daily rainfall data often contain outliers and imbalanced class distributions, which can degrade the performance of machine learning-based classification models. This study aims to conduct a preliminary analysis of the performance of several machine learning algorithms for daily rainfall classification in Jambi City by examining the effects of outlier removal. The algorithms evaluated include Support Vector Machine (RBF), K-Nearest Neighbor, Naive Bayes, Decision Tree, and Random Forest. Model performance was assessed using accuracy and macro F1-score metrics. The rainfall classes used in this study consist of four categories: no rain, light rain, moderate rain, and heavy rain. The results indicate that outlier removal improves the accuracy of all evaluated algorithms, with the most substantial improvement observed in the Decision Tree model with accuracy improved from 45.71% to 57.36% and macro F1-score from 28.99% to 38.78%. Overall, the implementation of outlier removal yields more balanced and representative rainfall classification results, potentially serving as a basis for future quantitative rainfall regression studies.

Keywords: Classification, Daily Rainfall, Machine Learning, Macro F1-Score, Outlier Removal

ABSTRAK

Curah hujan merupakan parameter meteorologi penting yang memiliki pengaruh signifikan terhadap berbagai sektor, terutama di wilayah tropis seperti Kota Jambi. Namun, data curah hujan harian sering mengandung pencilan dan distribusi kelas yang tidak seimbang, yang berpotensi menurunkan kinerja model klasifikasi berbasis pembelajaran mesin. Penelitian ini bertujuan untuk melakukan analisis pendahuluan kinerja beberapa algoritma pembelajaran mesin dalam klasifikasi curah hujan harian di Kota Jambi dengan meninjau pengaruh penghapusan pencilan. Algoritma yang digunakan meliputi *Support Vector Machine* (RBF), *K-Nearest Neighbor*, *Naive Bayes*, *Decision Tree*, dan *Random Forest*. Evaluasi model dilakukan menggunakan metrik akurasi dan skor F1 macro. Curah hujan dalam penelitian ini diklasifikasikan ke dalam empat kategori, yaitu tidak hujan, hujan ringan, hujan sedang, dan hujan lebat. Hasil penelitian menunjukkan bahwa penghapusan pencilan meningkatkan akurasi semua algoritma yang dievaluasi, dengan peningkatan paling signifikan terjadi pada model *Decision Tree*, di mana akurasi meningkat dari 45,71% menjadi 57,36% dan skor F1 Macro dari 28,99% menjadi 38,78%. Secara keseluruhan, implementasi penghapusan pencilan menghasilkan klasifikasi curah hujan yang lebih seimbang dan representatif, berpotensi menjadi dasar bagi penelitian regresi curah hujan kuantitatif di masa mendatang.

Kata kunci: Curah Hujan Harian, Klasifikasi, Pembelajaran mesin, Penghapusan Pencilan, Skor F1 Macro



1. Introduction

Jambi City is located in the eastern part of Sumatra within a lowland area dominated by Quaternary alluvial deposits. Geologically, the city is situated in the southern Sumatra Basin, which was formed by regional tectonic activity on Sumatra Island [1]. The hydrometeorological characteristics of Jambi City are influenced by its geological setting, flat topography, and the presence of major rivers such as the Batanghari River. Its humid tropical climate, characterized by relatively high rainfall and considerable weather variability throughout the year, represents one of the defining features of the city [2].

Rainfall is one of the most important meteorological parameters influencing many aspects of life, including agriculture, water resource management, flood disaster mitigation, and spatial planning [3], [4]. However, rainfall characteristics are highly variable and are influenced by numerous atmospheric factors, which makes modeling and analysis more complex [5], particularly in tropical regions such as Jambi City. In addition, rainfall data often contain extreme values or outliers, which may arise from very intense rainfall events, measurement errors, or inconsistencies in data recording [6]. All of these factors can significantly affect the performance of data-driven analytical models.

Rainfall forecasting remains highly important and has attracted the attention of governments, companies, risk management institutions, and academic organizations [7]. Machine learning approaches have been widely applied to analyze and model meteorological data, including rainfall [8]. In predictive analysis, machine learning algorithms play a crucial role by utilizing patterns from current and historical data to forecast future events [9], [10]. The performance of machine learning methods is often superior to that of conventional statistical techniques, as they are capable of identifying nonlinear relationships between various meteorological parameters and rainfall [11]. In operational studies, rainfall classification into specific categories is a commonly used approach [12]. One example is the classification of daily rainfall based on the standards of the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) [13]. Such classification methods are highly relevant for early warning systems that depend on rainfall intensity categories [14].

Nevertheless, the quality of the data used, including the presence of outliers, strongly affects statistical descriptions [15]. This issue also influences the performance of machine learning algorithms in rainfall classification. Outlier values can introduce bias during the model training process and reduce classification accuracy, particularly in datasets with imbalanced class distributions [16]. Therefore, a preliminary investigation is required to examine how the presence and removal of outliers affect the performance of various machine learning algorithms in daily rainfall classification.

In several previous studies, machine learning techniques have been applied to analyze and model rainfall data in various regions. In Malaysia, multiple algorithms such as Support Vector Machine (SVM), Random Forest, and Decision Tree have been employed [17]. Similar studies have been conducted in India using K-Nearest Neighbor (KNN), Random Forest, SVM, and Decision Tree methods [18]. Furthermore, in the Ischia Island region of Italy, a specialized machine learning approach using Random Forest was applied and proven effective for short-term rainfall prediction [19]. Machine learning has also been widely utilized in disaster-related studies associated with rainfall, such as flood prediction [20], and landslide susceptibility analysis [21]. In Indonesia, rainfall prediction studies have been conducted in several regions, including Bogor [22] and South Sumatra [23].

Based on this background, this study aims to conduct a preliminary evaluation of several machine learning algorithms for daily rainfall classification in Jambi City, with particular attention to the influence of outliers. The classification methods applied include SVM, KNN, Naive Bayes, Decision Tree, and Random Forest. Model performance is assessed using accuracy and macro F1-score to provide a comprehensive evaluation, especially under imbalanced class conditions. The effect of outlier removal on classification accuracy is further examined using a paired t-test to determine statistical significance, while Cohen's *d* is calculated to measure the magnitude of the outlier effect and the Shapiro–Wilk test is employed to verify normality assumptions. This study assumes that meteorological variables such as air temperature, humidity, atmospheric pressure, cloud cover, and wind speed are correlated with rainfall intensity and that machine learning models can learn these relationships from historical data. However, this study intentionally uses rainfall as a single feature to evaluate the statistical behavior of rainfall classification and the isolated impact of outlier removal. Other parameters will be incorporated in future multivariate modeling. The analysis is also limited to daily-scale classification, without detailed examination of within-class rainfall variability, and is intended as a preliminary foundation for future studies that adopt regression-based approaches for quantitative rainfall prediction.

2. Methodology

2.1. Data and Data Sources

The data used in this study were obtained from daily meteorological observations for Jambi City, sourced from the Online Data portal of the Indonesian Agency for Meteorology, Climatology, and Geophysics [24] (<https://dataonline.bmkg.go.id/>). This platform provides historical weather data based on observational records and reanalysis products and is widely used in climatological research and weather modeling studies. The collected dataset includes daily meteorological parameters that represent surface atmospheric conditions and can influence rainfall. Although the data source contains multiple variables, daily rainfall (RR) is the primary variable used in this preliminary machine learning analysis to evaluate the performance of each algorithm. The data span the period from 1 February 2024 to 31 December 2025 and are analyzed at a daily temporal resolution to ensure consistency in the analysis.

2.2. Data Preprocessing

Before the data were used for machine learning modeling, a preprocessing stage was conducted to ensure data quality. The first step involved selecting numerical variables relevant to the research objectives. In addition, missing values and data inconsistencies were examined. To avoid bias in the model training process, incomplete or invalid data were removed. The standards of the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) were then applied to classify daily rainfall, which was originally continuous data measured in millimeters, into rainfall intensity categories. This classification was intended to align the analysis with operational needs and categorical interpretation. The rainfall categories used were no rain (< 1 mm), light rain (1 to < 5 mm), moderate rain (5 to < 10 mm), heavy rain (10 to < 20 mm), very heavy rain (20 to < 50 mm), and extreme rain (> 50 mm) [13].

Subsequently, the dataset was divided into training and testing sets using the train–test split method, with 80% of the data allocated for training and 20% for testing. Feature standardization was performed using the StandardScaler to ensure that all features had comparable scales and to prevent any single feature from dominating the model learning process. The StandardScaler was applied only to the predictor features, while the target rainfall category label was not scaled because it represents a categorical classification output.

2.3. Outlier Detection and Removal

The performance of machine learning algorithms can be affected by the presence of outliers in rainfall data [25], particularly during pattern learning and decision-making processes. Therefore, the influence of outliers on the performance of classification models is analyzed in this study. In this study, outliers refer to statistical anomalies detected using the IQR method. These observations are not necessarily meteorological extreme rainfall events but values that lie outside the statistical distribution and may degrade machine learning model stability.

To identify outliers, the Interquartile Range (IQR) method was applied to the daily rainfall variable. The IQR is calculated using Equation (1) [26].

$$\begin{aligned} \text{IQR} &= Q_3 - Q_1 \\ \text{Outlier} &= \begin{cases} Q_3 + 1.5 \times \text{IQR} \\ Q_1 - 1.5 \times \text{IQR} \end{cases} \end{aligned} \quad (1)$$

Where Q_1 and Q_3 represent the first quartile and third quartile of the rainfall distribution, respectively. Values that fall outside the upper and lower IQR boundaries are considered outliers. After the outlier detection process, the dataset was divided into two analytical scenarios: data before outlier removal and data after outlier removal. Both scenarios were used to assess changes in model performance attributable to the presence of outliers.

2.4. Machine Learning Algorithms

This study employs several machine learning algorithms that are commonly used for classification problems. To ensure a fair performance comparison, all algorithms were trained and tested using the same data partitioning scheme. In addition, the effects of data conditions before and after outlier removal were considered for each algorithm. Support Vector Machine (SVM) is a supervised learning algorithm commonly used for classification tasks and is capable of modeling nonlinear relationships through the use of a radial basis function (RBF) kernel [27]. K-Nearest Neighbor (KNN) is a non-parametric classification method that assigns class labels based on the distance to the nearest samples in the feature space and is often employed as a baseline due to its simplicity and ease of implementation [28]. Naive Bayes is a probabilistic classifier derived from Bayes' theorem that assumes conditional independence among features given the class label and typically performs

well on datasets with moderate dimensionality [29]. The Decision Tree algorithm is an interpretable method that captures nonlinear relationships by recursively partitioning data into a tree-like structure and can handle both numerical and categorical variables [30]. Random Forest extends the Decision Tree approach by combining multiple trees in an ensemble framework to improve prediction stability and accuracy, while random feature selection at each split helps reduce overfitting and enhance generalization performance [31]. Each model used commonly accepted baseline parameters to ensure comparability and avoid overfitting due to excessive tuning. KNN used $k=5$, SVM used RBF kernel with default γ and C , and Decision Tree depth was unconstrained.

2.5. Evaluation Metrics

To evaluate the performance of the classification models, the two primary metrics used in this study are accuracy and the macro F1-score. Accuracy measures the overall proportion of correctly predicted instances, whereas the macro F1-score evaluates classification performance for each class independently and then averages the results across all classes. To compute the macro F1-score, the values of recall and precision are first calculated using Equations (2) and (3), respectively [32].

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

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

With TP denotes True Positive, FN denotes False Negative, and FP denotes False Positive. Accordingly, the macro F1-score is calculated using Equation (4) as follows [32].

$$F1_{macro} = \frac{1}{N} \sum_{i=1}^N \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (4)$$

Where N represents the number of classes and i indicates the class index. The macro F1-score was selected because the rainfall dataset exhibits class imbalance, where the ‘‘Heavy Rain’’ class dominates the observations. Unlike accuracy, the macro F1-score evaluates model performance equally across all classes by averaging the F1 value of each class without weighting by class frequency. Therefore, it provides a more reliable assessment of the model’s ability to detect minority rainfall events.

In addition, a confusion matrix is used to provide a more detailed representation of classification errors across rainfall categories [33]. To examine the impact of outliers on the performance of each machine learning algorithm, model results obtained before and after outlier removal are systematically compared.

2.6. Statistical Methodology Tests

Prior to conducting comparative tests, the normality assumption of the differences in model performance was evaluated using the Shapiro–Wilk test, which is an effective statistical method for detecting normality, particularly for small to medium sample sizes [34]. The differences in accuracy were found to be normally distributed ($p > 0.05$), allowing the use of parametric tests. Subsequently, a paired t-test was applied to determine whether there was a statistically significant difference between the mean model accuracies before and after the preprocessing treatment. The paired t-test evaluates two related conditions to test the null hypothesis that no meaningful difference exists between them [35]. The paired t-test was conducted using five paired observations corresponding to the five evaluated machine learning algorithms. In addition, Cohen’s d effect size was reported to quantify the practical magnitude of the observed performance differences by measuring the standardized difference between the paired conditions, indicating the magnitude of performance difference between model accuracies before and after outlier removal [36].

3. Results and Analysis

3.1. Data and Comparison Between Actual and Predicted Rainfall

The visualization results comparing actual and predicted rainfall indicate that all algorithms are predominantly influenced by the No Rain and Light Rain classes throughout the observation time index. Moderate Rain and Heavy Rain events occur far less frequently and appear sporadically, reflecting the natural characteristics of daily rainfall patterns in Jambi City and indicating a strong class imbalance in the dataset. This imbalance directly affects model prediction behavior, where most algorithms tend to achieve higher accuracy for majority classes compared to minority classes with low occurrence frequencies.

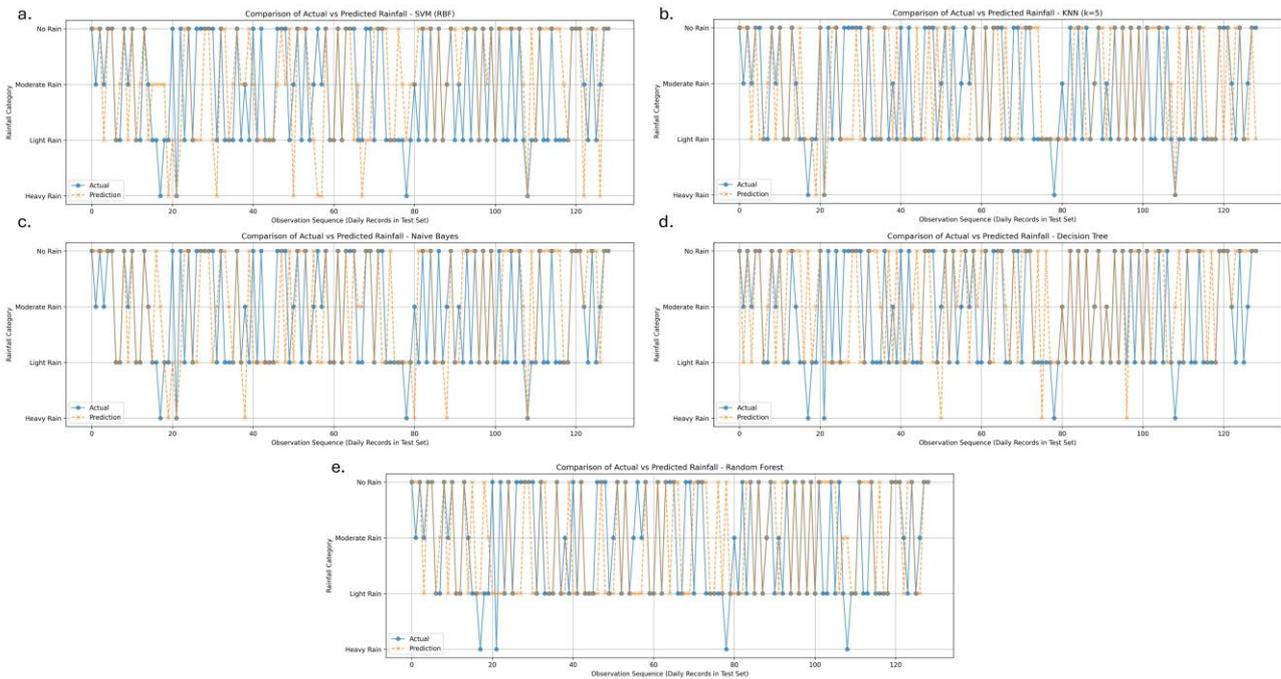


Figure 1. Comparison between actual and predicted rainfall using the a. SVM (RBF), b. KNN, c. Naive Bayes, d. Decision Tree, e. Random Forest.

Figure 1a shows the prediction results of the Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel. The model demonstrates relatively good performance in predicting the No Rain and Light Rain classes but still struggles to distinguish between Moderate Rain and Heavy Rain. Several high-intensity rainfall events are mapped to lower-intensity classes, indicating that although SVM is effective in global class separation, its sensitivity to minority classes remains limited in this dataset.

The KNN results in Figure 1b show a tendency for smoother predictions that closely follow the majority classes. At several time indices, Moderate Rain and Heavy Rain events are predicted as Light Rain or No Rain. The reliance of KNN on distance-based similarity causes classes with large sample sizes to dominate the classification process, thereby limiting sensitivity to high-intensity rainfall classes.

Figure 1c presents the results of the Naïve Bayes algorithm. The prediction pattern appears the simplest and is heavily clustered around the No Rain and Light Rain classes. Moderate Rain and Heavy Rain events are rarely predicted correctly. This behavior is associated with the conditional independence assumption of Naïve Bayes, which is less suitable for meteorological data where atmospheric variables are typically physically correlated.

As shown in Figure 1d, the comparison between actual and predicted rainfall again highlights the dominance of the No Rain and Light Rain classes across the observation period. Moderate Rain and Heavy Rain events remain sparse and irregular. This further emphasizes the strong class imbalance in the dataset and its direct impact on model predictions, where the Decision Tree algorithm tends to favor majority classes over low-frequency rainfall categories.

The Random Forest model in Figure 1e exhibits the most visually stable performance among all evaluated algorithms. Its predictions more consistently follow the actual rainfall data, including several Moderate Rain and Heavy Rain events. The ensemble nature of Random Forest, which combines multiple decision trees, enhances robustness against data variability and residual outlier effects. Random Forest demonstrated greater robustness because ensemble learning combines multiple decision trees trained on bootstrap samples. This reduces variance and limits the influence of extreme observations. Individual trees may be affected by outliers, but aggregation through majority voting stabilizes the final prediction. Nevertheless, a bias toward majority classes is still observable.

The visual comparison between actual and predicted rainfall categories indicates that all models are able to capture the general rainfall pattern, although mismatches still occur, particularly between the “Light Rain” and “No Rain” classes. This observation is consistent with the quantitative evaluation results. Before outlier removal, the model accuracy ranged from 43.57% to 52.14% and for macro F1-score is 28.13%–39.34%. These numerical findings confirm that the clearer alignment between actual and predicted labels seen in the plots.

Overall, these results confirm that accuracy alone is insufficient for evaluating the performance of daily rainfall classification models, particularly when applied to imbalanced datasets. The macro F1-score and visual analysis reveal that most algorithms continue to experience difficulty in detecting higher-intensity rainfall events. Therefore, this study emphasizes the importance of preliminary analysis to understand the influence of outliers and class distribution on machine learning performance before proceeding to more advanced model development.

3.2. Outlier Detection

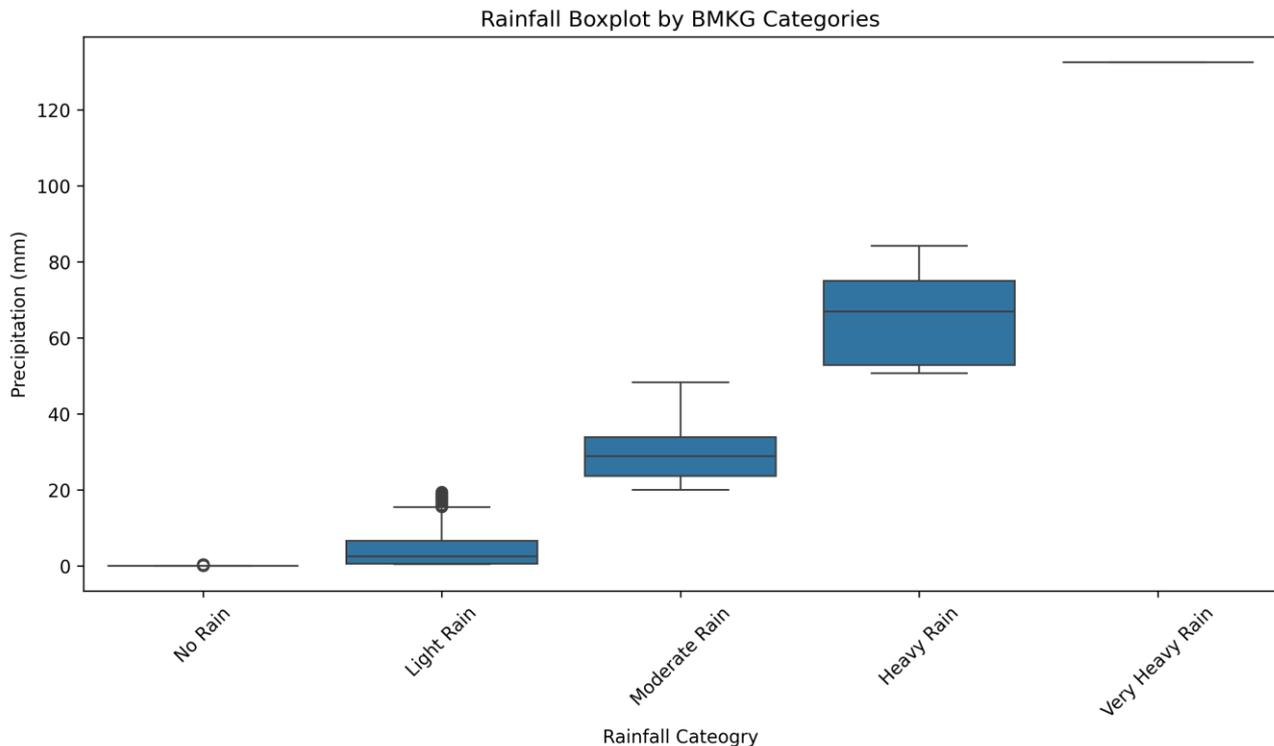


Figure 2. Boxplot visualization for identifying outliers in no rain and light rain data.

The rainfall boxplots by BMKG category in Figure 2 show distinct distribution characteristics across classes. The No Rain category exhibits tightly clustered values near zero with low variability, consistent with minimal precipitation conditions. Light Rain displays the highest variability and a strong right-skewed distribution, with numerous outliers indicating relatively intense rainfall events that still fall within this category, a common feature of tropical rainfall variability in Jambi City. Moderate Rain shows a more symmetric distribution with a moderate interquartile range and fewer outliers, suggesting relatively stable rainfall intensity. Heavy Rain is characterized by higher median values and wider variability but without a dominance of extreme outliers, indicating that BMKG thresholds effectively capture high-intensity rainfall. In contrast, Very Heavy Rain contains only a few extreme observations, resulting in a limited boxplot and highlighting the rarity of such events and the presence of class imbalance. From a modeling perspective, the concentration of outliers in the Light Rain class and differing variability across categories can affect classification performance, particularly for distance- and distribution-sensitive algorithms such as KNN and SVM, while tree-based methods are more robust. Therefore, careful outlier treatment is essential to improve model performance while preserving important climatological information on extreme rainfall events.

3.3. Confusion Matrix

The confusion matrices before and after outlier removal (Figures 3 and 4) show clear shifts in classification behavior across all algorithms. Before data cleaning, most models were able to identify the Light Rain class, though with high misclassification, mainly toward the No Rain class. After outlier removal, predictions across nearly all algorithms became increasingly dominated by the No Rain class, accompanied by a general decline in the recognition of Light Rain, Moderate Rain, and Heavy Rain. For SVM (RBF), outlier removal increased correct No Rain predictions but substantially increased the misclassification of Light Rain as No Rain, indicating a more conservative, majority-class-oriented decision boundary.

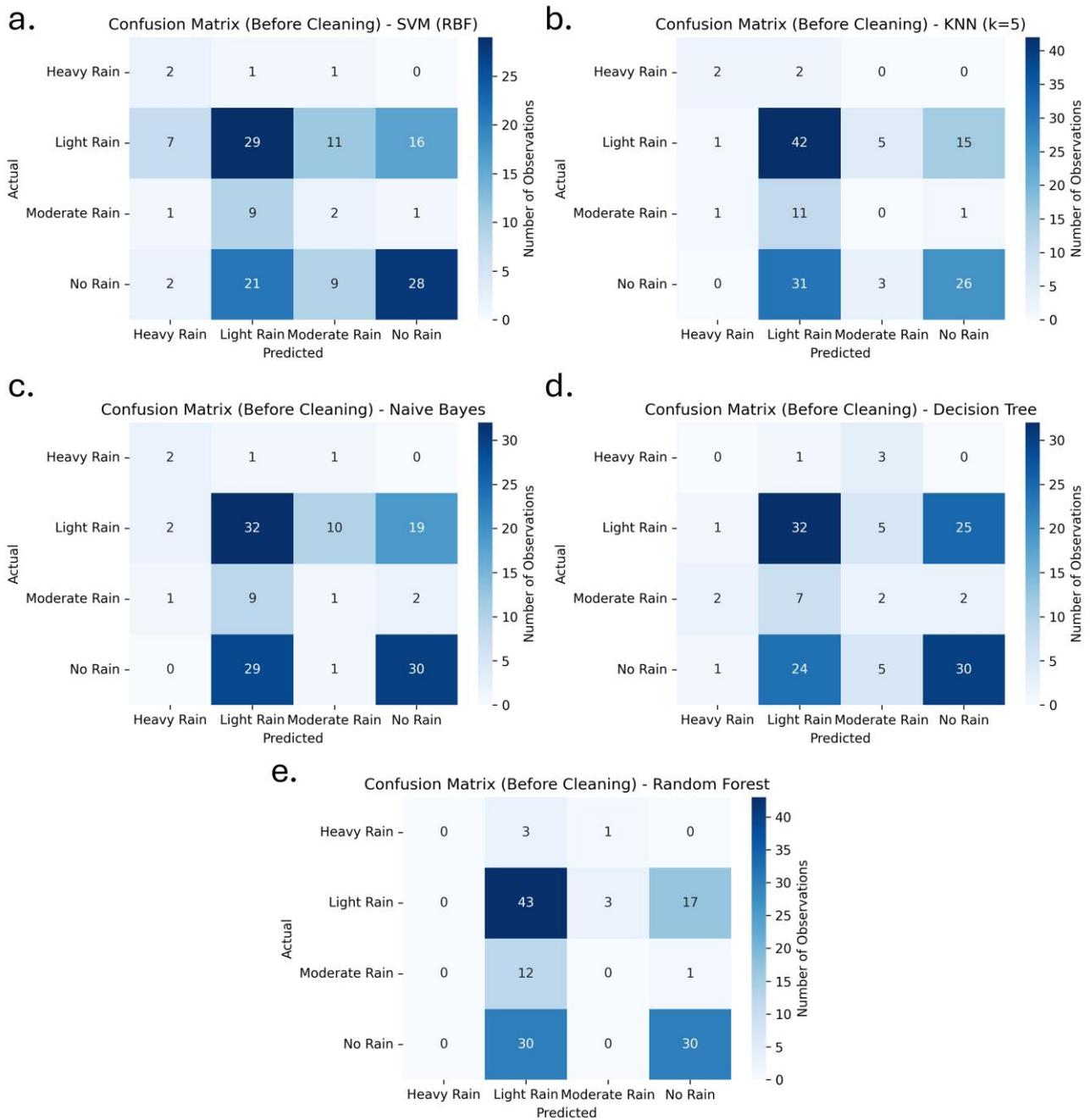


Figure 3. Confusion matrix before outlier removal.

A similar trend is observed in KNN, where after outlier removal predictions concentrate on No Rain and the detection of Moderate and Heavy Rain further deteriorates, reflecting KNN’s sensitivity to changes in data distribution. Naive Bayes also shifts strongly toward No Rain after outlier removal, suggesting that the removal of extreme rainfall values affects the representativeness of its probabilistic assumptions. In contrast, the Decision Tree retains relatively more balanced predictions, preserving some ability to recognize Light and Moderate Rain despite increased No Rain accuracy, indicating greater adaptability to distributional changes. Although Random Forest achieves higher accuracy after outlier removal, its predictions become heavily biased toward No Rain, resulting in a loss of minority-class sensitivity and reduced classification balance as reflected in macro F1-score. Overall, while outlier removal improves accuracy, it may reduce sensitivity to rainfall events—particularly Light Rain—highlighting the need for cautious treatment of extreme values and the joint use of accuracy and macro F1-score to achieve a more balanced and climatologically meaningful classification. Rainfall misclassification may also be influenced by seasonal variability, particularly during transitional monsoon periods when rainfall intensity fluctuates rapidly. Future studies should incorporate temporal features such as month or season to better capture climatic patterns.

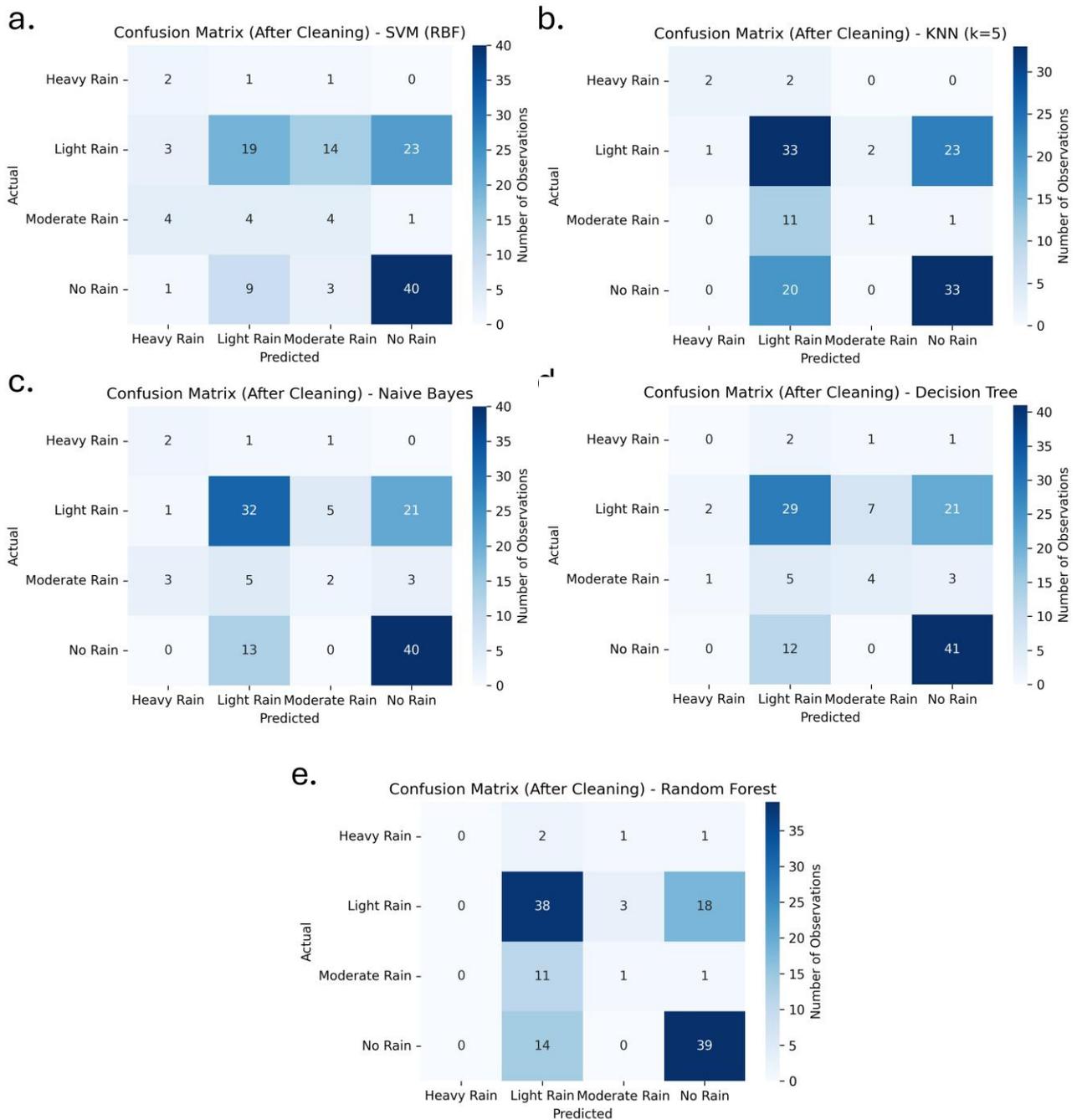


Figure 4. Confusion matrix after outlier removal.

3.4. Model Accuracy of Machine Learning Algorithms

Based on Figure 5, outlier removal consistently improves the accuracy of all classification algorithms, indicating that extreme values previously degraded model performance. SVM (RBF) shows a moderate accuracy increase, suggesting reduced noise supports better class separation with accuracy increased from 43.57% to 50.39%, while KNN (k = 5) rose from 50.00% to 53.49%, exhibits a clearer improvement, confirming its high sensitivity to data distribution and distance distortion caused by outliers. Naive Bayes demonstrates a substantial after outlier removal accuracy gain, indicating more stable and representative class probability estimation which accuracy is improved from 46.43% to 58.91%. The Decision Tree records one of the largest improvements accuracy from 45.71% to 57.36%, highlighting the strong influence of extreme values on split selection and tree stability, whereas Random Forest also benefits markedly as cleaner data enhances ensemble consistency and prediction accuracy which increased from 52.14% to 60.47%. Overall, these results emphasize that outlier removal is a crucial preprocessing step for daily rainfall classification; however, complementary metrics such as the macro F1-score are still required to ensure that accuracy gains reflect balanced performance across all rainfall classes rather than dominance of the majority class.

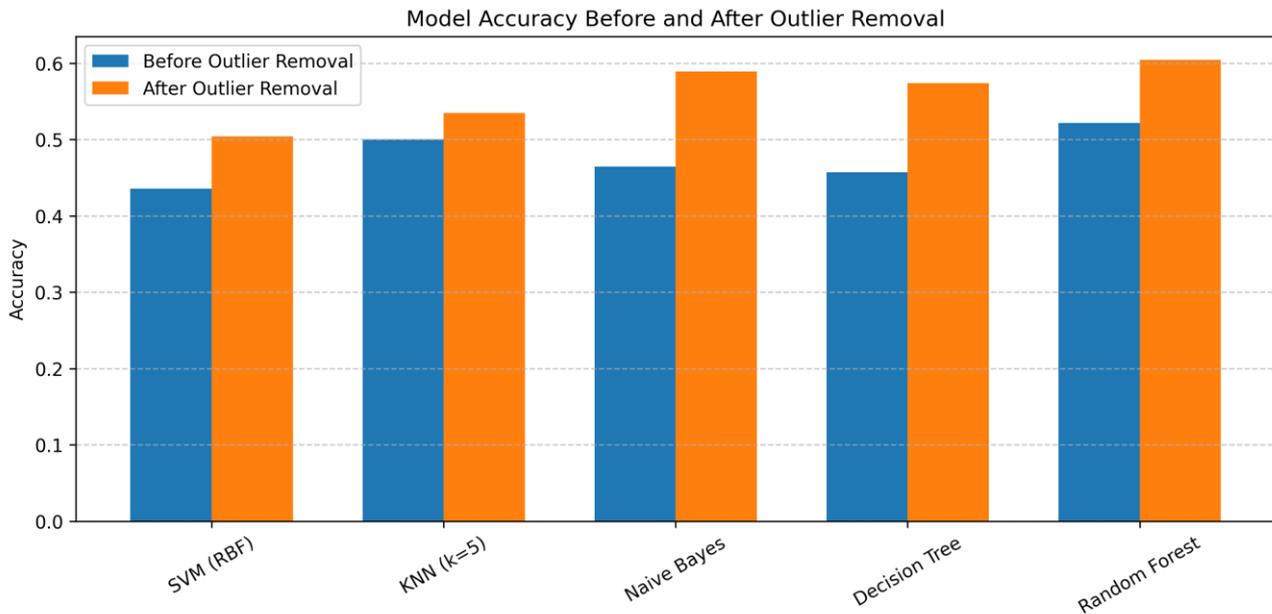


Figure 5. Comparison of model accuracy before and after outlier removal.

3.5. Relationship between Accuracy and Macro F1-Score after Outlier Removal

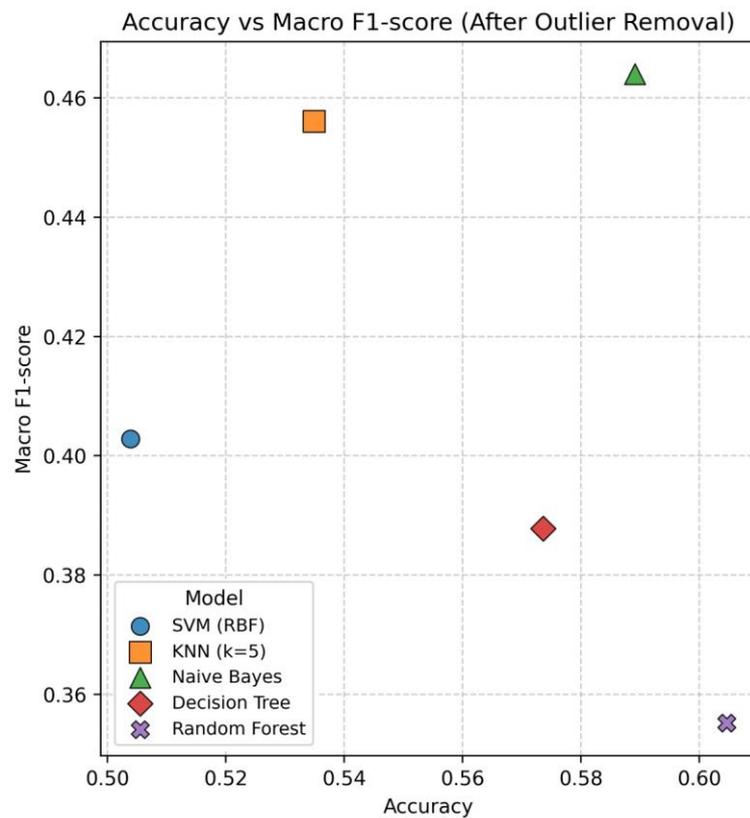


Figure 6. Comparison of accuracy and macro F1-score for each algorithm after outlier removal.

The relationship between accuracy and macro F1-score after outlier removal in Figure 6 shows that higher accuracy does not always correspond to a proportional improvement in macro F1-score, indicating that class-wise performance remains imbalanced for some models. Random Forest and Decision Tree achieve relatively high accuracy but lower macro F1-scores, suggesting a bias toward the majority rainfall class that reduces performance on minority classes, a behavior more clearly captured by the macro F1-score. In contrast, KNN ($k = 5$) attains a comparatively high macro F1-score at moderate accuracy, while Naive Bayes achieves the highest macro F1-score alongside high accuracy, reflecting more balanced classification across rainfall classes. Decision Tree also demonstrates a slightly better macro F1-score than Random Forest despite lower accuracy,

implying more balanced class partitioning. Conversely, SVM (RBF) records the lowest accuracy and macro F1-score, indicating continued difficulty in learning rainfall patterns, likely due to class imbalance and challenges in modeling the nonlinear characteristics of daily rainfall data. The macro F1-score of SVM (RBF) increased from 34.15% to 40.28%, KNN ($k = 5$) from 39.34% to 45.61%, Naive Bayes from 38.49% to 46.40%, Decision Tree from 28.99% to 38.78%, and Random Forest from 28.13% to 35.51%.

3.6. Statistical Test for Comparing Model Accuracy

The statistical tests were conducted using paired accuracy values obtained before and after outlier removal for each classification algorithm. A total of five paired observations ($n = 5$), corresponding to SVM (RBF), KNN ($k = 5$), Naive Bayes, Decision Tree, and Random Forest, were included in the analysis.

Table 1. Results of the normality test, paired t-test, and effect size analysis.

Statistical Test	Parameter	Value	Interpretation
Shapiro–Wilk	Statistic	0.9498	The distribution of accuracy differences is close to normal $p > 0.05$, indicating that the normality assumption is satisfied
	P-value	0.7354	
Paired T-Test	T-statistic	-5.2155	A significant difference in mean accuracy is observed $p < 0.05$, indicating a statistically significant difference
	P-value	0.0064	
Effect Size Cohen’s d	d	2.3324	Very large effect size (strong practical impact)

Based on Table 1, the Shapiro–Wilk test produces a p-value of 0.7354 ($p > 0.05$), indicating that the accuracy differences before and after outlier removal are normally distributed and that a paired t-test is appropriate. The paired t-test yields a t-statistic of -5.2155 with a p-value of 0.0064 ($p < 0.05$), confirming a statistically significant increase in model accuracy after the treatment, with the negative t-value indicating higher post-treatment accuracy. The effect size analysis further shows a Cohen’s d of 2.3324, classified as a very large effect, demonstrating that the improvement is not only statistically significant but also practically substantial. Overall, these results confirm that outlier removal leads to a valid, significant, and strongly impactful improvement in the accuracy of rainfall classification models.

4. Conclusions and Recommendations

The results show that outlier removal consistently improves the performance of all evaluated classification algorithms, as indicated by increases in both accuracy and F1-score. Accuracy improvements were observed for SVM (RBF) from 43.57% to 50.39% ($\Delta = 6.82\%$), KNN ($k = 5$) from 50.00% to 53.49% ($\Delta = 3.49\%$), Naive Bayes from 46.43% to 58.91% ($\Delta = 12.49\%$), Decision Tree from 45.71% to 57.36% ($\Delta = 11.65\%$), and Random Forest from 52.14% to 60.47% ($\Delta = 8.32\%$). Similar trends were found in macro F1-score, with SVM (RBF) increasing from 34.15% to 40.28%, KNN from 39.34% to 45.61%, Naive Bayes from 38.49% to 46.40%, Decision Tree from 28.99% to 38.78%, and Random Forest from 28.13% to 35.51%. The largest gains were achieved by Naive Bayes and Decision Tree, indicating high sensitivity to outliers in rainfall data, while Random Forest and KNN, despite higher final accuracy, exhibited moderate F1-scores, suggesting a tendency toward majority-class dominance and highlighting that improved accuracy does not always imply balanced classification across all rainfall classes.

This study is limited to an initial analysis that uses rainfall as the sole classification feature. Future work should incorporate additional meteorological variables—such as air temperature, humidity, atmospheric pressure, and wind speed—to better capture inter-variable relationships. To address the persistent class imbalance indicated by F1-score results, the application of the Synthetic Minority Over-sampling Technique (SMOTE) during preprocessing is strongly recommended, as it can improve minority class representation and promote more balanced and robust model performance. Overall, these enhancements are expected to improve both classification accuracy and interpretability while providing deeper insights into the contributions of meteorological variables in the study area.

References

- [1] Julikah, G. Rahmat, and M. B. Wiranatanegara, “Subsurface Geological Evaluation of the Central Sumatra Basin in Relation to the Presence of Heavy Oil,” *Sci. Contrib. Oil Gas*, vol. 44, no. 1, pp. 65–81, 2021, doi: 10.29017/scog.44.1.491.
- [2] E. Yanfatriani *et al.*, “Extreme Rainfall Trends and Hydrometeorological Disasters in Tropical Regions:

- Implications for Climate Resilience,” *Emerg. Sci. J.*, vol. 8, no. 5, pp. 1860–1874, 2024, doi: 10.28991/ESJ-2024-08-05-012.
- [3] J. Wang, Z. You, P. Song, and Z. Fang, “Rainfall’s impact on agricultural production and government poverty reduction efficiency in China,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–21, 2024, doi: 10.1038/s41598-024-59282-2.
- [4] R. Rahayu, S. A. Mathias, S. Reaney, G. Vesuviano, R. Suwarman, and A. M. Ramdhan, “Impact of land cover, rainfall and topography on flood risk in West Java,” *Nat. Hazards*, vol. 116, no. 2, pp. 1735–1758, 2023, doi: 10.1007/s11069-022-05737-6.
- [5] V. Kumar, N. Kedam, O. Kisi, S. Alsulamy, K. M. Khedher, and M. A. Salem, “A Comparative Study of Machine Learning Models for Daily and Weekly Rainfall Forecasting,” *Water Resour. Manag.*, vol. 39, no. 1, pp. 271–290, 2025, doi: 10.1007/s11269-024-03969-8.
- [6] C. W. Wu and F. N. F. Chou, “An inverse-problem approach to detect outliers in rainfall measurements of ground gauges for robust reservoir flood control operation,” *J. Hydrol.*, vol. 620, no. PA, p. 129360, 2023, doi: 10.1016/j.jhydrol.2023.129360.
- [7] P. K. Das, R. L. Sahu, and P. C. Swain, “Comparative analysis of machine learning models for rainfall prediction,” *J. Atmos. Solar-Terrestrial Phys.*, vol. 264, no. August, p. 106340, 2024, doi: 10.1016/j.jastp.2024.106340.
- [8] N. Nicholls, “Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation,” *Nat. Hazards*, vol. 23, no. 2–3, pp. 137–155, 2001, doi: 10.1023/A:1011130223164.
- [9] Z. A. Dwiyantri and C. Prianto, “Prediksi Cuaca Kota Jakarta Menggunakan Metode Random Forest,” *J. Tekno Insentif*, vol. 17, no. 2, pp. 127–137, 2023, doi: 10.36787/jti.v17i2.1136.
- [10] I. Hapsari and S. Pandya Wisesa, “Evaluasi Model Prediksi Curah Hujan Berbasis Machine Learning di Kota Bandung,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 11, no. 2, pp. 136–143, 2025, doi: 10.25077/teknosi.v11i2.2025.136-143.
- [11] O. Ejike, D. Ndzi, and M. Z. Shakir, “Comparative Study of Machine Learning-Based Rainfall Prediction in Tropical and Temperate Climates,” *Climate*, vol. 13, no. 8, pp. 1–27, 2025, doi: 10.3390/cli13080167.
- [12] C. D. Usman and U. Sudibyo, “Klasifikasi Curah Hujan di Kota Semarang Menggunakan Machine Learning,” *Pros. SAINTEK Sains dan Teknol.*, vol. 1, no. 1, pp. 1–5, 2022, [Online]. Available: https://www.researchgate.net/publication/388748074_Klasifikasi_Curah_Hujan_di_Kota_Semarang_Menggunakan_Machine_Learning/link/67a465ab207c0c20fa7b4e06/download?_tp=eyJjb250ZXh0Ijp7InBhZ2UiOiJwdWJsaWNhdGlvbiIsInByZXZpb3VzUGFnZSI6bnVsbH19
- [13] BMKG, *Penyediaan Dan Penyebarluasan Peringatan Dini Cuaca Ekstrem Sesaat (Nowcasting) 1 (Satu) Sampai Dengan 3 (Tiga) Jam Ke Depan Di Lingkungan Badan Meteorologi, Klimatologi, Dan Geofisika*. 2023, pp. 1–11. [Online]. Available: <https://jdih.bmkg.go.id/common/dokumen/2023sopbmkg025.pdf>
- [14] D. T. Utari, G. R. P. Palage, F. Fadhlirobby, and A. B. Nuswantoro, “COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR RAINFALL CLASSIFICATION IN YOGYAKARTA,” *BAREKENG J. Math. Its Applications*, vol. 264, no. 4, pp. 2765–2776, 2024, doi: 10.1016/j.jastp.2024.106340.
- [15] M. Yaşar and F. Dikbaş, “Multivariate Outlier Detection in Precipitation Series by Using Two-Dimensional Correlation,” no. January, pp. 0–18, 2023, doi: <https://doi.org/10.21203/rs.3.rs-399196/v2>.
- [16] S. Lusito, A. Pugnana, and R. Guidotti, *Solving imbalanced learning with outlier detection and features reduction*, vol. 113, no. 8. Springer US, 2024. doi: 10.1007/s10994-023-06448-0.
- [17] W. M. Ridwan, M. Sapitang, A. Aziz, K. Faizal, A. Najah, and A. El-shafie, “Rainfall forecasting model using machine learning methods : Case study,” *Ain Shams Eng. J.*, 2020, doi: 10.1016/j.asej.2020.09.011.
- [18] V. Guhan, A. D. Raju, R. Krishna, and K. Nagaratna, “Evaluating weather trends and forecasting with machine learning : Insights from maximum temperature , minimum temperature , and rainfall data in India,” *Dyn. Atmos. Ocean.*, vol. 110, no. March, p. 101562, 2025, doi: 10.1016/j.dynatmoce.2025.101562.
- [19] F. Taromideh, G. F. Santonastaso, and R. Greco, “Journal of Hydrology : Regional Studies Rainfall nowcasting by integrating radar and rain gauge data with machine learning for Ischia Island , Italy,” *J. Hydrol. Reg. Stud.*, vol. 58, no. March, p. 102273, 2025, doi: 10.1016/j.ejrh.2025.102273.
- [20] S. Jang, J. Yoo, Y. Lee, and B. Kim, “Progress in Disaster Science Flood prediction in urban areas based on machine learning considering the statistical characteristics of rainfall,” *Prog. Disaster Sci.*, vol. 26, no. March, p. 100415, 2025, doi: 10.1016/j.pdisas.2025.100415.
- [21] H. Wen, F. Yan, J. Huang, and Y. Li, “Interpretable machine learning models and decision-making mechanisms for landslide hazard assessment under different rainfall conditions,” *Expert Syst. Appl.*, vol. 270, no. January, p. 126582, 2025, doi: 10.1016/j.eswa.2025.126582.

- [22] F. I. Arassah, K. Sadik, B. Sartono, and P. Sofan, "Optimizing Machine Learning for Daily Rainfall Prediction in Bogor : A Statistical Downscaling Approach," vol. 5, no. 6, pp. 7006–7018, 2025, doi: <https://doi.org/10.59188/eduvest.v5i6.51307>.
- [23] A. Rahmannisa, M. Ariska, S. M. Siahaan, and I. Seprina, "Implementation of Machine Learning for Rainfall Prediction in Smoke-Prone Areas of South Sumatra," vol. IX, no. 2, pp. 89–103, 2025, doi: <https://doi.org/10.19109/h8s3w172>.
- [24] BMKG, "Data Online - Direktorat Data dan Komputasi BMKG." <https://dataonline.bmkg.go.id/>
- [25] C. S. Kim, C. R. Kim, and K. H. Kok, "Outlier Detection in Hydrological Data Using Machine Learning: A Case Study in Lao PDR," *Water (Switzerland)*, vol. 17, no. 21, pp. 1–28, 2025, doi: 10.3390/w17213120.
- [26] M. Okirya and J. Du Plessis, "Evaluating Bias Correction Methods Using Annual Maximum Series Rainfall Data from Observed and Remotely Sensed Sources in Gauged and Ungauged Catchments in Uganda," *Hydrology*, vol. 12, p. 113, 2025, doi: <https://doi.org/10.3390/hydrology12050113>.
- [27] C. Meng, Y. Hu, Y. Zhang, and F. Guo, "PSBP-SVM: A Machine Learning-Based Computational Identifier for Predicting Polystyrene Binding Peptides," vol. 8, no. March, pp. 1–9, 2020, doi: 10.3389/fbioe.2020.00245.
- [28] M. A. Azeem and S. Dev, "A performance and interpretability assessment of machine learning models for rainfall prediction in the Republic of Ireland," *Decis. Anal. J.*, vol. 12, no. June, p. 100515, 2024, doi: 10.1016/j.dajour.2024.100515.
- [29] M. Hassan *et al.*, "Machine Learning-Based Rainfall Prediction : Unveiling Insights and Forecasting for Improved Preparedness," *IEEE Access*, vol. 11, no. November, pp. 132196–132222, 2023, doi: 10.1109/ACCESS.2023.3333876.
- [30] V. Kumar, V. K. Yadav, and E. S. Dubey, "Rainfall Prediction using Machine Learning," no. May, 2022, doi: <https://doi.org/10.22214/ijraset.2022.42876>.
- [31] Q. B. Pham, T. C. Yang, C. M. Kuo, H. W. Tseng, and P. S. Yu, "Combing random forest and least square support vector regression for improving extreme rainfall downscaling," *Water (Switzerland)*, vol. 11, no. 3, 2019, doi: 10.3390/w11030451.
- [32] M. C. H. Lee, J. Braet, and J. Springael, "applied sciences Performance Metrics for Multilabel Emotion Classification : Comparing Micro , Macro , and Weighted F1-Scores," *Appl. Sci.*, vol. 14, no. 9863, 2024, doi: <https://doi.org/10.3390/app14219863>.
- [33] T. K. Aulia, W. A. Arifin, and M. Rudi, "Klasifikasi Kualitas Air Budidaya Ikan Nila Menggunakan Support Vector Machine," pp. 1842–1852, 2025, doi: 10.33364/algorithm/v.22-2.2356.
- [34] S. S. Shapiro, M. B. Wilk, and B. T. Laboratories, "An analysis of variance test for normality (complete samples)!," 1965, doi: <http://www.math.utah.edu/~morris/Courses/ShapiroWilk.pdf>.
- [35] Abraham S. Fischler School of Education, "Module 6 : t Tests Module 6 Overview," Nova Southeastern University. [Online]. Available: <https://app.nova.edu/toolbox/instructionalproducts/Statistics & SPSS/Module 6 t tests.pdf>
- [36] O. Rainio, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.