https://doi.org/10.32815/jitika.1252

Volume 20, Issue 2, 2026

Jurnal Ilmiah Teknologi Informasi Asia

Journal Homepage: https://jurnal.asia.ac.id/index.php/jitika

Article

Comparative robustness and interpretability analysis of MLP and random forest for multi-class weather classification in tropical regions

Yoyok Dewantoro [a,c,*], Joko Triloka [a], Ahmad Zarkoni [b,c], Tomy Ivan Sugiharto [b]

[a] *Department of Informatics Engineering, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, 35141, Indonesia*
[b] *Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, 65144, Indonesia*
[c] *Meteorological, Climatological, and Geophysical Agency (BMKG), Malang, 65165, Indonesia*

***Abstract*—The stability of machine learning models remains a critical challenge in weather type classification, particularly when applied to low-variance meteorological datasets characterized by correlated atmospheric parameters. Neural network models, such as the Multi-Layer Perceptron (MLP), are known to capture nonlinear relationships effectively but may be sensitive to training–testing data partitioning. In contrast, ensemble methods like Random Forests (RFs) are designed to reduce variance through aggregation. This study systematically evaluates the stability and generalization capability of MLP and RF classifiers for multi-class weather classification (sunny, cloudy, rainy) using historical meteorological data. Experiments were conducted under three data split scenarios (80:20, 70:30, 60:40) and validated using 5-fold and 10-fold cross-validation. While MLP achieved accuracy above 96% across all scenarios, RF consistently outperformed MLP with accuracy between 98% and 99%. Importantly, cross-validation results reveal that RF demonstrates superior stability, with standard deviation values ranging from 0.00 to 0.01, compared to 0.01 for MLP. These findings confirm that ensemble-based methods provide more robust and consistent performance for meteorological classification tasks characterized by multivariate dependence and limited variance variability.**

***Keywords*—machine learning; meteorological data; multi-layer perceptron; random forest; weather classification.**

1. Introduction

Weather conditions constitute dynamic and complex atmospheric phenomena that exert significant influence on various strategic sectors, including agriculture, transportation, tourism, disaster mitigation, and regional planning (Shimada, 2024). In tropical regions, weather variability is largely governed by nonlinear interactions among multiple meteorological parameters such as air temperature, relative humidity, atmospheric pressure, wind speed, precipitation, and cloud cover (Simbo et al., 2023). These complex, interdependent relationships pose substantial challenges for accurate weather classification, particularly when relying on conventional analytical methods.

Traditional weather analysis methods are commonly based on classical statistical techniques and deterministic rule-based thresholds derived from predefined meteorological criteria (Yang et al., 2024). While such approaches offer interpretability and simplicity, their effectiveness is often limited when dealing with nonlinear patterns and multivariate dependencies inherent in atmospheric systems (Bhowmick et al., 2023; Mawalagedara et al., 2025). The increasing availability of long-term historical meteorological datasets, along with advances in computational capabilities, has therefore driven the adoption of machine learning techniques as a more flexible and data-driven alternative for weather classification and prediction (H. Zhang et al., 2025).

Recent studies have explored a wide range of machine

* Corresponding author.
E-mail Address: dewantoro.2321210014p@mail.darmajaya.ac.id (Y. Dewantoro)

Author E-mail(s): YD (dewantoro.2321210014p@mail.darmajaya.ac.id), JT (joko.triloka@darmajaya.ac.id), AZ (ahmad.zarkoni@bmkg.go.id), TIS (4240605210004@student.uin-malang.ac.id)
Digital Object Identifier 10.32815/jitika.1252
Manuscript submitted 15 January 2026; revised 16 February 2026; accepted 5 March 2026.
ISSN: 2580-8397(O), 0852-730X(P).

learning algorithms for weather-related applications, including Support Vector Machines, Decision Trees, K-Nearest Neighbor, Artificial Neural Networks, ensemble learning methods, and gradient boosting models (Ejike et al., 2025; Safia et al., 2023). Although more recent "black-box" models such as XGBoost and deep learning architectures have demonstrated strong predictive performance, their high computational cost, increased model complexity, and reduced interpretability may limit their applicability in operational forecasting environments where robustness, reproducibility, and computational efficiency are critical considerations (Yang et al., 2024).

Within this context, Multi-Layer Perceptron (MLP) and Random Forest (RF) remain relevant and widely adopted machine learning models for meteorological applications. MLP is capable of capturing complex nonlinear relationships through its layered neural architecture and nonlinear activation functions, making it well-suited for modeling multivariate atmospheric interactions (Cahyani et al., 2025; Wiguna et al., 2024). Meanwhile, Random Forest, as an ensemble-based decision tree method, offers strong robustness against overfitting, stability under noisy and correlated features, and relatively low sensitivity to hyperparameter tuning (Feng et al., 2024; Sun et al., 2024). Importantly, both models provide a practical balance between classification accuracy and computational efficiency, which is particularly advantageous for operational weather classification systems that require consistent performance across varying data conditions.

Despite the extensive use of MLPs and Random Forests in weather-related studies, several challenges remain inadequately addressed in the recent literature. In particular, systematic comparative analyses that explicitly evaluate model robustness under multiple training–testing data split scenarios and cross-validation schemes are still relatively limited, especially for multi-class weather classification problems (Aftab et al., 2024; Senior-Williams et al., 2024). Many existing studies report model performance using a single data partition, which may not adequately reflect model stability and generalization capability when applied to real-world meteorological data characterized by temporal variability and uncertainty.

Moreover, recent works published between 2024 and 2025 have emphasized the importance of model reliability, robustness, and operational feasibility in applied meteorology, rather than focusing solely on peak predictive accuracy (Shimada, 2024; Yang et al., 2024; C. J. Zhang et al., 2025). These studies highlight the need for classification frameworks that not only achieve high accuracy but also demonstrate consistent performance across different data configurations and validation strategies. However, empirical studies that comprehensively address these aspects using well-established yet operationally viable models, such as MLPs and Random Forests, remain relatively scarce.

Based on these considerations, this study aims to classify weather types into three categories: sunny, cloudy, and rainy, using historical meteorological parameters within a supervised machine learning framework. The study implements Multi-Layer Perceptron and Random Forest classifiers. It evaluates their performance using multiple metrics, including accuracy, precision, recall, F1-score, error-based metrics, receiver operating characteristic (ROC) analysis, and cross-validation. The main contributions of this study are threefold: (1) providing a systematic and up-to-date comparative analysis of MLP and Random Forest for multi-class weather classification, (2) evaluating model robustness across multiple training–testing data split scenarios, and (3) assessing model generalization capability through 5-fold and 10-fold cross-validation. The findings are expected to contribute to the development of reliable and operationally feasible machine learning-based weather classification systems and to serve as a reference for future research in applied meteorology and intelligent environmental modeling.

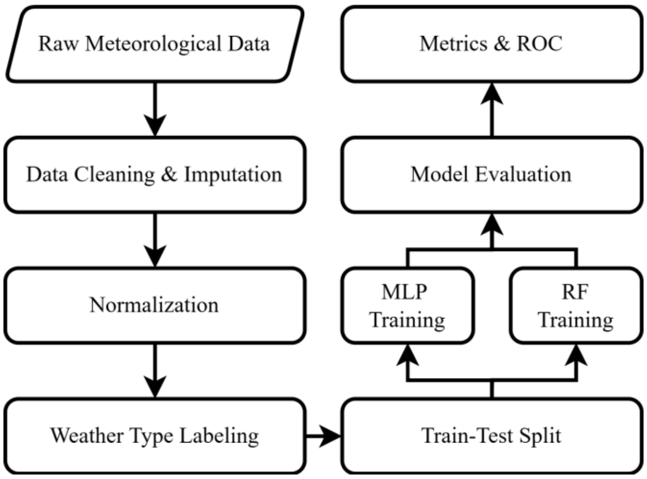


Fig. 1. Workflow of the Proposed Method.

2. Method

2.1. Research Framework and Workflow

This study adopts a supervised machine learning framework to classify weather types based on historical meteorological parameters (Safia et al., 2023). The methodological framework is designed systematically and in a structured manner to ensure reproducibility, transparent model configuration, and consistent evaluation across the implemented algorithms. The overall procedure consists of data acquisition, data preprocessing, weather type labeling, data partitioning, model training using Multi-Layer Perceptron (MLP) and Random Forest (RF), and comprehensive model performance evaluation.

To enhance technical clarity and methodological transparency, the entire data processing and modeling pipeline is depicted in a workflow diagram (Fig. 1). This visualization replaces purely descriptive procedural explanations. It provides an explicit computational representation of the proposed methodology.

2.2. Data Description and Meteorological Variables

The dataset used in this study consists of historical meteorological records collected over a nine-year period, from 2015 to 2024. Each observation represents atmospheric conditions at a specific time and includes several meteorological parameters commonly used in weather analysis. These parameters are treated as predictor variables (*input features*), while the weather type is defined as the target class (*output variable*).

The meteorological variables used in this study, along with their descriptions and units of measurement, are summarized in

Table 1. The use of a long temporal range is intended to capture interannual and seasonal variability, thereby improving the generalization capability of the classification models.

2.3. Weather Type Labeling Strategy

Weather type labeling represents a critical stage in the supervised learning process, as it directly influences the quality of model training. In this study, weather conditions are categorized into three classes: sunny, cloudy, and rainy. The labeling process is conducted using predefined and physically interpretable meteorological criteria derived from domain knowledge, primarily considering precipitation, relative humidity, and cloud cover (Tsagalidis & Evangelidis, 2022).

This rule-based labeling strategy ensures that each weather class corresponds to distinct atmospheric characteristics while maintaining interpretability and consistency. The distribution of samples across the three classes is further analyzed to identify potential class imbalance, which could affect model learning and evaluation.

2.4. Data Preprocessing

Data preprocessing is performed to enhance data quality and ensure compatibility with machine learning algorithms. Missing values arising from measurement limitations are handled using appropriate statistical imputation techniques, such as mean or median substitution, depending on the distributional characteristics of each variable (Novogroder Idan, 2024).

Subsequently, feature normalization is applied to address scale differences among meteorological variables, which is particularly important for neural network-based models. In this study, min–max normalization is employed, as defined in Equation (1):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x denotes the original feature value, $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values of the feature, respectively, and x' is the normalized value.

The preprocessed dataset is subsequently divided into training and testing subsets using multiple data partitioning scenarios to evaluate model robustness, as described in Table 2.

2.5. Data Partitioning Scheme

To assess the stability of the classification models across different training data proportions, the dataset is split into training and test sets using three data split scenarios: 80:20, 70:30, and 60:40. The detailed data partitioning scheme is presented in Table 2.

These scenarios are employed consistently for both MLP and Random Forest models to ensure a fair and objective performance comparison.

2.6. Multi-Layer Perceptron (MLP) Classifier

The primary classification model implemented in this study is the Multi-Layer Perceptron (MLP), which is a feedforward artificial neural network capable of modeling complex nonlinear relationships among meteorological variables (Cahyani et al., 2025). The MLP architecture consists of an input layer corresponding to the meteorological features, two hidden layers, and an output layer representing the weather type classes. The net input to a neuron in the hidden layer is computed using Equation (2):

$$z_j = \sum_{i=1}^{n} w_{ij} x_i + b_j \quad (2)$$

The nonlinear activation function applied to each hidden neuron is the Rectified Linear Unit (ReLU), defined in Equation (3):

$$f(z_j) = max(0, z_j) \quad (3)$$

Model training is performed using the backpropagation algorithm with the Adam optimizer to minimize the categorical

Table 1. Meteorological variables used in this study

Variable	Description	Unit
Temperature	Near-surface air temperature	°C
Humidity	Relative humidity	%
Wind speed	Average wind speed	m/s
Precipitation	Daily precipitation	mm
Cloud cover	Cloud coverage level	oktas
Atmospheric pressure	Mean sea-level pressure	hPa
UV index	Ultraviolet radiation index	–
Visibility	Horizontal visibility	km
Season	Seasonal indicator	categorical

Table 2. Data Separation Scheme

Scenario	Training Data	Testing Data
Split 1	80%	20%
Split 2	70%	30%
Split 3	60%	40%

Table 3. Hyperparameter configuration of MLP and Random Forest models

Model	Hyperparameter	Value
MLP	Number of hidden layers	2
	Neurons per layer	64-32
	Activation function	ReLU
	Output activation	Softmax
	Optimizer	Adam
	Learning rate	0.001
	Random state	42
	Max iter	1000
Random Forest	Number of trees ($n_{estimators}$)	100
	Criterion	Gini
	Max depth	5
	Min samples split	5
	Min samples leaf	10
	Random state	42

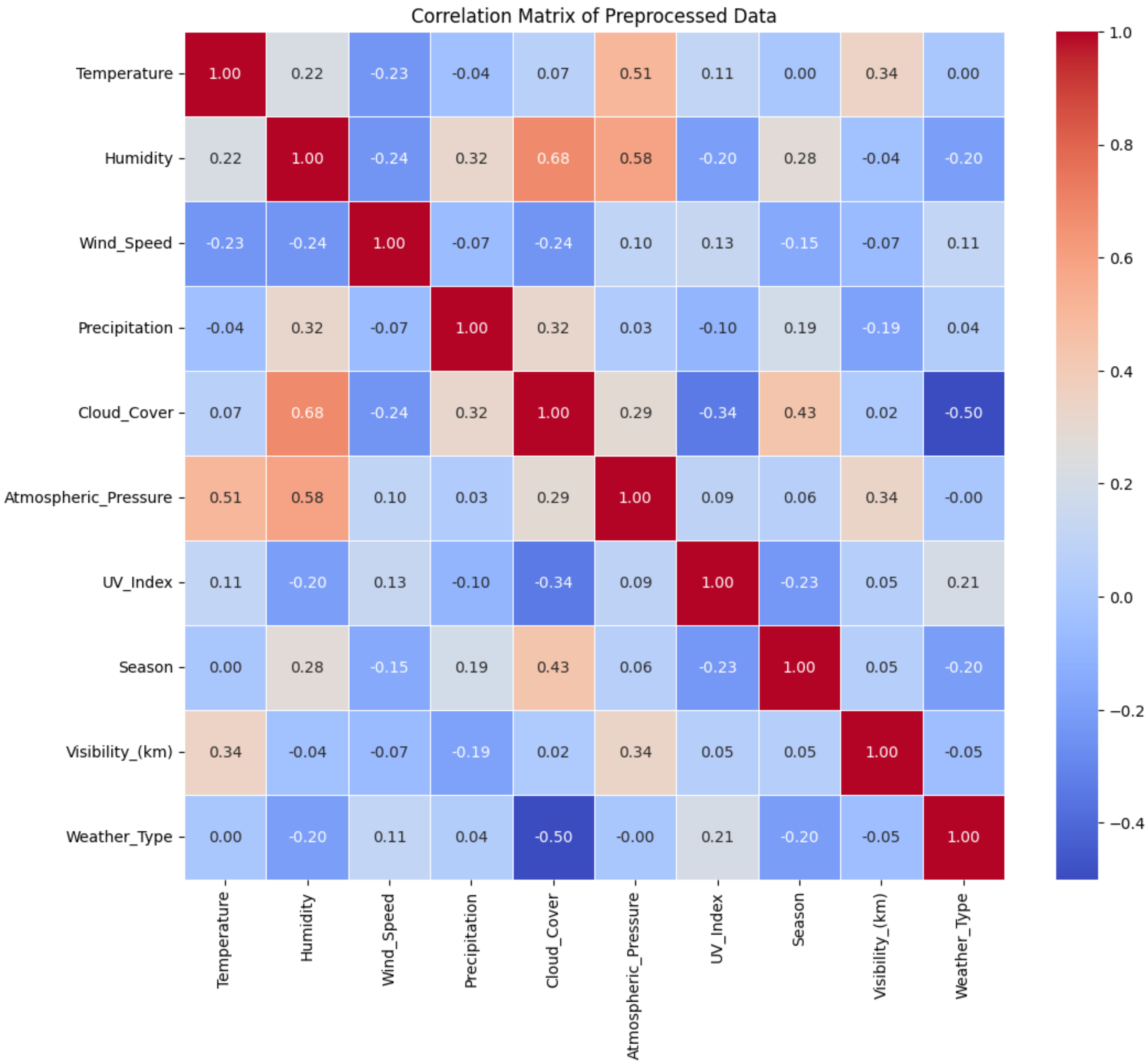


Fig. 2. Correlation Matrix.

cross-entropy loss function. The complete set of MLP hyperparameters, determined through empirical experimentation to balance predictive performance and computational efficiency, is summarized in Table 3.

2.7. Random Forest Classifier

As a benchmark model, this study employs the Random Forest (RF) classifier, an ensemble learning method based on multiple decision trees. Random Forest constructs a collection of decision trees using randomly selected subsets of training data and features, and the final prediction is obtained through a majority voting mechanism (Feng et al., 2024).

This ensemble strategy improves model stability and reduces overfitting, particularly for meteorological datasets characterized by correlated features and inherent variability. The Random Forest model is trained using the same data partitioning schemes as the MLP to ensure consistency in performance evaluation. The detailed configuration of Random Forest hyperparameters is also provided in Table 3.

This table is provided to ensure full experimental reproducibility and transparency.

2.8. Model Evaluation Metrics

Model performance is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Accuracy is defined in Equation (4):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

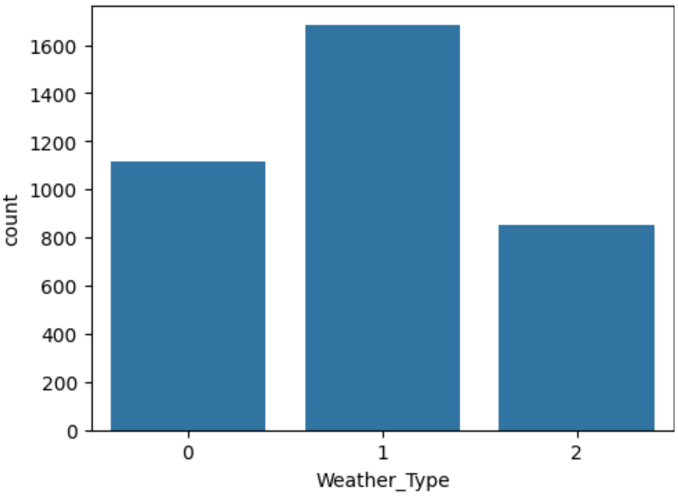


Fig. 3. Original Target (y) Distribution.

Table 4. Overall Classification Performance

Model	MLP	Random Forest
Best Data Split	80:20	70:30
Accuracy (%)	98.22	99.09
Precision	High	Very High
Recall	High	Very High
F1-score	High	Very High

Table 5. Error Metrics Comparison

Model	Data Split	MAE	RMSE
MLP	**80:20**	**0.02**	**0.13**
	70:30	0.03	0.17
	60:40	0.03	0.20
Random Forest	80:20	0.01	0.10
	70:30	**0.01**	**0.10**
	60:40	0.01	0.11

Precision, recall, and F1-score are defined in Equations (5)–(7), respectively:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{F1} - \text{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

In addition, the *confusion matrix*, *Receiver Operating Characteristic* (ROC) curves (Abdulla et al., 2022), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) are used to analyze classification performance and prediction errors (Aksan et al., 2025). To further evaluate model stability and generalization capability, 5-fold and 10-fold cross-validation schemes are applied.

3. Result

3.1. Data Characteristics and Target Distribution

The relationships among meteorological variables were initially examined using a correlation matrix (Fig. 2). The matrix shows varying degrees of linear correlation among atmospheric parameters, reflecting the inherent interdependence of meteorological processes. Notably, humidity and cloud cover exhibit a strong positive correlation, indicating their close physical association in atmospheric moisture dynamics.

The distribution of the target classes 0 (sunny), 1 (cloudy), and 2 (rainy) is shown in Fig. 3. The dataset shows a relatively balanced class distribution, with no single weather type dominating the observations. This balanced distribution supports unbiased model training and enables performance metrics to be interpreted without the need for additional class-balancing techniques.

3.2. Overall Classification Performance

The classification performance of the Multi-Layer Perceptron (MLP) and Random Forest (RF) models was evaluated across multiple data-split scenarios. For clarity and conciseness, only the best-performing configuration for each model is reported, and stability across splits is assessed via cross-validation.

Table 4 summarizes the overall classification performance of both models. The MLP classifier achieved its highest accuracy of 98.22% under the 80:20 data split scenario. In comparison, the Random Forest classifier achieved a superior accuracy of 99.09% with a 70:30 data split, with consistently high precision, recall, and F1-score across all weather classes.

3.3. Error Metrics Comparison

To further quantify predictive performance, error-based metrics were evaluated. Table 5 presents the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for both models.

The Random Forest model yields lower error values, indicating more accurate and stable predictions than the MLP classifier.

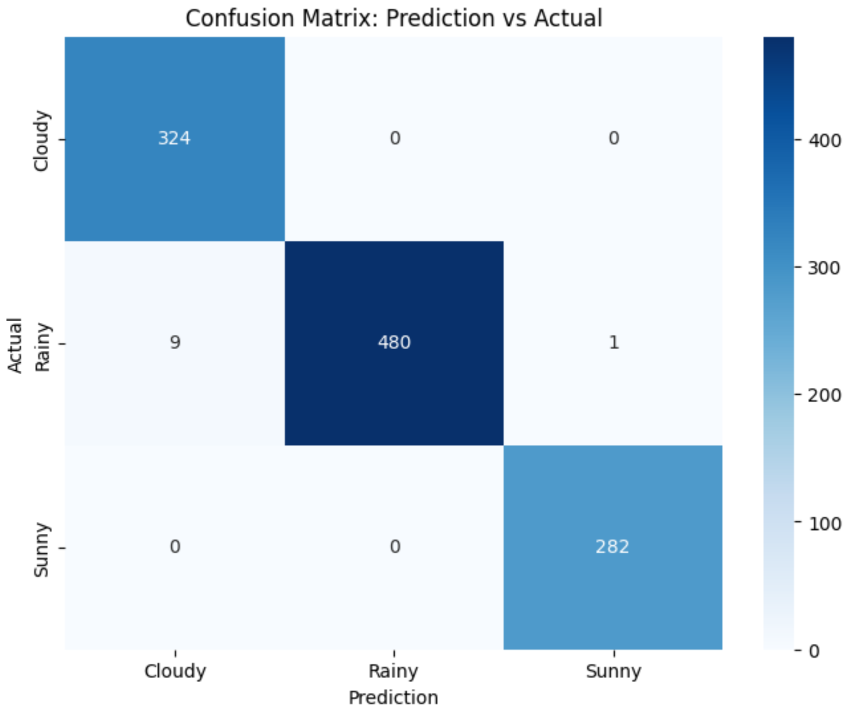
3.4. Confusion Matrix and ROC Analysis

The confusion matrix and Receiver Operating Characteristic (ROC) curve for the Random Forest model under its best-performing configuration are shown in Fig. 4. The confusion matrix indicates minimal misclassification across all weather categories, while the ROC curve demonstrates strong discrimination, with curves approaching the upper-left corner of the plot.

These results confirm that the Random Forest classifier effectively distinguishes between sunny, cloudy, and rainy conditions with high confidence.

3.5. Feature Importance Analysis of the Random Forest Model

To enhance model interpretability, feature importance analysis was conducted using the best Random Forest classifier.



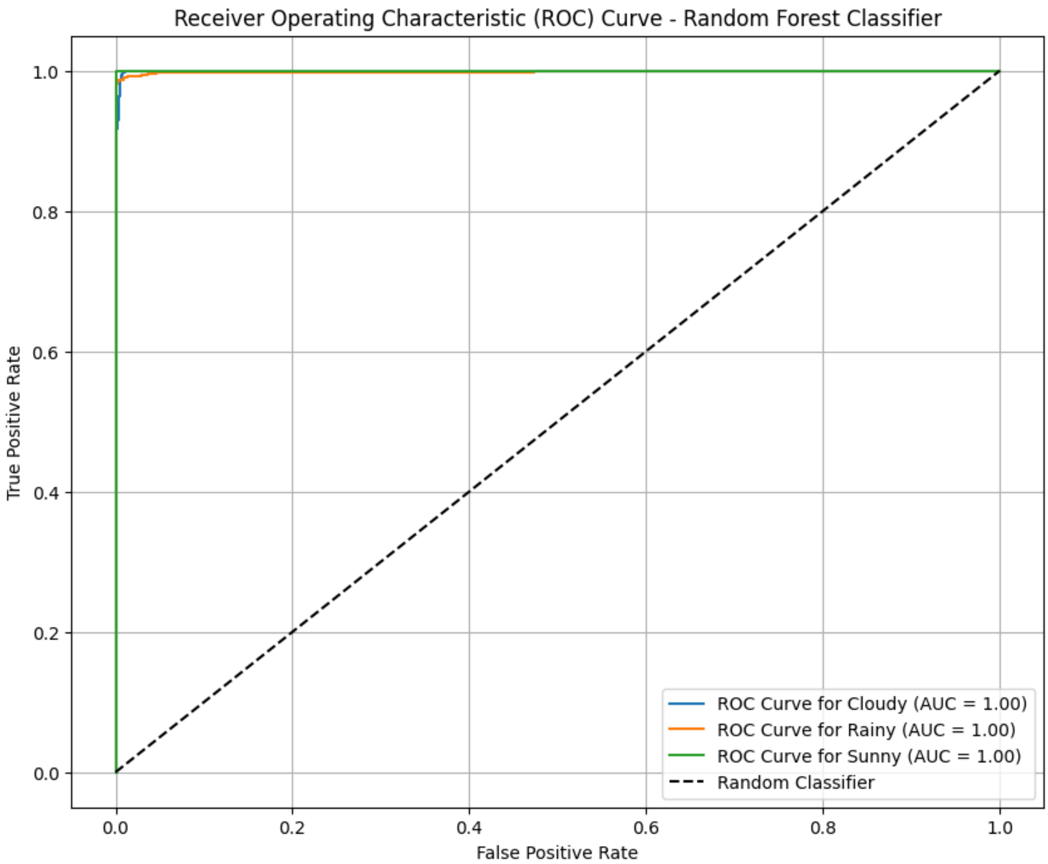


Fig. 4. Confusion matrix and ROC curve for the Random Forest best model.

The relative importance of each meteorological variable is illustrated in Fig. 5.

The results indicate that cloud cover, relative humidity, and precipitation are the most influential features driving the classification process. These variables contribute substantially more to model decisions compared to temperature, wind speed, atmospheric pressure, UV index, and visibility.

The dominance of cloud cover and humidity reflects fundamental atmospheric processes governing weather conditions. High humidity is closely associated with cloud

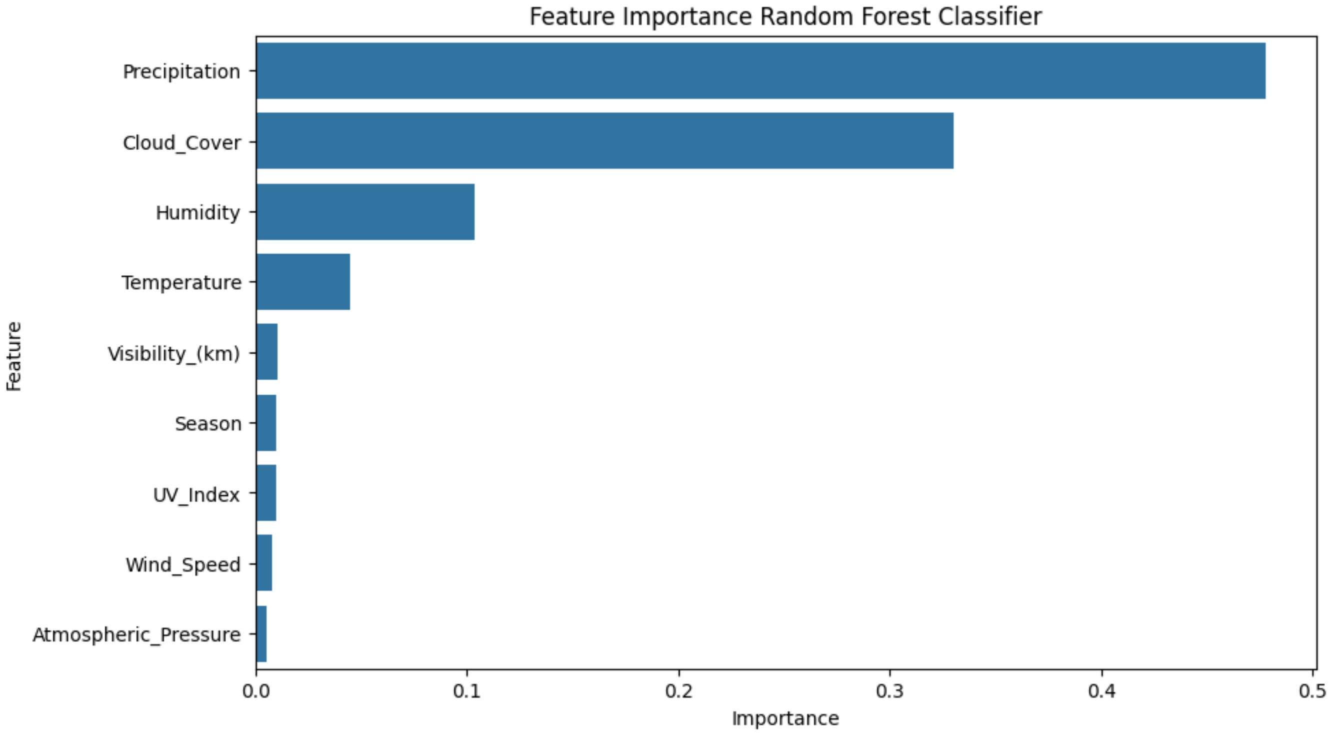


Fig. 5. Feature importance analysis.

Table 6. Schemes 5-Fold and 10-Fold Cross-Validation

Model	Data Split	5-Fold		10-Fold	
		Average accuracy	Standard deviation	Average accuracy	Standard deviation
Multi-Layer Perceptron	80:20	96.58%	0.01	97.06%	0.01
	70:30	96.21%	0.01	96.01%	0.01
	60:40	96.07%	0.01	96.03%	0.01
Random Forest	80:20	98.36%	0.00	98.32%	0.00
	70:30	98.24%	0.00	98.20%	0.00
	60:40	98.22%	0.00	98.27%	0.01

formation, while increased cloud cover is a direct indicator of atmospheric instability. Precipitation further reinforces class separation by clearly distinguishing rainy conditions from non-rainy weather types.

Conversely, variables such as wind speed and atmospheric pressure exhibit lower importance scores, suggesting a more indirect role in short-term weather classification. Overall, the feature importance results demonstrate that the Random Forest model not only achieves high predictive accuracy but also aligns well with established meteorological principles, thereby enhancing the physical interpretability of the classification outcomes.

3.6. Cross-Validation Results

To further assess model stability, cross-validation experiments were conducted using 5-fold and 10-fold schemes. The average accuracy and standard deviation for each model and data split scenario are summarized in Table 6.

For the MLP classifier, the average cross-validation accuracy ranged from 96.07% to 97.06%, with a standard deviation of approximately 0.01 across all configurations. In contrast, the Random Forest classifier achieved higher average accuracy values, ranging from 98.20% to 98.36%, with very small standard deviations.

These cross-validation results provide an objective measure of model performance consistency across multiple training and testing partitions.

3.7. Explainable Artificial Intelligence (XAI) Analysis Using SHAP

To enhance the interpretability of the Random Forest classifier and address the need for explainable machine learning in meteorological applications, this study integrates Shapley Additive Explanations (SHAP) analysis. SHAP is a model-agnostic interpretability framework derived from cooperative game theory that quantifies the contribution of each feature to individual predictions and overall model output.

While traditional feature importance scores in Random Forest measure average impurity reduction, SHAP provides a more rigorous attribution mechanism by estimating the marginal contribution of each meteorological variable across all possible feature combinations. This approach enables both global and local interpretability of the classification model.

The SHAP summary plot for the best-performing Random Forest model (70:30 split) is presented in Fig. 6. The figure illustrates the distribution of SHAP values for each meteorological variable across all samples.

The SHAP analysis confirms that cloud cover, relative humidity, and precipitation are the three most influential variables driving weather type classification. These variables exhibit the largest absolute SHAP values, indicating strong contributions to decision boundaries within the ensemble trees.

High cloud cover values produce positive SHAP contributions toward the cloudy and rainy classes, whereas low cloud cover shifts predictions toward the sunny class. Similarly, elevated relative humidity increases the likelihood of rainfall, while lower humidity favors sunny conditions.

Precipitation emerges as a decisive separator between rainy and non-rainy classes. Even moderate increases in precipitation generate substantial positive SHAP values for the rainy category, reinforcing the model's reliance on physically meaningful atmospheric indicators.

3.8. Meteorological Interpretation of SHAP Patterns

From a meteorological perspective, the SHAP findings align with established atmospheric dynamics. The strong contribution of relative humidity and cloud cover reflects the thermodynamic relationship between moisture content and cloud formation. High humidity enhances condensation, leading to cloud formation and potential precipitation. Consequently, the model correctly assigns positive contributions to these variables when predicting cloudy or rainy conditions.

The interaction between humidity and cloud cover, previously observed in the correlation matrix (Fig. 1), is evident in the SHAP distributions. Their combined high SHAP values indicate that the Random Forest model internally constructs decision rules that mirror real-world atmospheric processes, such as high moisture plus dense cloud coverage, increasing the probability of rainfall.

Interestingly, variables such as wind speed and atmospheric pressure exhibit smaller SHAP magnitudes. This suggests that while these parameters contribute to atmospheric variability, their direct influence on short-term categorical weather classification is comparatively weaker within the studied dataset. Pressure variations, for instance, may be more relevant for synoptic-scale analysis than for immediate weather-type labeling.

The integration of SHAP significantly strengthens the scientific transparency of the proposed classification framework. Rather than functioning as a purely "black-box" predictor, the Random Forest model demonstrates explainable decision logic that aligns with meteorological principles.

For operational meteorological agencies, such as automated weather station (AWS) systems, interpretability is critical for trust and deployment. The SHAP analysis confirms that the model's predictions are not arbitrary but are grounded in physically interpretable atmospheric relationships.

Overall, incorporating SHAP analysis enhances this study's methodological contribution by combining high predictive performance with explainable artificial intelligence. This integration addresses common concerns regarding machine learning interpretability in environmental modeling and provides a more robust foundation for practical implementation.

4. Discussion

The results of this study indicate that both the Multi-Layer Perceptron (MLP) and Random Forest (RF) classifiers achieve

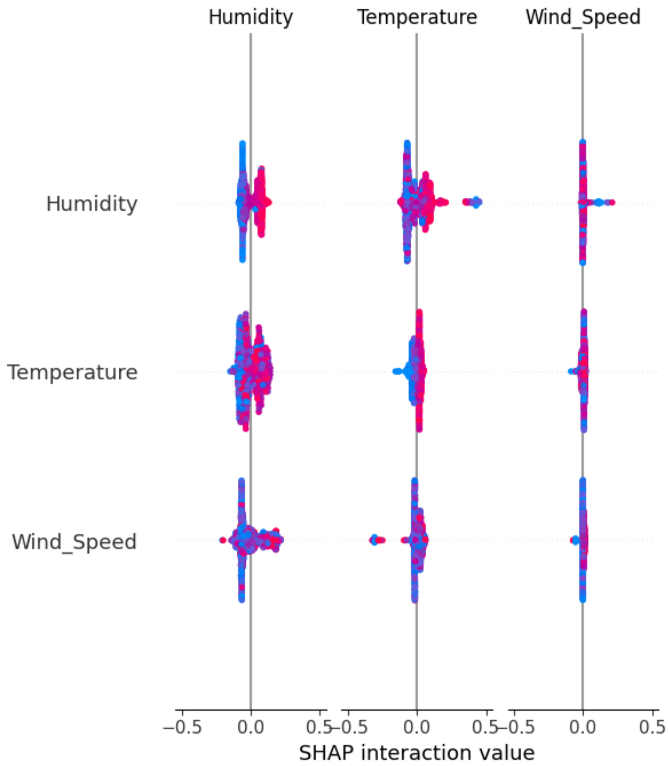


Fig. 6. SHAP summary plot for the Random Forest classifier.

high performance in weather-type classification using meteorological parameters. However, a deeper examination beyond overall accuracy reveals meaningful differences in robustness, stability, and interpretability between the two models.

The correlation matrix presented in Fig. 1 shows that several meteorological variables exhibit moderate to strong linear relationships, particularly between humidity and cloud cover. This finding is consistent with atmospheric processes, as increased cloud formation is commonly associated with higher air moisture content. From a classification perspective, such correlations imply that certain weather classes share overlapping feature spaces. This is especially evident for the Cloudy and Rainy categories, which often display similar humidity levels and cloud coverage, differing primarily in precipitation intensity.

Under these conditions, the MLP model demonstrates strong capability in modeling nonlinear relationships among variables. Nevertheless, the performance trend across different data split scenarios indicates moderate sensitivity to the proportion of training data. As the size of the training set decreases, accuracy declines gradually. This behavior can be explained by neural networks' gradient-based optimization mechanism. When input features are highly correlated, the optimization process may be sensitive to weight initialization and prone to local minima, particularly when class boundaries are not sharply separated, as in the transition between Cloudy and Rainy conditions.

In contrast, the Random Forest classifier exhibits more stable performance across all training–testing configurations. The ensemble learning mechanism underlying Random Forest plays a central role in this robustness. By constructing multiple decision trees using random subsets of data and features, the model reduces variance and prevents any single correlated variable from dominating classification. Consequently, correlated features such as humidity and cloud cover are distributed across the ensemble, enabling more reliable discrimination between overlapping weather classes.

The confusion matrix analysis further supports this interpretation. Although both models achieve low misclassification rates, Random Forest consistently shows fewer errors in distinguishing between the Cloudy and Rainy categories. This suggests that the tree-based partitioning strategy is particularly effective at capturing threshold-based decision rules, such as precipitation presence combined with high humidity, which are essential for separating similar atmospheric states.

Given that Random Forest achieves accuracy values approaching 99% in certain scenarios, the possibility of overfitting must be carefully considered. To address this concern, cross-validation experiments using 5-fold and 10-fold schemes were conducted. The results show that Random Forest maintains average accuracy values above 98% with minimal standard deviation across folds. The relatively small difference between hold-out testing performance and cross-validation results indicates that the model demonstrates strong generalization capability rather than memorizing specific training instances.

Meanwhile, the MLP model shows slightly lower average cross-validation accuracy than in a single hold-out evaluation. This difference suggests a mild tendency toward overfitting when trained on a fixed partition, which is consistent with the characteristics of neural network models applied to datasets with correlated inputs. The use of cross-validation in this study, therefore, provides additional assurance that the reported performance reflects genuine predictive ability.

From a methodological standpoint, the findings highlight the suitability of ensemble-based learning methods for meteorological classification tasks characterized by correlated variables and gradual atmospheric transitions. While MLP remains effective in capturing nonlinear feature interactions, Random Forest offers greater stability, reduced sensitivity to data partitioning, and improved robustness in distinguishing physically similar weather categories.

Overall, the analysis confirms that model selection in weather classification should consider not only accuracy metrics but also stability, interpretability, and resilience to correlated input features. In the context of the present dataset and experimental design, Random Forest provides a more reliable and operationally robust solution for multi-class weather type classification.

5. Conclusion

This study systematically evaluated the performance of the Multi-Layer Perceptron (MLP) and Random Forest (RF) classifiers for multi-class weather-type classification using nine years of historical meteorological data. The experimental results demonstrate that both models achieve high predictive performance in classifying weather conditions into sunny, cloudy, and rainy categories. However, Random Forest consistently outperformed MLP across all data split scenarios and cross-validation schemes, achieving accuracy levels of 98%-99% with lower error rates and minimal performance variance. These findings confirm that ensemble-based learning methods are particularly well-suited for handling the nonlinear, multivariate, and correlated characteristics of meteorological data.

From a methodological standpoint, this study contributes a structured comparative evaluation across multiple training–testing splits and validation strategies, providing robust evidence of model stability and generalization. The superior stability of Random Forest suggests that tree-based ensemble methods effectively mitigate variance and overfitting, especially in atmospheric datasets where inter-feature correlations and natural variability are prominent.

Despite these promising findings, several concrete limitations must be acknowledged. First, the classification framework relies exclusively on tabular historical meteorological variables and does not incorporate spatial, satellite-based, or radar-derived atmospheric features, which may contain additional predictive information. Second, the MLP architecture implemented in this study represents a conventional feedforward neural network and does not explicitly model temporal dependencies inherent in sequential weather observations. Advanced sequence-aware architectures, such as Temporal Fusion Transformers (TFT) or hybrid recurrent–attention models, could potentially capture dynamic temporal interactions more effectively and improve predictive robustness. Third, although SHAP-based interpretability analysis has been integrated to enhance model transparency, the current explainability assessment remains limited to global feature attribution. More advanced local interpretability analyses, interaction effect exploration, or comparative XAI benchmarking across multiple models could further strengthen decision transparency and operational reliability. Future research should therefore integrate model-agnostic interpretability frameworks, such as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-Agnostic Explanations), to quantify feature contributions and enhance transparency for meteorological practitioners.

Beyond methodological contributions, this study has clear operational implications. The proposed Random Forest-based classification framework can be implemented as an automated weather-labeling system for meteorological stations, particularly Automatic Weather Stations (AWS) that operate without continuous human observation. By transforming real-time meteorological sensor readings into categorized weather types, the model can serve as a decision-support component within operational forecasting systems at the Meteorological, Climatological, and Geophysical Agency (BMKG). Such an implementation could accelerate data standardization, reduce manual labeling workload, and support downstream applications such as early warning systems, climatological reporting, and data-driven public information services.

In summary, this research confirms that ensemble-based machine learning approaches, particularly Random Forest, provide a robust and practical solution for weather type classification based on meteorological parameters. Future work should extend the framework by incorporating spatio-temporal deep learning architectures, integrating XAI-based interpretability mechanisms, and validating the model across multiple geographic regions and real-time operational environments. These extensions would not only enhance predictive capability but also strengthen scientific transparency and operational trustworthiness in applied meteorological systems.

Data Availability

The dataset supporting this study has been deposited in Zenodo and is publicly available at: https://doi.org/10.5281/zenodo.18257312.

Declaration of Competing Interest

The authors declare that they have no known financial or personal competing interests that could have influenced the work reported in this article.

Authors' Contributions

YD: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Visualization. JT: Methodology, Validation, Resources, Writing – Review & Editing, Supervision, Project Administration, Funding acquisition. AZ: Methodology, Software, Data Curation, Review & Editing. TIS: Software, Validation, Formal analysis, Visualization.

Acknowledgements

The authors would like to express their gratitude to the research partners who have indirectly contributed by facilitating the execution of this study, particularly Institut Informatika dan Bisnis Darmajaya, for their roles in data analysis and processing. Appreciation is also extended to Badan Meteorologi, Klimatologi, Geofisika (BMKG) for providing the data essential for this research.

References

Abdulla, N., Demirci, M., & Ozdemir, S. (2022). Design and evaluation of adaptive deep learning models for weather forecasting. *Engineering Applications of Artificial Intelligence*, *116*, 105440. https://doi.org/10.1016/j.engappai.2022.105440

Aftab, B., Wang, Z., Wang, S., & Feng, Z. (2024). Application of a Multi-Layer Perceptron and Markov Chain Analysis-Based Hybrid Approach for Predicting and Monitoring LULCC Patterns Using Random Forest Classification in Jhelum District, Punjab, Pakistan. *Sensors*, *24*(17), 5648. https://doi.org/10.3390/s24175648

Aksan, F., Suresh, V., & Janik, P. (2025). PV Generation Prediction Using Multilayer Perceptron and Data Clustering for Energy Management Support. *Energies*, *18*(6), 1378. https://doi.org/10.3390/en18061378

Bhowmick, R., Trepanier, J. C., & Haberlie, A. M. (2023). Classification Analysis of Southwest Pacific Tropical Cyclone Intensity Changes Prior to Landfall. *Atmosphere*, *14*(2), 253. https://doi.org/10.3390/atmos14020253

Cahyani, N., Putri, W. A., & Irsyada, R. (2025). Improving Multiclass Rainfall Prediction with Multilayer Perceptron and SMOTE: Addressing Class Imbalance Challenges. *Brilliance: Research of Artificial Intelligence*, *4*(2), 901–908. https://doi.org/10.47709/brilliance.v4i2.5203

Ejike, O., Ndzi, D., & Shakir, M. Z. (2025). Comparative Study of Machine Learning-Based Rainfall Prediction in Tropical and Temperate Climates. *Climate*, *13*(8), 167. https://doi.org/10.3390/cli13080167

Feng, J., Toth, Z., Zhang, J., & Peña, M. (2024). Ensemble forecasting: A foray of dynamics into the realm of statistics. *Quarterly Journal of the Royal Meteorological Society*, *150*(762), 2537–2560. https://doi.org/10.1002/qj.4745

Mawalagedara, R., Ray, A., Das, P., Watson, J., Pal, A. K., Duffy, K., Bhatia, U., Aldrich, D. P., & Ganguly, A. R. (2025). Non-linear dynamical approaches for characterizing multi-sector climate impacts under irreducible uncertainty. *Npj Climate and Atmospheric Science*, *8*(1), 329. https://doi.org/10.1038/s41612-025-01208-4

Novogroder Idan. (2024). *Data Preprocessing in Machine Learning: Steps & Best Practices*. LakeFS. https://lakefs.io/blog/data-preprocessing-in-machine-learning/

Safia, M., Abbas, R., & Aslani, M. (2023). Classification of Weather Conditions Based on Supervised Learning for Swedish Cities. *Atmosphere*, *14*(7), 1174. https://doi.org/10.3390/atmos14071174

Senior-Williams, J., Hogervorst, F., Platen, E., Kuijt, A., Onderwaater, J., Tervo, R., John, V. O., & Okuyama, A. (2024). The Classification of Tropical Storm Systems in Infrared Geostationary Weather Satellite Images Using Transfer Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 5234–5244. https://doi.org/10.1109/JSTARS.2024.3365852

Shimada, U. (2024). Tropical Cyclone Intensity Forecasting with Three Multiple Linear Regression Models and Random Forest Classification. *Journal of the Meteorological Society of Japan*, *102*(5), 555–573. https://doi.org/10.2151/jmsj.2024-030

Simbo, R. T., Gogra, A. B., Kawa, Y. K., & Moiwo, P. J. (2023). Seasonal Effect of Weather Elements on Water Table Fluctuation in Potable Wells in Kono District, Eastern Sierra Leone. *Open Journal of Applied Sciences*, *13*(11), 2198–2209. https://doi.org/10.4236/ojapps.2023.1311171

Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, *237*, 121549. https://doi.org/10.1016/j.eswa.2023.121549

Tsagalidis, E., & Evangelidis, G. (2022). Exploiting Domain Knowledge to Address Class Imbalance in Meteorological Data Mining. *Applied Sciences (Switzerland)*, *12*(23), 12402. https://doi.org/10.3390/app122312402

Wiguna, S., Adriano, B., Mas, E., & Koshimura, S. (2024). Evaluation of Deep Learning Models for Building Damage Mapping in Emergency Response Settings. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 5651–5667. https://doi.org/10.1109/JSTARS.2024.3367853

Yang, R., Hu, J., Li, Z., Mu, J., Yu, T., Xia, J., Li, X., Dasgupta, A., & Xiong, H. (2024). Interpretable machine learning for weather and climate prediction: A review. *Atmospheric Environment*, *338*, 120797. https://doi.org/10.1016/j.atmosenv.2024.120797

Zhang, C. J., Chen, M. S., Ma, L. M., & Lu, X. Q. (2025). Deep Learning and Wavelet Transform Combined with Multichannel Satellite Images for Tropical Cyclone Intensity Estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 4711–4735. https://doi.org/10.1109/JSTARS.2025.3531448

Zhang, H., Liu, Y., Zhang, C., & Li, N. (2025). Machine Learning Methods for Weather Forecasting: A Survey. *Atmosphere*, *16*(1), 82. https://doi.org/10.3390/atmos16010082

Yoyok Dewantoro is a dedicated professional who uniquely merges academic pursuit with practical expertise. Currently advancing his knowledge as a student in the field of Informatics Engineering, he simultaneously holds a critical operational role as a Climatology Forecaster. This dual engagement positions him at the intersection of data technology and atmospheric sciences, where he leverages his growing technical skills in computing and data analysis to enhance the precision and methodology of weather forecasting and climate monitoring.

Joko Triloka is an integral part of the academic community at the Darmajaya Institute of Informatics and Business (Institut Informatika dan Bisnis Darmajaya), a leading private higher education institution in Bandar Lampung, Lampung Province. In his capacity as a lecturer and educator, he actively shapes the next generation of professionals in informatics and business disciplines. His role encompasses not only teaching but also contributing to the institute's academic and intellectual environment, which plays a significant role in the regional educational landscape.

Ahmad Zarkoni holds a Master's degree in Informatics from the State Islamic University of Maulana Malik Ibrahim Malang and serves as a practitioner at the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG), specifically at the Malang Geophysical Station. His expertise and research areas encompass seismic data analysis, machine learning applications for geophysical studies, and the development of disaster mitigation systems.

Tomy Ivan Sugiharto is pursuing a Master's degree in Informatics at the State Islamic University of Maulana Malik Ibrahim Malang. His research interests center on machine learning, deep learning, and geophysics, with a particular focus on leveraging artificial intelligence for data analysis, modeling, and interpretation within geophysical domains. He continues to advance interdisciplinary research integrating computational intelligence with earth sciences to develop modern geophysical technologies and analytical frameworks.