Agromet 40 (1): 13-27, 2026

Agromet
journal.ipb.ac.id/index.php/agromet
ISSN : 0126 - 3633

Designing A Rainfall Potential Prediction Model Based on Atmospheric Stability in Airport Areas

Yus Prihatinina[1, 2], Perdinan[1*], Akhmad Faqih[2], Supari[2]

[1] Department of Geophysics and Meteorology, IPB University, Bogor, Indonesia.
[2] Indonesian Agency of Meteorology, Climatology, and Geophysics (BMKG)

ARTICLE INFO

Received
15 September 2025

Revised
27 February 2026

Accepted for Publication
31 March 2026

Published
8 May 2026

doi: 10.29244/j.agromet.40.1.13-27

Correspondence:
Perdinan
Department of Geophysics and Meteorology, IPB University, Bogor, Indonesia.

Email: perdinan@apps.ipb.ac.id

ABSTRACT

Rainfall forecasting in airport areas plays a crucial role in ensuring flight safety and operational reliability, yet local characteristics effect on rainfall intensity are overlooked. This study aims to develop a rainfall occurrence and intensity prediction model using Artificial Neural Network (ANN) based on atmospheric stability indices including Showalter Index (SI), Lifted Index (LI), K Index (KI), Total Totals (TT), Convective Available Potential Energy (CAPE), Convective Inhibition (CIN), and Precipitable Water (PW). Data were obtained from radiosonde observations, Automated Weather Observing System (AWOS), and Meteorological Aerodrome Report (METAR) in Soekarno-Hatta International Airport January 2020-December 2024 at 00 and 12 UTC. We used two stage analysis (1) binary rainfall occurrence prediction and (2) three-class rainfall intensity prediction based on selected feature. Our results reveal that ANN is capable to simulate rainfall occurrence with high accuracy (>0.78), outperforming the minimal-feature model (0.738) and all other configurations. Physically, the atmospheric indices can be grouped into three categories: stability-related indices (SI, LI, KI, TT), energy-related indices (CAPE, CIN), and moisture-related indices (PW), representing key factors influencing convective rainfall in tropical regions. However, the model's applicability may be limited in study site due to local climatic characteristics and temporal constraints of the dataset. These findings highlight the importance of selecting physically relevant atmospheric parameters and implementing robust data preprocessing to enhance rainfall prediction accuracy in operational aviation contexts.

KEYWORDS

artificial neural network, aviation meteorology, atmospheric stability, airport weather forecasting, rainfall prediction

1. INTRODUCTION

Accurate rainfall forecasting is vital for airport safety and smooth operations, particularly during critical takeoff and landing phases. Operational tools like METAR and AWOS provide a real-time weather information for aviation activities (Alves et al., 2024; Shankar et al., 2025). However, their applicability for rainfall prediction in tropical regions remains limited due to complex interplays of local topography and atmospheric factors (Hidayat and Taufik, 2025; Wu and Takemi, 2023), that can trigger localized and short-lived convective events that are hard to capture in real time.

Rainfall formation is governed by complex interactions between atmospheric circulation, humidity, and vertical instability (Abdulkadir et al., 2020; Bruno Soares et al., 2018). Although atmospheric stability is a primary driver of convective clouds (Umakanth et al., 2020), extreme states of stability or instability often complicate standard predictive models (Cimini et al., 2015; Sanchez et al., 2024). This limitation is especially relevant for airport-scale forecasting, in which rainfall is develop rapidly and vary strongly over short distances (Ferrett et al., 2021; Sham et al., 2025).

Various approaches have been used to predict rainfall and convective weather, including statistical methods, numerical weather prediction, and machine learning. Linear regression and correlation analysis are commonly used to examine relationships between atmospheric predictors and rainfall because they provide interpretable measures of association. However, rainfall is generated by complex and non-linear interactions among atmospheric variables (An et al., 2023), which can reduce the effectiveness of traditional linear approaches.

Numerical models like Weather Research and Forecasting (WRF) try to represent these processes physically, but their performances depend heavily on how key physics are parameterized and how the model is configured (Fonseca et al., 2015; Taufik and Haikal, 2024). Meanwhile, advanced deep learning methods are increasingly used in meteorology, but they are commonly designed for large gridded datasets, such as radar, satellite, or numerical model outputs, to learn complex spatiotemporal rainfall patterns (An et al., 2025; Chen et al., 2020). These constraints suggest the need for a computationally practical data-driven approach that can be applied to station-based airport observations.

Artificial Neural Networks (ANNs) are relevant in this context because they can approximate non-linear functions and learn relationships between meteorological predictors and rainfall outcomes (Haidar and Verma, 2018). Compared with purely linear approaches, ANNs provide greater flexibility in representing continuous and non-linear relationships among atmospheric stability indices, making them suitable for rainfall prediction using station-scale atmospheric predictors (Çınaroğlu and Unutulmaz, 2019). Given the station-based nature of the present dataset and the focus on physically interpretable stability indices, ANN was selected as an appropriate and computationally efficient method for airport-level rainfall forecasting.

Previous studies have emphasized the significance of atmospheric stability indices in predicting convective phenomena. High CAPE values are closely linked to thunderstorm occurrences in tropical and subtropical regions (Brooks and Dotzek, 2008; Emanuel, 1994). Various machine learning approaches have also been applied to related problems. For instance, Guijo-Rubio et al., (2020a) employed Support Vector for Ordinal Regression (SVOR) to predict CB clouds at Madrid-Barajas Airport and achieved better results than conventional TAF forecasts, Manzato (2005) utilized ANN with radiosonde and lightning data, demonstrating the benefit of selective feature inputs to minimize overfitting. Similarly, cumulonimbus cloud prediction using radiosonde data from Surabaya, showing improved accuracy when CAPE was included among the predictors (Putra et al., 2021).

Yet, most of these models were designed for regional or synoptic contexts and rarely incorporated aviation-specific datasets such as AWOS, METAR, and radiosonde in an integrated framework, limiting their direct applicability to airport-level rainfall forecasting. This study integrates multiple atmospheric stability indices (SI, LI, KI, TT, CAPE, CIN, and PW) within a dual-stage Artificial Neural Network (ANN) framework designed to separately predict rainfall occurrence and intensity. The systematic optimization and validation across training, validation, and testing phases aim to enhance prediction reliability, while the incorporation of a feature selection process combining statistical correlation, pairwise visualization, and Principal Component Analysis (PCA) improves model efficiency and interpretability.

2. MATERIALS AND METHODS

2.1 Study Area

This research used case study at the Soekarno-Hatta International Airport, located in Cengkareng, Tangerang City, Banten Province (Figure 1). The airport is equipped with radiosonde observation facilities and Automated Weather Observing Systems (AWOS), which provide vertical atmospheric profiles and local meteorological observations for aviation operations.

The study site is situated in a lowland area below 15 m above sea level and experiences a tropical monsoon climate. Annual rainfall is around 2000 mm, which is ~28% lower than West Java (Primartono et al., 2025). The rainy season generally occurs on December–February (DJF) and dry season during June–August (JJA) (Siswanto et al., 2023).

2.2 Datasets

This study used a five-year dataset from 2020 to 2024 consisting of predictor and target variables. The predictor variables were derived from radiosonde observations and represented atmospheric stability conditions, while the target variables were rainfall occurrence and rainfall intensity obtained from AWOS and METAR reports at Soekarno-Hatta International Airport. METAR is a coded aviation weather report issued at regular intervals, commonly every 30 minutes, and contains information on airport weather conditions, including present weather and rainfall intensity.

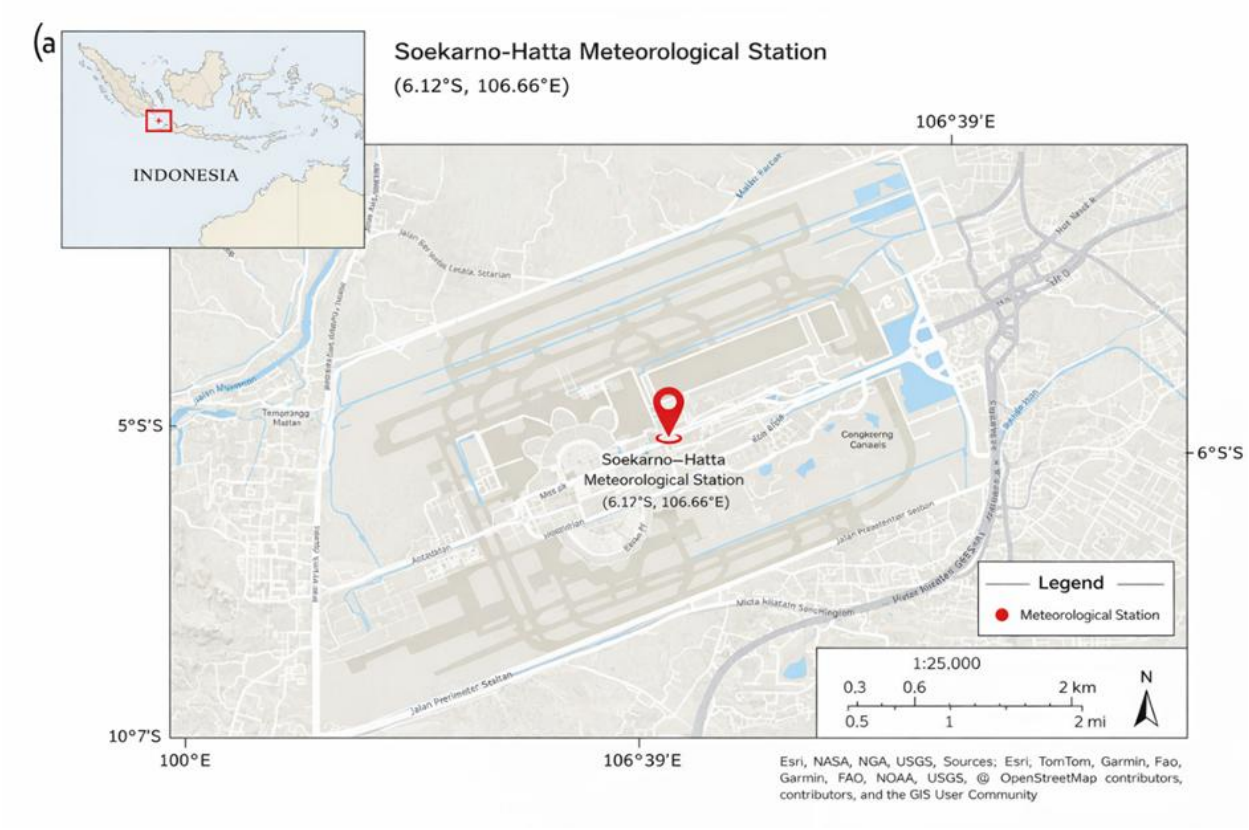


Figure 1 Study site at Soekarno-Hatta International Airport, Tangerang Indonesia

In this study, rainfall occurrence was identified from rainfall related METAR codes, while rainfall intensity was classified based on the reported rain categories. Following ICAO Annex 3, rainfall intensity was categorized into three classes: light rain, moderate rain, and heavy rain (ICAO, 2018). Light rain refers to rainfall intensity below 2.5 mm h^{-1}, moderate rain refers to rainfall intensity between 2.5 and 10 mm h^{-1}, and heavy rain refers to rainfall intensity above 10 mm h^{-1}.

The predictor data consisted of radiosonde-derived atmospheric stability indices, including the Lifted Index (LI), Showalter Index (SI), K-Index (KI), Total Totals Index (TT), Convective Available Potential Energy (CAPE), Convective Inhibition (CIN), and Precipitable Water (PW). LI and SI describe atmospheric instability in the lower-to-middle troposphere, where lower or negative values indicate stronger convective potential. KI and TT combine information on temperature and humidity in the lower and middle atmosphere to estimate the likelihood of convective rainfall or thunderstorms. CAPE represents the amount of buoyant energy available for convection, while CIN indicates the energy barrier that suppresses convective development. PW represents the total water vapor content in an atmospheric column and provides an indication of moisture availability for cloud formation and rainfall (Sanchez et al., 2024).

2.3 Data Analysis

2.3.1 Model pre-processing

The research began with the collection and processing of data to construct predictor and target datasets, as shown in Figure 2. The initial step included descriptive statistical analysis to assess data completeness and quality. Additionally, inter-variable correlation tests were conducted to identify potential multicollinearity or redundancy. Relationships among atmospheric stability indices were analyzed using pairwise plots or scatterplot matrices to visualize the potential correlations and linkages between variables. This analysis used a correlation matrix and pairwise visualizations to observe the interrelation patterns among the indices, as well as their respective relationships to rainfall occurrence and intensity.

Beyond correlation, the relative contribution of each variable was analyzed using Principal Component Analysis (PCA) to assess the extent to which each index explains data variance and how inter-variable relationships are structured (Jolliffe dan Cadima 2016). In meteorology, PCA, also known as Empirical Orthogonal Functions, has been widely used to identify dominant variability modes in atmospheric parameters (Hannachi et al. 2007), aiding in data interpretation. In this study, PCA was used purely as an exploratory tool, not for dimensionality reduction, ensuring that each index's physical significance is preserved.

Variables showing very high correlation with others could introduce bias during model training and increase the risk of overfitting. Therefore, feature selection was guided by three considerations: (1) strength of relationship to the target (rainfall occurrence/intensity), (2) high inter-feature correlation, and (3) meteorological theory. Some variables were retained despite weak statistical correlation if they were physically important in rainfall formation. Conversely, variables with low contribution both statistically and meteorologically were removed to simplify the model and improve generalization.

Figure 2 summarizes the step-by-step process used to evaluate and select atmospheric stability indices, starting with data quality checks, correlation analysis, pairwise visualization, and PCA diagnostics, followed by meteorological physics reasoning. The final feature selection is based on three criteria: (1) statistical relevance to the target rainfall occurrence or intensity, (2) redundancy or multicollinearity between indices, and (3) physical relevance to convective processes. This workflow integrates statistical and meteorological considerations to ensure that the selected features are both quantitatively significant and physically meaningful for rainfall prediction.

The dataset was split into three subsets: training data, validation data, and testing data. A total of five years of data (2020–2024) were used in this study, with the first three years (2020–2022) allocated for training, and two years (2023-2024) for independent testing. This temporal division ensures that the model generalizes across different seasonal conditions and avoids temporal leakage between datasets. For the training data, class imbalance was addressed using the SMOTE (Synthetic Minority Oversampling Technique) method. Several studies have shown that combining SMOTE with neural networks significantly improves AUC and F1-score metrics (Aditya Gumilar et al., 2022; Kandasamy et al., 2025). Additionally, hybrid approaches involving ANN have been reported to enhance the accuracy of annual rainfall prediction (Hall et al., 1999). Therefore, in this study, input integration principles and neural architecture design were considered in model development.

Because the rainfall prediction framework consists of two independent Artificial Neural Network (ANN) models, the SMOTE balancing process was applied separately to each. The first model, designed to predict rainfall occurrence, used a binary target (rain and no rain) and applied SMOTE to address the imbalance between rainy and non-rainy observations. The second model, developed to classify rainfall intensity, was trained only on rainfall cases with three categorical

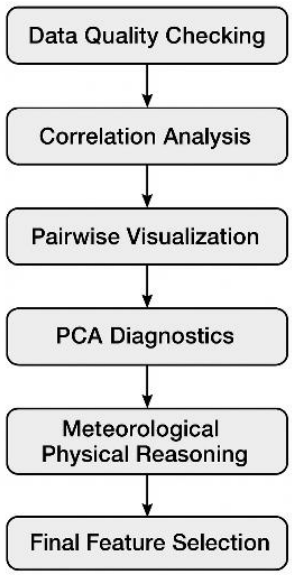


Figure 2 Feature selection workflow

targets (RA−, RA, and RA+). In this case, SMOTE was applied exclusively within the rainfall subset to balance the distribution of light, moderate, and heavy rainfall events.

A custom adaptive SMOTE function was implemented using a k-nearest neighbours approach (k = 5–7) with a conservative sampling strategy based on the average class count. This ensured that synthetic samples were generated without oversaturating minority classes. The method produced balanced training datasets for each model while maintaining the physical representativeness of the meteorological data and avoiding overfitting.

The training and validation data were then normalized to a range of (-1, 1) to ensure that all variables were on a uniform scale, using Equation (1).

$$x' = 2 \cdot \frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1 \qquad (1)$$

In the implementation of the Artificial Neural Network (ANN), selecting the appropriate number of hidden layers and neurons is crucial for achieving a balance between generalization capability and model complexity. An architecture that is too large may lead to overfitting, while an architecture that is too small may lack the capacity to represent complex patterns in atmospheric data. A study by Kim et al. (2017) showed that the use of optimization techniques based on adaptive learning rates can improve convergence and training stability in neural network models, particularly when working with medium-sized weather datasets (Kim et al., 2017).

2.3.2 Model Design

The predictive approach used in this study consists of two stages. The first stage is the Rain Occurrence Model, which determines whether rainfall will occur or not. If the prediction indicates no rain, the process stops there. However, if rainfall is predicted, it proceeds to the second stage, identified as the Rainfall Intensity Model, which classifies the rainfall into three categories: light, moderate, and heavy.

To ensure optimal performance of both models, a grid search optimization was implemented to identify the best combination of ANN hyperparameters. A total of 45 configurations were tested, combining hidden layer structures ([10], [20], [10,5], [20,10], [30,15]), learning rates (0.01, 0.1, 0.3), and training algorithms (trainlm, trainbr, trainscg). Activation functions used were tansig for hidden layers and softmax for outputs. Early stopping was applied to prevent overfitting. This architectural design ensures that the ANN can effectively capture the nonlinear relationships among atmospheric stability indices while maintaining computational efficiency. The summary of hyper-parameter configurations is presented in Table 1.

Table 1 Hyperparameter configurations

Parameter Category	Tested Values (Grid Search)
Hidden Layer Structure	[10], [20], [10,5], [20,10], [30,15]
Learning Rate	0.01, 0.1, 0.3
Training Function (Optimizer)	trainlm, trainbr, trainscg
SMOTE Oversampling	N = 200%, k = 5–7

The rain occurrence prediction model in this study aims to determine the likelihood of rainfall based on atmospheric conditions derived from upper-air observations. Rain occurrence is the result of complex physical atmospheric processes involving interactions between temperature, humidity, and air pressure across various layers of the troposphere (Stensrud 2007). In this study, the rain occurrence model is built using seven atmospheric parameters grouped into three categories: atmospheric stability, atmospheric energy, and water vapor content. These parameters were selected based on their physical relevance to the processes of convection and rainfall formation (Emanuel 1994).

Unlike the occurrence model, the rainfall intensity model places more emphasis on input parameters that are more focused and representative of the rain formation process in terms of intensity scale, namely atmospheric energy and water vapor content, as these two factors play a key role in determining the strength of convection that produces rainfall of varying intensities (Wilks 2006).

All training, validation, and testing processes were conducted separately based on two main observation times, 00 UTC and 12 UTC, to capture the different atmospheric characteristics at each observation hour. Through this separation, the resulting model better represents the atmospheric dynamics of each observation time. This stepwise approach and time-based separation are expected to simplify the modeling complexity and enhance the accuracy of rainfall prediction.

Table 2 Atmospheric parameters

Prediction	Variable Category	Parameters
Rainfall Occurrence Model	Atmospheric Stability	LI, SI, TT, KI
	Atmospheric Energy	CAPE, CIN
	Water Vapor Content	PW
Rainfall Intensity Model	Atmospheric Energy	CAPE, CIN
	Water Vapor Content	PW

To minimize temporal dependencies, the 00 UTC and 12 UTC data sets were modeled independently to reflect different diurnal atmospheric patterns. Furthermore, by dividing the data set chronologically (three years for training, one year for validation, and one year for testing), autocorrelation between consecutive observations can be avoided. This temporal separation ensures that model evaluation reflects true generalization, not memorization of temporally adjacent samples.

2.3.3 Model Evaluation

The performance of the model was evaluated in stages using validation and test data. For the rainfall occurrence model, evaluation metrics included accuracy, F1-score (Equation 2), confusion matrix, and ROC-AUC. For the rainfall intensity model, the evaluation included F1-score per class, MSE (Equation 3), RMSE (Equation 4), and Brier Score (Equation 5). The results from the validation stage served as the basis for model optimization and calibration to improve prediction accuracy. After calibration, the model was re-evaluated using the test data to assess its performance on previously unseen data

$$F1 = 2.\frac{Precision.Recall}{Precision + Recall} \quad (2)$$

Precision = TP / (TP + FP) and Recall = TP / (TP + FN), where TP is true positive, FP is false positive, and FN is false negative.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2} \quad (4)$$

Where y_i is the actual value, $\hat{y}_i$ is the prediction value, and n is the number of samples.

$$BS = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2 \quad (5)$$

Where p_i is the predicted probability, o_i is the binary observation (1 for rainfall, 0 for no rainfall), and n is the number of samples.

Following model validation, probability calibration was performed to refine the classification thresholds. Using the validation dataset, optimal cutoffs were determined by maximizing Youden's J statistic from the ROC curve, ensuring balanced sensitivity and specificity. This process was implemented using a MATLAB-based function that derived optimal thresholds for each rainfall class based on probabilistic model outputs. The calibrated thresholds improved decision reliability and probability interpretation for operational rainfall forecasting.

Model performance evaluation in weather prediction is not limited to accuracy metrics alone, but must also consider the reliability of the generated probabilities. This is crucial since operational decisions, such as flight risk mitigation, depend on the confidence level of the predictions. Thupeng et al. (2023) emphasized that model testing based on probability calibration and probabilistic metrics, such as the Brier Score, can provide a more comprehensive overview of the reliability of weather prediction models in real-world scenarios (Thupeng et al., 2023).

To assess the statistical robustness of the final model, 95% confidence intervals were estimated using a non-parametric bootstrap procedure with 1,000 resamples of the independent test dataset. Confidence intervals were computed for key performance metrics, including Accuracy, Precision, Recall, F1-score, AUC, and Brier Score. In addition, predictive (epistemic) uncertainty was quantified using Monte Carlo Dropout with 100 stochastic forward passes during inference. These complementary approaches ensure that model performance is interpreted not only through point estimates but also with reliable uncertainty bounds. The entire modeling process was conducted separately for 00 UTC and 12 UTC to ensure the model captures the differences in atmospheric characteristics at different observation times.

3. RESULTS AND DISCUSSION

3.1 Data Analysis and Feature Selection

The dataset used in this study consists of observational data from the period 2020 to 2024. The data comprise a combination of atmospheric parameters and rainfall occurrence observations, divided into two main observation times: 00 UTC and 12 UTC. All data were collected at 12-hour intervals, resulting in a total of 2,878 training data for rainfall occurrence, 1,218 training data for rainfall intensity, 682 validation data, and 546 testing data.

Table 3 Summary of observational data distribution for period 2020-2024

Model	Time	Data Count	Class Distribution
Rainfall Occurrence	00 UTC	2253	1327 No rain, 926 Rain
Rainfall Occurrence	12 UTC	1853	1151 No Rain, 702 Rain
Rainfall Intensity	00 UTC	926	187 RA-, 729 RA, 10 RA+
Rainfall Intensity	12 UTC	702	114 RA-, 576 RA, 12 RA+

Based on the Spearman correlation analysis (Table 4), the PW variable shows the highest positive correlation with rainfall occurrence, followed by KI, which has a fairly strong relationship. CIN and TT exhibit moderate positive correlations, whereas LI shows a weak but significant negative correlation. CAPE only demonstrates a weak positive relationship, while SI shows a relatively stronger negative correlation than LI. These results indicate that PW, KI, and CIN have the potential to be more relevant predictors in modeling. In contrast, variables with low correlations need further evaluation to determine their contribution to the model.

Pairwise plots in Figure 3 provide an initial overview for feature selection by showing relationships between variables, distributions, and value differences based on rainfall categories. High correlations, such as between KI and PW or TT and SI, indicate potential multicollinearity. Therefore, one of the variables in each pair may be eliminated to avoid redundancy.

Conversely, variables showing a clear difference in median values across rainfall categories, such as PW, KI, and CAPE, are worth retaining due to their potential discriminative power. Skewed distributions, such as in CIN, suggest the need for transformation to meet the

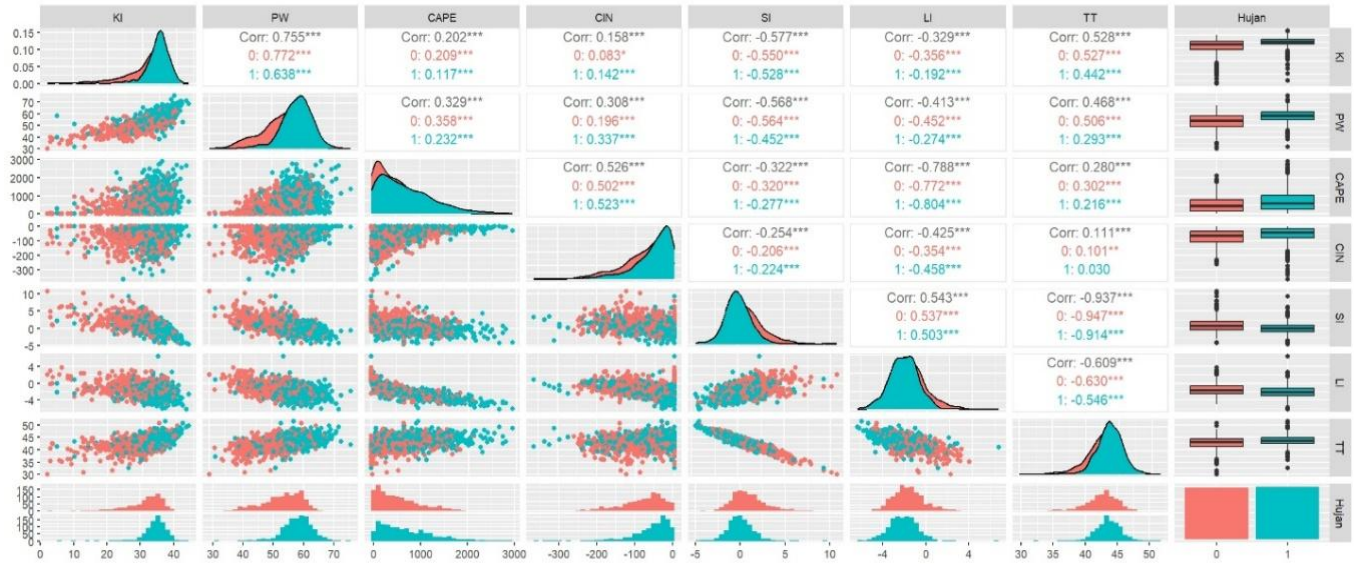


Figure 3 Analysis of relationships, distributions, and differences in atmospheric variable values based on rainfall occurrence, for the period 2020–2024.

model assumptions. Additionally, the color-coded scatter plots reveal a fairly clear separation between rain and no-rain cases for several variables, making them strong candidates for the model. Overall, the pairwise plots help identify relevant features, reduce redundancy, and assess data distribution before modeling.

Table 4 Spearman Correlation between Parameters and Rainfall Occurrence (2020–2024)

Index	Spearman's ρ	P-value	Description
PW	0.394	<0.001	Highest correlation, positive
KI	0.316	<0.001	Fairly strong
CIN	0.233	<0.001	Moderate
TT	0.205	<0.001	Moderate
LI	-0.167	<0.001	Weak negative, significant
CAPE	0.161	<0.001	Weak
SI	-0.252	<0.001	Negative correlation, relatively strong

The PCA analysis in Figure A1(a) reveals a fairly distinct distribution of samples in the space defined by the two principal components, indicating that most of the data variability can be explained using these two dimensions alone. Meanwhile, Figure A1(b) illustrates the contribution of each variable to these components, with LI, SI, CAPE, and CIN showing the most significant influence. Relationships between variables are also evident, such as the positive correlation between KI and PW, and the negative correlation between LI and CIN. In general, the contribution of each variable appears relatively balanced, suggesting that feature elimination requires further consideration based on statistical correlations and meteorological relevance.

Based on feature selection results from Spearman correlation, pairwise visualization, and PCA analysis (Table A1), variables with the lowest contribution and relevance were gradually eliminated: LI was removed first due to its weak correlation and because its stability information was deemed adequately represented by SI; TT was the next to be eliminated, as it plays a greater role in extreme weather conditions rather than general rainfall events; KI was then excluded due to its moisture, stability information being partly redundant with the combination of SI and CAPE; lastly, CIN was considered for elimination due to its relatively small statistical contribution in this dataset and its stronger influence on the initiation of extreme convection. From this elimination process, the final retained feature combination consists of SI, PW, and CAPE. These three variables complement each other: SI indicates whether the air tends to rise, CAPE shows the potential strength of convection, and PW reflects the amount of water vapor available to produce rainfall.

The Showalter Index (SI) is an atmospheric stability parameter that calculates the temperature difference between an air parcel lifted adiabatically and the surrounding environment at the 500 hPa level. Low or negative SI values indicate a highly unstable atmosphere, which supports the development of medium to large-scale convection. Studies by Showalter (1953) and Huntrieser et al. (1997) show that SI has a strong correlation with the likelihood of thunderstorms, especially in regions with moderate to high humidity. Although SI is related to the Lifted Index

(LI), it is often considered a more conservative indicator for detecting severe weather potential in tropical regions (Huntrieser et al., 1997; Showalter, 1953).

Precipitable Water (PW) represents the total water vapor content in an atmospheric column, which serves as the primary ingredient for cloud and rain formation. Research by Trenberth et al. (2003) and Holloway & Neelin (2009) indicates that high PW increases the probability of convective rainfall, particularly in tropical areas with atmospheric circulation that supports convergence (Holloway and Neelin, 2009; Trenberth et al., 2003). A rise in PW is also directly associated with the intensity of heavy rainfall in coastal regions (Yoo et al., 2011).

Convective Available Potential Energy (CAPE) is a key indicator of the potential energy for the development of large convective clouds. The study by Seeley & Romps (2015) found that CAPE is strongly correlated with the intensity of thunderstorms and extreme rainfall in tropical regions (Seeley and Romps, 2015). Johnson & Ciesielski (2000) further noted that sufficiently high CAPE levels often trigger localized rainfall in tropical areas, even though daily CAPE variations are relatively small (Johnson and Ciesielski, 2000). Considering the complementary physical roles of SI (instability), CAPE (convective energy), and PW (moisture), these three parameters are retained as the main predictors in the model.

3.2 ANN Model Performance and Evaluation

Once the final feature combination was obtained, each variable elimination stage produced a different set of inputs, which were then used to build five scenarios of rainfall occurrence and intensity prediction models, as shown in Table A2. Each scenario represents a feature combination at a specific elimination stage, starting from a model using all initial variables to one that uses only the three main features selected in the final stage. Evaluation of all these scenarios was conducted to compare the effect of the number and types of input variables on the model's performance, aiming to determine whether reducing features can maintain or even improve prediction performance.

Based on the evaluation results, Model 3 demonstrated the most balanced performance among the five tested scenarios. This model achieved the highest test accuracy with the lowest precision and Brier Score for rainfall occurrence prediction. The balance between precision and recall indicates that the model can provide fairly accurate rainfall predictions without generating excessive false alarms, similar to the findings of Devi et al. (2016), who tested various ANN architectures for daily rainfall prediction (Devi et al., 2016). The advantage of Model 3 can be attributed to the combination of complementary variables: SI represents the tendency of air to rise (instability), KI and CIN contribute to measuring moisture and convective inhibition, CAPE represents the potential for convective energy, and PW represents the availability of water vapor. Meteorologically, these five indices together form a predictive framework that covers aspects of atmospheric energy, moisture, and stability, thereby enhancing prediction quality.

Table 5 Performance model 3 for rain occurrence with 95% confidence intervals

Metric	Value	95% CI (Bootstrap)
Accuracy	0.782	0.7418 - 0.8114
Precision	0.665	0.5948 - 0.7259
Recall	0.686	0.6257 - 0.7558
F1-Score	0.684	0.6216 - 0.7287
AUC	0.806	0.7503 - 0.8326
Brier Score	0.201	0.1728 - 0.2372

Table 5 shows the statistical uncertainty of the selected final model (Model 3) for rainfall occurrence. The 95% confidence intervals are relatively narrow (e.g., Accuracy = 0.782, 95% CI: 0.7418–0.8114; AUC = 0.806, 95% CI: 0.7503–0.8326), indicating stable and reliable performance on the independent test dataset. The confidence intervals also indicate that the numerical differences between the five candidate models in Table A2 may not be statistically significant, as several metrics show overlapping ranges of uncertainty. Therefore, the selection of Model 3 is based not only on point estimates but also on its stability and meteorological interpretability.

For rainfall intensity classification, Model 3 also yielded the best multi-class accuracy, with the highest F1-score for the heavy rain category (RA+) on the test data. This demonstrates its ability to capture meteorological signals associated with extreme rainfall events, despite the relatively small number of RA+ samples. The low Brier Score across most categories also indicates that the predicted probabilities are well-calibrated, making the model more reliable for operational needs. Meanwhile, Model 5, which used the minimal feature set, did produce the lowest RMSE for intensity, but the decrease in rainfall occurrence accuracy was still considerable, which made it less ideal for early warning systems at airports.

The performance of Model 3 on the test data is visualized through the confusion matrix (Figure A1), which shows a balance between True Positives (TP) and True Negatives (TN). The confusion matrix on the test data in Figure A1(a) indicates that the rainfall

occurrence model has a good balance between TP and TN, with relatively low misclassification rates (False Positives and False Negatives). In the multi-class rainfall intensity model in Figure A1(b), the light rain (RA-) and moderate rain (RA) categories are predicted fairly well, while heavy rain (RA+) remains difficult to detect consistently. This limitation is due to the very small number of RA+ data samples, which restricts the model's ability to learn the characteristic patterns of heavy rainfall.

The robustness and generalization of the rainfall occurrence model were further evaluated using ROC curve analysis, as shown in Figure 4(a)**.** Both validation and test datasets exhibited consistent ROC curve shapes with AUC values of 0.773 and 0.789**,** respectively, indicating strong discrimination ability and minimal overfitting tendency. The close AUC values between validation and testing suggest that the model performs stably across unseen data. Model 3**,** which integrates five key atmospheric indices (SI, KI, CAPE, CIN, PW), demonstrated the best generalization capability, while the minimal-feature model (SI, CAPE, PW) achieved comparable results (AUC = 0.713–0.786), confirming that the ANN architecture remained robust even under reduced input complexity. This consistency indicates that the early stopping, cross validation, and SMOTE balancing procedures were effective in preventing overfitting.

Meanwhile, a separate analysis based on observation time (Figure 7b) showed AUC values >0.77 for both time periods (00 UTC and 12 UTC). This indicates that the ANN model is capable of recognizing different atmospheric patterns between morning and afternoon observations without experiencing a significant drop in performance.

3.3 Meteorological Interpretation and Results Comparison

The combination of SI, CAPE, and PW forms a predictive framework that encompasses the three main components of convective rainfall development in tropical regions: atmospheric instability (the tendency of air to rise), available convective energy, and moisture content. Beyond the theoretical roles of SI, CAPE, and PW, this study identifies several empirical patterns that are specific to the Soekarno–Hatta airport environment. In our dataset, light and moderate rain events commonly occur under moderately negative SI values (-1 to -3), CAPE in the range of 300–800 J/kg, and PW above 45 mm, whereas heavy rain (RA+) typically appears only when PW exceeds 55 mm and CAPE surpasses approximately 900 J/kg. These thresholds differ slightly from typical tropical convection values noted in earlier studies, suggesting that the airport's coastal microclimate plays a significant role in modifying atmospheric instability behavior (Petersen and Rutledge, 2001; Williams and Renno, 1993). The combined land–sea breeze circulation common in coastal regions enhances low-level moisture convergence, which may raise PW values and simultaneously reduce the SI and CAPE thresholds required for convective initiation (Zhu et al., 2017). This mechanism is reflected in the model's misclassification patterns, where RA+ cases with marginal SI or CAPE values tend to be predicted as moderate rainfall.

To connect these findings with operational forecasting, we propose an operational interpretation of the SI–CAPE–PW interactions. In a TAFOR workflow, the two-stage ANN model developed in this study can serve as a decision-support tool that highlights high-risk convective scenarios based on these indices. The model's relatively low Brier Scores indicate reliable probability outputs, supporting probabilistic guidance for balancing false alarms and missed events crucial in aviation forecasting (WMO, 2015). This integration pathway underscores the practical value of the model beyond numerical performance.

A dedicated comparison with previous studies was conducted to contextualize the performance of the proposed ANN model. Manzato (2005), who employed a large set of sounding-derived indices and lightning observations for short-term thunderstorm forecasting, reported ROC–AUC values typically ranging between 0.70–0.80, depending on storm type. Meanwhile, using five instability indices (SI, LI, KI, TT, CAPE) for cumulonimbus prediction in Surabaya, reported an accuracy of 80% and a FAR of 17% (Putra et al., 2021). In comparison, our best ANN model achieves an AUC of 0.8238 and an accuracy of 0.7821 for rainfall occurrence, indicating comparable or slightly improved performance despite using a more compact and physically interpretable feature set. Importantly, these studies differ substantially in input variables and spatial/temporal scales: (Manzato, 2005) operates at a mesoscale convective environment, (Putra et al., 2021) focuses on cloud-type detection at a daily timescale, while our model targets airport-scale rainfall prediction using aviation-specific datasets (AWOS, METAR, radiosonde).

Our study contributes a feature-selection framework and dataset integration strategy specifically tailored for airport forecasting. Unlike SVM or tree-based methods, ANN better captures the nonlinear interactions between stability, moisture, and energy parameters. Nevertheless, we acknowledge that a formal benchmarking exercise involving multiple algorithms remains an important direction for future research.

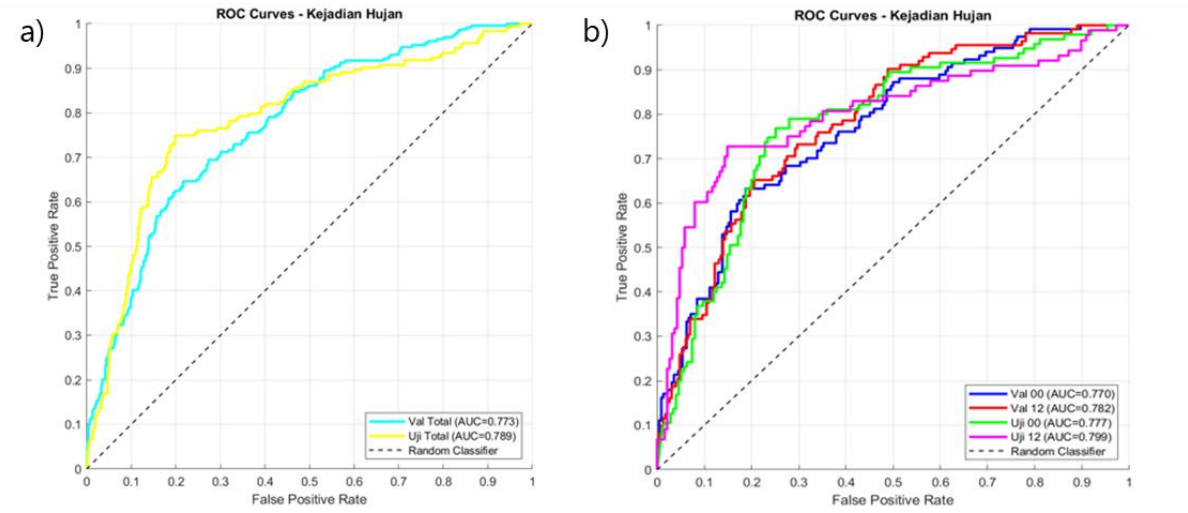


Figure 4 (a) ROC curve of the ANN model for rain event prediction on combined test data, (b) ROC curve of the ANN model for rain event prediction based on observation time

Accordingly, statements implying superiority have been softened, and the results are framed as providing comparable performance with added operational relevance and interpretability. These findings highlight the importance of feature selection based on both statistical and meteorological analyses to optimize model performance while preserving interpretability, particularly in the context of airport weather forecasting. However, several limitations remain:

1. The number of heavy rain (RA+) samples is limited, which affects the model's sensitivity to this category; adding more data could improve its detection.
2. Meso-scale dynamic variables, such as wind shear, have not been considered.
3. The potential for accuracy improvement through more complex ANN architectures or ensemble methods has not yet been explored.

These limitations have several quantitative and practical implications for the model's performance. The small number of RA+ samples leads to consistent under detection of heavy rainfall, resulting in lower recall and wider confidence intervals for this category. The limited representation of mesoscale processes such as wind shear and boundary-layer convergence also contributes to increased prediction variance in borderline convective cases. In terms of applicability, the model is expected to generalize best to airports with similar tropical coastal characteristics, and therefore, deployment in different climatic regimes would require retraining or recalibration. To address these issues, future research could explore ensemble strategies (e.g., averaging multiple ANN runs or combining ANN with tree-based models) and hybrid approaches that integrate radar or satellite-based predictors, as well as techniques such as cost-sensitive learning or synthetic oversampling to improve RA+ representation during training.

4. CONCLUSION

This study successfully developed a rainfall occurrence and intensity prediction model based on Artificial Neural Networks (ANN) using atmospheric stability index variables. Evaluation results show that the best-performing model is Model 3, which uses the combination of SI, KI, CAPE, CIN, and PW variables. This model achieved the highest performance for rainfall occurrence prediction, with an accuracy of 0.7821, AUC of 0.8238, precision of 0.6649, recall of 0.7049, and Brier Score of 0.2013 on the test data. For rainfall intensity prediction, this model also excelled with an accuracy of 0.6186, RMSE of 0.7017, MSE of 0.492, and an F1-Score for RA+ (heavy rain) of 0.4961, indicating a good balance between detecting low and high intensity events. In addition, a model using the minimal combination of variables (SI, CAPE, and PW) also produced competitive results, with rainfall occurrence accuracy of 0.7381 and rainfall intensity accuracy of 0.5752, offering a practical alternative for data-limited operational environments.

This study demonstrates that a suite of predictors based on physical conditions (SI, CAPE, PW, complemented by KI and CIN) captures the dominant convection signatures relevant to airport-scale rainfall. This suggests that convection processes influencing rainfall in tropical aviation environments can be effectively characterized using a limited but meteorologically meaningful set of features.

These findings further demonstrate that the two-stage ANN structure (rainfall occurrence followed by

intensity classification) improves operational relevance by mirroring decision-making workflows in aviation weather The forecasting, where the detection of rainfall initiation typically precedes intensity assessment. This framework provides a structured means of integrating METAR, AWOS, and upper-air data into a localized, airport-focused prediction system. A key scientific implication of this work is that reduced-variable models retain interpretability while sustaining acceptable predictive skill. This highlights a quantifiable trade-off between model complexity, computational efficiency, and operational feasibility,

Despite its advantages, the model has limitations. The RA+ class remains underrepresented, leading to larger uncertainty in intensity classification. The exclusion of mesoscale dynamical factors (e.g., wind shear, storm-relative flow) may limit the model's ability to represent organized convection. Furthermore, the model was trained using data from a single tropical airport, constraining its external validity.

For future work, we recommend methodological enhancements including the use of ensemble or hybrid learning architectures, the implementation of temporal cross-validation strategies such as year-wise or rolling-origin schemes to evaluate temporal generalizability, and the incorporation of uncertainty quantification tools such as prediction intervals, reliability diagrams, and calibrated probability scoring systems. Expanding the dataset to multiple airports with different climatic regimes would also enable a more robust assessment of model transferability. Overall, this study contributes to the literature by demonstrating that a systematically selected, physically interpretable subset of atmospheric stability indices can support statistically reliable, operationally applicable ANN-based rainfall prediction at the airport scale.

ACKNOWLEDGEMENT

The authors sincerely acknowledge the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) and the Soekarno-Hatta International Airport meteorological station for providing radiosonde, AWOS, and METAR observations used in this study. The authors also thank all individuals and institutions who provided technical support and constructive input during the preparation of this research.

REFERENCES

Aditya Gumilar, Sri Suryani Prasetiyowati, Yuliant Sibaroni, 2022. Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification). J. RESTI (Rekayasa Sist. Teknol. Inf.) 6, 481–490. https://doi.org /10.29207/resti.v6i3.4142

Alves, D., Mendonça, F., Mostafa, S.S., Morgado-Dias, F., 2024. Low tropospheric wind forecasts in aviation: the potential of deep learning for terminal aerodrome forecast bulletins. Pure and Applied Geophysics 181, 2265–2276.

An, P., Li, Y., Ye, W., Fan, X., 2023. How Well Does Weather Research and Forecasting (WRF) Model Simulate Storm Rashmi (2008) Itself and Its Associated Extreme Precipitation over the Tibetan Plateau at the Same Time? Atmosphere. https://doi.org/10.3390/atmos14101479

An, S., Oh, T.-J., Sohn, E., Kim, D., 2025. Deep learning for precipitation nowcasting: A survey from the perspective of time series forecasting. Expert Systems with Applications 268, 126301.

Brooks, H.E., Dotzek, N., 2008. The spatial distribution of severe convective storms and an analysis of their secular changes, in: Diaz, H.F., Murnane, R.J. (Eds.), Climate Extremes and Society. Cambridge University Press, pp. 35–53. https://doi.org /10.1017/CBO9780511535840.006

Chen, L., Cao, Y., Ma, L., Zhang, J., 2020. A deep learning-based methodology for precipitation nowcasting with radar. Earth and Space Science 7, e2019EA000812.

Cimini, D., Nelson, M., Güldner, J., Ware, R., 2015. Forecast indices from a ground-based microwave radiometer for operational meteorology. Atmos. Meas. Tech. 8, 315–333. https://doi.org/10.5194 /amt-8-315-2015

Çınaroğlu, E., Unutulmaz, O., 2019. A Data Mining Application of Local Weather Forecast for Kayseri Erkilet Airport. Politeknik Dergisi 22, 103–113. https://doi.org/10.2339/politeknik.391801

Devi, S.R., Arulmozhivarman, P., Venkatesh, C., Agarwal, P., 2016. Performance comparison of artificial neural network models for daily rainfall prediction. International Journal of Automation and Computing 13, 417–427. https://doi.org/ 10.1007/s11633-016-0986-2

Emanuel, K.A., 1994. Atmospheric Convection. Oxford University Press.

Ferrett, S., Frame, T.H., Methven, J., Holloway, C.E., Webster, S., Stein, T.H., Cafaro, C., 2021. Evaluating convection-permitting ensemble forecasts of precipitation over Southeast Asia. Weather and Forecasting 36, 1199–1217.

Fonseca, R.M., Zhang, T., Yong, K.T., 2015. Improved simulation of precipitation in the tropics using a modified BMJ scheme in the WRF model. Geo-scientific Model Development. https://doi.org /10.5194/gmd-8-2915-2015

Guijo-Rubio, D., Casanova-Mateo, C., Sanz-Justo, J., Gutiérrez, P.A., Cornejo-Bueno, S., Hervás, C.,

Salcedo-Sanz, S., 2020. Ordinal regression algorithms for the analysis of convective situations over Madrid-Barajas airport. Atmospheric Research 236, 104798. https://doi.org/10.1016/j.atmosres.2019.104798

Haidar, A., Verma, B., 2018. A novel approach for optimizing climate features and network parameters in rainfall forecasting. Soft Comput 22, 8119–8130. https://doi.org/10.1007/s00500-017-2756-7

Hall, T., Brooks, H.E., Doswell, C.A., 1999. Precipitation Forecasting Using a Neural Network. Wea. Forecasting 14, 338–345. https://doi.org/10.1175/15200434(1999)014%253C0338:PFUANN%253E2.0.CO;2

Hidayat, R., Taufik, M., 2025. Bias Correction of CMIP6 Models for Assessment of Wet and Dry Conditions Over Sumatra. Agromet 39, 33–39. https://doi.org/10.29244/j.agromet.39.1.33-39

Holloway, C.E., Neelin, J.D., 2009. Moisture Vertical Structure, Column Water Vapor, and Tropical Deep Convection. Journal of the Atmospheric Sciences 66, 1665–1683. https://doi.org/10.1175/2008JAS2806.1

Huntrieser, H., Schiesser, H.H., Schmid, W., Waldvogel, A., 1997. Comparison of Traditional and Newly Developed Thunderstorm Indices for Switzerland. Wea. Forecasting 12, 108–125. https://doi.org/10.1175/15200434(1997)012%253C0108:COTAND%253E2.0.CO;2

Johnson, R.H., Ciesielski, P.E., 2000. Rainfall and Radiative Heating Rates from TOGA COARE Atmospheric Budgets. J. Atmos. Sci. 57, 1497–1514. https://doi.org/10.1175/1520-0469(2000)057%253C1497:RARHRF%253E2.0.CO;2

Kandasamy, O., N., M., E., S., R., R., Kannan, B., C., P., 2025. Rainfall prediction using artificial neural networks and machine learning algorithms over the Coimbatore region. Journal of Water and Climate Change 16, 946–959. https://doi.org/10.2166/wcc.2025.576

Kim, S., Hong, S., Joh, M., Song, S., 2017. DeepRain: ConvLSTM Network for Precipitation Prediction using Multichannel Radar Data. https://doi.org/10.48550/arXiv.1711.02316

Manzato, A., 2005. The use of sounding-derived indices for a neural network short-term thunderstorm forecast. Weather and forecasting 20, 896–917.

Petersen, W.A., Rutledge, S.A., 2001. Regional Variability in Tropical Convection: Observations from TRMM. J. Climate 14, 3566–3586. https://doi.org/10.1175/15200442(2001)014%253C3566:RVITCO%253E2.0.CO;2

Putra, R.M., Fibriantika, E., Herawati, H., Kusumayanti, Y., Afriani, E., Hidayanti, A., Wiujiana, A., Swastiko, W.A., Andrianto, D., 2021. Cumulonimbus cloud prediction based on machine learning approach using radiosonde data in Surabaya, Indonesia. IOP Conf. Ser.: Earth Environ. Sci. 724, 012047. https://doi.org/10.1088/1755-1315/724/1/012047

Sanchez, J.L., Garcia, E., Marcos, J.L., Dessens, J., 2024. Formation of big and giant drops inside mediterranean convective cells.

Seeley, J.T., Romps, D.M., 2015. Why does tropical convective available potential energy (CAPE) increase with warming? Geophysical Research Letters 42. https://doi.org/10.1002/2015GL066199

Sham, F.A.F., El-Shafie, A., Jaafar, W.Z.B.W., Adarsh, S., Sherif, M., Ahmed, A.N., 2025. Improving rainfall forecasting using deep learning data fusing model approach for observed and climate change data. Sci Rep 15, 27872. https://doi.org/10.1038/s41598-025-13567-2

Shankar, A., Sahana, B.C., Kumar, A., Anand, A., 2025. Evaluating wind persistence and nowcasting of wind in aviation services: Harnessing deep learning for enhanced terminal aerodrome forecast. Theoretical and Applied Climatology 156, 467.

Showalter, A.K., 1953. A Stability Index for Thunderstorm Forecasting. Bulletin of the American Meteorological Society 34, 250–252. https://doi.org/10.1175/1520-0477-34.6.250

Siswanto, S., Nuryanto, D.E., Ferdiansyah, M.R., Prastiwi, A.D., Dewi, O.C., Gamal, A., Dimyati, M., 2023. Spatio-temporal characteristics of urban heat Island of Jakarta metropolitan. Remote Sensing Applications: Society and Environment 32, 101062. https://doi.org/10.1016/j.rsase.2023.101062

Taufik, M., Haikal, M., 2024. A Preliminary Study on the Parameter Configuration of Weather Research Forecasting in Tropical Peatland, Central Kalimantan. Agromet 38, 49–57. https://doi.org/10.29244/j.agromet.38.1.49-57

Thupeng, W.M., Sivasamy, R., Daman, O.A., 2023. RAINFALL SERIES FORECASTING MODELS BY ARIMA, NN, AND HOMM METHODS. ADAS 91, 83–98. https://doi.org/10.17654/0972361724007

Trenberth, K.E., Dai, A., Rasmussen, R.M., Parsons, D.B., 2003. The Changing Character of Precipitation. Bulletin of the American Meteorological Society 84, 1205–1218. https://doi.org/10.1175/BAMS-84-9-1205

Umakanth, N., Satyanarayana, G.Ch., Simon, B., Kumar, P., Rao, M.C., 2020. Impact of convection and stability parameters on lightning activity over Andhra Pradesh, India. Acta Geophys. 68, 1845–

1866. https://doi.org/10.1007/s11600-020-00479-0

Williams, E., Renno, N., 1993. An Analysis of the Conditional Instability of the Tropical Atmosphere. Mon. Wea. Rev. 121, 21–36. https://doi.org/10.1175/1520-0493(1993)121%253C0021:AAOTCI%253E2.0.CO;2

WMO, 2015. Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8).

Wu, P.-Y., Takemi, T., 2023. Impacts of mountain topography and background flow conditions on the predictability of thermally induced thunderstorms and the associated error growth. Journal of the Atmospheric Sciences 80, 1177–1199.

Yoo, C., Feldstein, S., Lee, S., 2011. The impact of the Madden-Julian Oscillation trend on the Arctic amplification of surface air temperature during the 1979-2008 boreal winter: IMPACT OF MJO ON POLAR AMPLIFICATION. Geophys. Res. Lett. 38, n/a-n/a. https://doi.org/10.1029/2011GL049881

Zhu, L., Meng, Z., Zhang, F., Markowski, P.M., 2017. The influence of sea- and land-breeze circulations on the diurnal variability in precipitation over a tropical island. Atmos. Chem. Phys. 17, 13213–13232. https://doi.org/10.5194/acp-17-13213-2017

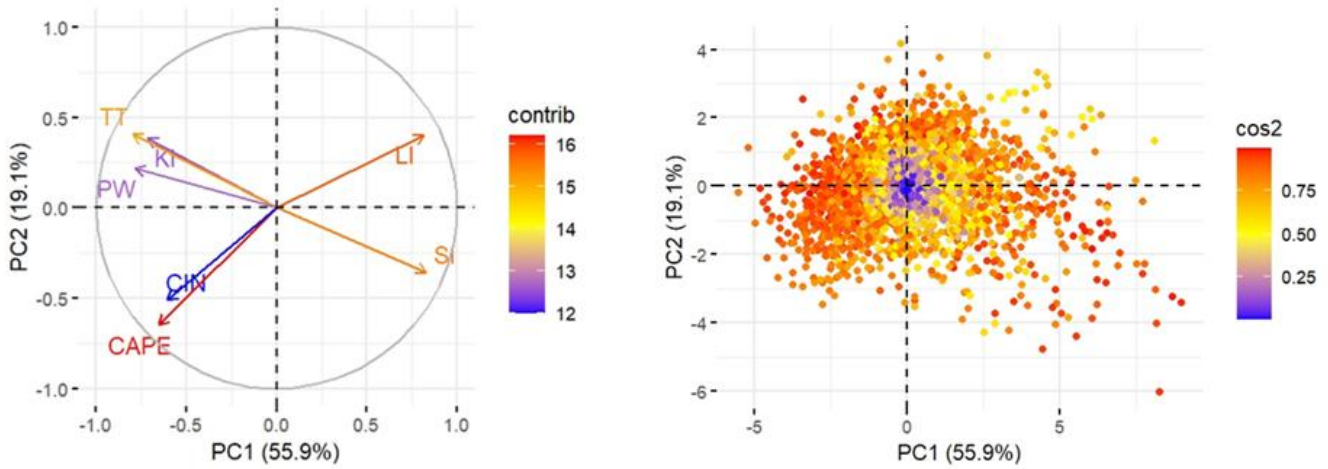
ANNEX

Table A1 Feature Selection Analysis

Elimination Order	Variable	Statistical Reason	Meteorological Reason
1	LI	Weak negative correlation (ρ = -0.167), significant in principal components but adequately represented by stronger SI.	SI sufficiently represents stability
2	TT	Moderate correlation but lower than KI, with a similar role to LI	Stability is already represented by SI/KI
3	KI	Fairly strong correlation (ρ = 0.316) but lower role in principal components compared to SI and tend to overlap	Stability remains well-covered by SI.
4	CIN	Moderate correlation (ρ = 0.233) but contributes negatively to the second principal component; its energetic effect can be covered by CAPE	Energetic component is still represented by CAPE
Retained	SI	Strong role in principal components, significant correlation, and main stability variable.	Determines whether the air tends to rise (instability)
Retained	CAPE	Strong contribution to the second component, weak correlation, but physically meaningful.	Represents convective energy potential
Retained	PW	Highest correlation (ρ = 0.394), strong positive role in principal components, and not redundant	Represents the atmospheric water vapor content

Table A2 ANN Model evaluation results

MATRIKS	Model 1 f(SI, LI, KI, TT, CAPE, CIN, PW)	Model 2 f(SI, KI, TT, CAPE, CIN, PW)	Model 3 f(SI, KI, CAPE, CIN, PW)	Model 4 f(SI, CAPE, CIN, PW)	Model 5 f(SI,CAPE,PW)
	Uji-Total	Uji-Total	Uji-Total	Uji-Total	Uji-Total
KEJADIAN HUJAN:					
Accuracy	0.771	0.769	0.782	0.780	0.738
Precision	0.633	0.636	0.665	0.658	0.589
Recall	0.754	0.727	0.705	0.716	0.727
F1-Score	0.688	0.679	0.684	0.686	0.650
AUC	0.814	0.846	0.824	0.789	0.786
Brier Score	0.210	0.203	0.201	0.206	0.247
INTENSITAS HUJAN:					
Accuracy	0.596	0.555	0.619	0.618	0.575
RMSE	0.707	0.776	0.711	0.698	0.798
MSE	0.500	0.603	0.505	0.487	0.637
F1-Score RA-	0.458	0.494	0.485	0.537	0.497
F1-Score RA	0.699	0.647	0.723	0.704	0.680
F1-Score RA+	0.143	0.083	0.222	0.125	0.080
F1-Score macro	0.433	0.408	0.477	0.455	0.419
F1-Score weighted	0.574	0.563	0.606	0.619	0.578
brier_score RA-	0.270	0.269	0.251	0.245	0.280
brier_score RA	0.231	0.243	0.225	0.231	0.221
brier_score RA+	0.064	0.076	0.069	0.068	0.075
brier_score macro	0.188	0.196	0.182	0.181	0.192
brier_score weighted	0.247	0.253	0.234	0.235	0.251



(a) (b)

Figure A1 (a) Distribution of individual samples in the space of the first two principal components of PCA (b) Variable contribution plot in the first two principal components of PCA