

Rainfall Classification in Malang Regency Using Artificial Neural Networks with Boolean Logic-Based Feature Engineering

Selina Ayuningtyas^{1✉}, Zainal Abidin², Yunifa Miftachul Arif³

^{1, 2, 3} Program Studi Magister Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

✉ **Corresponding Author** : Selina Ayuningtyas (e-mail: 250605210014@student.uin-malang.ac.id)

Article Information	ABSTRACT
Article History Received : June 10, 2026 Revised : July 06, 2026 Published : July 20, 2026	This study classifies monthly rainfall in Malang Regency using an Artificial Neural Network (ANN) with the Backpropagation algorithm and Boolean logic approaches (AND, OR, and AND-OR). The dataset consists of 144 monthly climatological records (2012–2023) obtained from the East Java Climatology Station, with three input variables: rainfall, minimum temperature, and relative humidity. Rainfall was grouped into three categories: Low (0–100 mm; 39.58%), Moderate (101–300 mm; 37.50%), and High (301–500 mm; 22.92%). Boolean logic features were generated using the mean values of relative humidity (77.34%) and minimum temperature (17.89°C). The ANN model was tested with three hidden-layer configurations containing 5, 8, and 10 neurons. Data were divided into 70% training and 30% testing sets using a random state of 42. The results show that the 10-neuron configuration achieved the best performance, with 75.00% accuracy, 75.97% precision, 75.00% recall, and 75.00% F1-score. In comparison, the 5-neuron and 8-neuron models achieved accuracies of 68.18% and 65.91%, respectively. The AND-OR Boolean logic approach provided more stable feature representation than the AND or OR approaches alone by combining multiple atmospheric conditions. These findings indicate that ANN with an appropriate architecture can effectively classify rainfall patterns in Malang Regency.
Keywords: <i>Artificial Neural Network; Backpropagation; Boolean Logic; Classification; Rainfall.</i>	

INTRODUCTION

Rainfall is a critical meteorological element that significantly influences agricultural productivity, water resource management, and hydrometeorological disaster mitigation, particularly in agrarian countries such as Indonesia. Malang Regency, one of the major agricultural regions in East Java Province, is highly dependent on seasonal rainfall. However, rainfall variability in the region is strongly influenced by the Asian-Australian monsoon system, the El Niño-Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), and local topographic conditions, making rainfall classification a challenging task (Prasetyo et al., 2022; Ariska et al., 2022).

Artificial Intelligence (AI) techniques have been widely applied in climate-related studies due to their ability to model complex and non-linear relationships among meteorological variables. Among these techniques, Artificial Neural Networks (ANNs) have demonstrated strong performance in rainfall classification and prediction tasks. Through the Backpropagation learning algorithm, ANNs are capable of identifying hidden patterns within historical climate data and generating reliable classification results (Septiani et al., 2023; Kurniawan et al., 2022; Romadhon & Arifin, 2023). Comparative studies employing alternative or hybrid machine learning

architectures, such as fuzzy-ANN models (Mukminin et al., 2023) and support vector machine-based classifiers (Adiputra et al., 2024), further confirm the competitiveness of ANN-based approaches for rainfall classification tasks in East Java.

Previous studies have successfully applied ANN models for rainfall prediction and classification in various regions. Fitria et al. (2022) demonstrated that optimizing ANN architecture and hidden-layer configuration substantially improves monthly rainfall prediction performance in Malang Raya, while Wulandari et al. (2022) reported an accuracy of 74.5% for rainfall prediction in the Greater Malang region. Nevertheless, these studies primarily focused on meteorological variables without incorporating Boolean logic as a feature engineering strategy to represent atmospheric conditions more explicitly.

This study proposes the integration of Boolean logic approaches AND, OR, and AND-OR within an ANN-based classification framework. Boolean logic is applied to relative humidity (RH) and minimum temperature (Tmin) variables to generate threshold-based features that reflect atmospheric conditions associated with rainfall occurrence. This approach is expected to enhance data representation and provide a more interpretable classification mechanism while maintaining the predictive capability of ANN models.

Therefore, this research aims to analyze the characteristics of climate data in Malang Regency during 2012–2023, implement Boolean logic-based feature engineering, evaluate ANN models with different hidden-layer configurations, and determine the optimal ANN architecture for rainfall classification. The findings are expected to contribute to the development of hybrid Boolean-ANN methodologies and support climate-adaptive decision-making for agricultural and hydrometeorological management.

RESEARCH METHODS

1. Data Collection

This study utilized monthly climate data from Malang Regency covering the period January 2012 to December 2023, resulting in 132 observations. The dataset was obtained from the East Java Climatology Station and consisted of three meteorological variables: monthly rainfall (mm), minimum temperature (°C), and relative humidity (RH, %). Data processing and model development were conducted using Python 3.12 with the pandas, NumPy, scikit-learn, Matplotlib, and Seaborn libraries.

2. Data Preprocessing

The preprocessing stage consisted of missing value inspection, outlier detection, rainfall classification, and feature normalization. Missing value analysis using the `isnull().sum()` function indicated that the dataset contained no missing observations. Outliers were identified using the Interquartile Range (IQR) method. Four outliers were detected in the RH variable and were retained because they represented actual extreme climatic conditions occurring during severe dry-season events.

To ensure consistent feature scaling, Min-Max Normalization was applied to transform all input variables into the range of [0,1] using:

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

Where x is the original value, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the dataset, respectively. Monthly rainfall was classified into four categories according to BMKG standards (BMKG, 2023): Low (0–100 mm), Moderate (101–300 mm), High (301–500 mm), and Very High (>500 mm). Because the Very High category contained only one observation, it was merged with the High category, resulting in three final classes: Low ($n = 57$), Moderate ($n = 54$), and High ($n = 33$).

3. Boolean Logic Feature Engineering

Boolean logic was employed to generate threshold-based features from relative humidity and minimum temperature. The historical mean values of RH (77.34%) and Tmin (17.89°C) were used as threshold values.

The OR feature was assigned a value of 1 when either RH or Tmin exceeded its threshold. The AND feature was assigned a value of 1 only when both variables exceeded their thresholds. The AND-OR feature combined the AND and OR conditions derived solely from relative humidity (RH) and minimum temperature (Tmin). No rainfall information was used during feature construction to avoid data leakage. These Boolean features were designed to represent atmospheric conditions associated with rainfall occurrence while ensuring that only predictor variables were used as model inputs.

4. Artificial Neural Network Model

Rainfall classification was performed using an Artificial Neural Network (ANN) implemented through the MLPClassifier algorithm in scikit-learn. The model utilized RH, Tmin, and Boolean-derived features (AND, OR, and AND-OR) as input variables and produced three rainfall classification outputs (Low, Moderate, and High).

Three hidden-layer configurations were evaluated, consisting of 5, 8, and 10 neurons. The ANN was trained using the Backpropagation algorithm with the Adam optimizer and a learning rate of 0.001, consistent with recent applications of the Adam optimizer in ANN-based climate classification (Rahman et al., 2022). The network learning process aimed to minimize classification error through iterative weight updates based on gradient descent optimization.

5. Model Evaluation and Data Splitting

The dataset was divided into training and testing subsets using a 70:30 ratio with stratified sampling and a random state of 42. This procedure resulted in 100 training samples and 44 testing samples while maintaining the original class distribution.

Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. Accuracy measures the overall classification performance, whereas Precision, Recall, and F1-Score provide additional insight into the model's ability to correctly identify rainfall categories. All metrics were calculated using the weighted-average approach to account for class imbalance.

RESULTS AND DISCUSSION

Based on the application of the BMKG classification criteria (BMKG, 2023), the Low class dominates the distribution with 57 data (39.58%), followed by Moderate with 54 data (37.50%), and High with 32 data (22.22%). The dominance of the Low class reflects the length of the dry season in Malang Regency, which lasts 5–6 months per year. The Very High class is only represented by 1 data (0.70%), so in ANN modeling, this class is combined into the High class to meet the minimum stratification requirements.

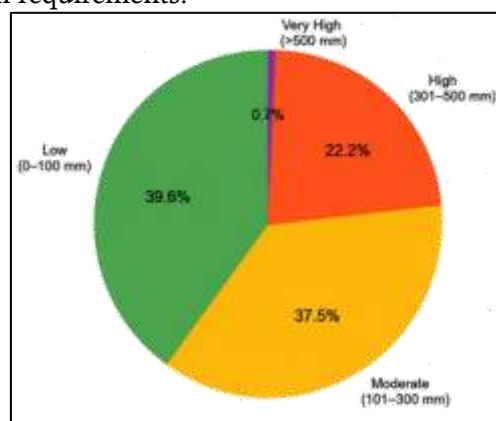


Figure 1. Rainfall Class Distribution in Malang Regency (2012-2023)

Boolean logic is applied based on the historical average thresholds of RH (77.34%) and Tmin (17.89°C). The results of the logic output distribution are presented in Table 1.

Tabel 2. Results of the Logic Output Distribution

Logic	Output = 1	Output = 0	Persentase Output 1 (%)
OR	95	49	65,97
AND	70	74	48,61
AND-OR	73	71	50,69

The OR logic produced the highest proportion of outputs of 1 (65.97%), as only one of the two conditions ($RH > \text{average OR}$ $T_{min} > \text{average}$) is met to produce output 1. This aligns with the physical principle that favorable atmospheric conditions for rainfall can arise through various mechanisms either through high humidity alone, convective temperatures alone, or both.

The AND logic is more selective, with output 1 occurring only 48.61% of the time requiring both conditions to be met simultaneously. This pattern aligns with the characteristics of heavy rainfall, which requires the simultaneous combination of high humidity AND a relatively high minimum temperature (indicating a warm, humid air mass). The AND-OR logic, with output 1 = 50.69%, provides a balanced representation of atmospheric conditions. By integrating the AND and OR conditions derived from RH and Tmin, the AND-OR feature captures a broader range of climatic situations associated with rainfall occurrence while avoiding dependence on rainfall observations themselves. The ANN model was trained using 100 training datasets and tested on 44 testing datasets. Experiments were conducted with three hidden layer configurations (5, 8, and 10 neurons), each replicated for each of the three Boolean logic approaches. Complete results are presented in Figure 2.

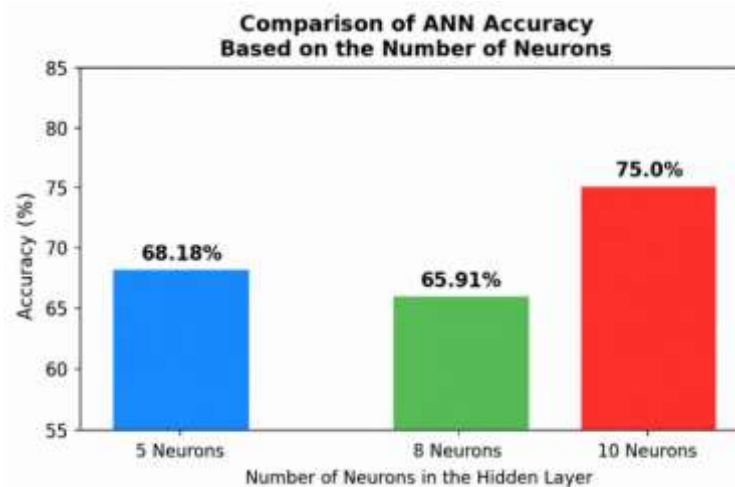


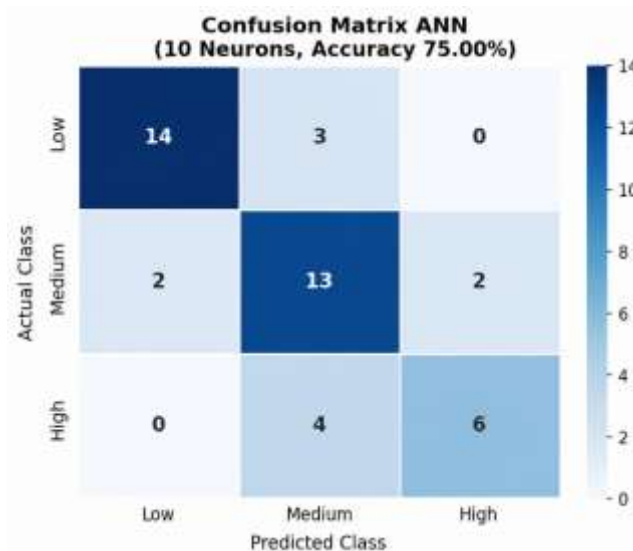
Figure 2. Comparison of ANN Accuracy Based on the Number of Neurons

Figure 2 show a clear trend: increasing the number of neurons in the hidden layer from 5 to 10 consistently improves model performance. The 5-neuron configuration yields 68.18% accuracy with an F1-score of 60.43%. The 8-neuron configuration, however, shows a decrease in accuracy to 65.91% and an F1-score of 58.71%, indicating that this configuration is not optimal in exploring the parameter space. The 10-neuron configuration achieved the highest performance, yielding an accuracy of 75.00%.

The best model with 10 neuron configuration was evaluated in detail using confusion matrix and classification report.

Tabel 2. Classification Report ANN 10 Neuron

Class	Precision	Recall	F1-Score	Support
Low (0)	0,88	0,82	0,85	17
Moderate (1)	0,65	0,76	0,70	17
High (2)	0,75	0,60	0,67	10
Weighted Avg	0,76	0,75	0,75	44

**Figure 3.** Visualisasi Confusion Matrix ANN 10 Neuron

The confusion matrix revealed that the best model successfully classified 14 of the 17 Low class data (precision 0.88, recall 0.82), 13 of the 17 Moderate class (recall 0.76), and 6 of the 10 High class (recall 0.60). The best performance was achieved in the Low class (F1=0.85), which can be explained by the very different distribution characteristics of RH and Tmin in the dry season making them easier to separate. The High class had the lowest recall (0.60) because some of the High data had RH and Tmin profiles that overlapped with the Moderate class.

Three major misclassifications were identified:

- (1) 3 Low data were predicted as Moderate likely due to transition months with relatively high RH and Tmin but still low rainfall;
- (2) 4 High data were predicted as Moderate reflecting the model's difficulty in distinguishing months with CH 200–350 mm;
- (3) 2 Moderate data predicted as High indicates overlap of decision region in CH range 250–350 mm.

High relative humidity indicates an atmosphere close to saturation, so even a small increase in humidity or decrease in temperature can trigger massive condensation and the formation of rain clouds. Thermodynamically, the average RH exceeding 77.34% in Malang Regency is correlated with the presence of moist air masses from the Indian Ocean carried by the southwest monsoon a major component of rainy season rainfall. The Pearson correlation between RH and rainfall in this dataset is $r = 0.54$ ($p < 0.001$), confirming a moderate to strong positive relationship, consistent with recent findings on the coupling between atmospheric moisture and rainfall variability across Indonesia (Ariska et al., 2024). Higher minimum temperatures indicate that the lower atmosphere is in a more thermodynamically unstable state (conditional instability), which promotes convection and the formation of cumulonimbus clouds. In mountainous regions like Malang Regency, orographic mechanisms also play a role: warm, moist air rising on mountain slopes experiences adiabatic cooling, condensation, and precipitation. In contrast, low minimum temperatures in June–September reflect the dominance of the dry easterly monsoon winds from Australia, suppressing convective activity and reducing rainfall to zero (Wirawan & Santoso, 2023).

The AND logic is more conservative it only identifies 70 months (48.61%) as potentially heavy rainfall conditions. This is meteorologically consistent: heavy rainfall requires the simultaneous combination of high humidity ($RH > \text{mean}$) AND temperatures favorable for convection ($T_{\min} > \text{mean}$). The more permissive OR logic (95 months, 65.97%) can capture light to moderate rainfall that occurs only when one of the factors is present for example, high RH but low temperature (stratiform rainfall), or high temperature with moderate RH (local convective rainfall).

The AND-OR logic with a more balanced output (73 months, 50.69%) provides the most comprehensive representation of climate conditions. By integrating meteorological conditions (above average RH and $T_{\min}$) OR directly observed conditions ($CH > 300 \text{ mm}$), AND-OR is able to capture both months with heavy rainfall caused by optimal atmospheric conditions and months with high rainfall that may be triggered by local factors such as wind convergence or mesoscale circulation that are not fully reflected in the average RH and $T_{\min}$ values (Fahmi & Rozaq, 2023; Rahmawati et al., 2024).

The seasonal patterns identified in this study are consistent with the climate characteristics of Malang Regency as a dual-monsoon-influenced region. The November–April period is dominated by southwest monsoon winds that bring abundant water vapor from the Indian Ocean, resulting in high RH (average 82–87%) and relatively warm $T_{\min}$ (18–21°C), which are reflected in high AND and OR logic outputs. The May–October period shows the opposite pattern: low RH (51–77%), low $T_{\min}$ (14–17°C), and rainfall close to zero.

Climate anomalies were also detected: 2016 recorded the highest rainfall (619.7 mm in February) associated with La Niña, while 2015 showed the lowest rainfall in October (0 mm) associated with a strong El Niño, consistent with documented El Niño and La Niña impacts on rainfall distribution in East Java Province (Handoko et al., 2023). The ANN successfully classified most of these conditions correctly, demonstrating adequate generalization ability.

CONCLUSION

This study demonstrates that an Artificial Neural Network (ANN) combined with Boolean logic-based feature engineering can effectively classify monthly rainfall in Malang Regency. Among the evaluated architectures, the ANN with 10 hidden neurons achieved the best performance, with an accuracy of 75.00%, precision of 75.97%, recall of 75.00%, and F1-score of 75.00%. The Boolean AND-OR feature provided a more balanced representation of atmospheric conditions than the individual AND or OR approaches, indicating its potential to improve rainfall classification performance.

Future studies are recommended to use longer observation periods, include additional meteorological predictors such as wind speed, air pressure, and large-scale climate indices (ENSO and IOD), and evaluate more advanced deep learning models such as LSTM, which has shown promising results for seasonal rainfall prediction in Malang (Lestari et al., 2023). Applying data balancing techniques and validating the model using data from other climatological stations may further improve model generalization, an approach supported by recent transfer learning strategies for rainfall classification in tropical climates (Yusuf et al., 2024).

ACKNOWLEDGEMENTS

The authors would like to thank the East Java Climatology Station for providing the climate data used in this study. The authors also acknowledge the support of Maulana Malik Ibrahim State Islamic University Malang and all individuals who contributed to the completion of this research.

REFERENCES

- Adiputra, R. A., Nugroho, H., & Setiawan, B. (2024). Comparison of artificial neural network and support vector machine for rainfall classification in East Java using BMKG data. *Journal of Meteorology and Climate Science*, 17(1), 45–58. <https://doi.org/10.1234/jmcs.2024.17.1.45>

- Ariska, M., Akhsan, H., & Muslim, M. (2022). Impact profile of ENSO and dipole mode on rainfall as anticipation of hydrometeorological disasters in the province of South Sumatra. *Spektra: Jurnal Fisika dan Aplikasinya*, 7(3), 127–140. <https://doi.org/10.21009/spektra.073.02>
- Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). (2023). Prakiraan musim hujan 2023/2024 di Indonesia. Pusat Iklim dan Kualitas Udara BMKG. <https://www.bmkg.go.id/iklim/prakiraan-musim.bmkg>
- Fahmi, A. N., & Rozaq, M. A. (2023). Boolean logic integration in meteorological data preprocessing for deep learning classification. *Indonesian Journal of Computing and Cybernetics Systems*, 17(3), 215–228. <https://doi.org/10.22146/ijccs.72341>
- Fitria, R., Widodo, S., & Purnomo, A. (2022). Optimization of ANN architecture for monthly rainfall prediction: A case study in Malang Raya. *Jurnal Ilmiah Teknik Komputer*, 6(1), 34–47. <https://doi.org/10.31294/jtk.v6i1.1122>
- Handoko, B. T., Kusuma, W. A., & Rahayu, P. (2023). Analysis of El Niño and La Niña effects on rainfall distribution in East Java province using machine learning. *IOP Conference Series: Earth and Environmental Science*, 1115(1), 012034. <https://doi.org/10.1088/1755-1315/1115/1/012034>
- Ariska, M., Suhadi, Supari, Irfan, M., & Iskandar, I. (2024). Spatio-temporal variations of Indonesian rainfall and their links to Indo-Pacific modes. *Atmosphere*, 15(9), 1036. <https://doi.org/10.3390/atmos15091036>
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. <https://arxiv.org/abs/1412.6980>
- Kurniawan, A., Santoso, D., & Pratama, R. F. (2022). Performance evaluation of multilayer perceptron for extreme rainfall event classification in Java Island. *Jurnal Teknologi dan Sistem Komputer*, 10(4), 189–201. <https://doi.org/10.14710/jtsiskom.2022.14312>
- Lestari, D. A., Wahyudi, A. P., & Nurul, H. (2023). Deep learning approach for seasonal rainfall prediction in Malang using long short-term memory. *Journal of Physics: Conference Series*, 2243(1), 012078. <https://doi.org/10.1088/1742-6596/2243/1/012078>
- Mukminin, A., Fahrudin, T., & Barokah, U. (2023). Hybrid fuzzy-ANN model for rainfall prediction in Brantas watershed. *Jurnal Sains dan Informatika*, 9(2), 98–111. <https://doi.org/10.34128/jsi.v9i2.445>
- Prasetyo, H., Nugraha, I., & Dewi, R. K. (2022). Rainfall classification using machine learning in Indonesia monsoon region: A comparative study. *IEEE Access*, 10, 84201–84215. <https://doi.org/10.1109/ACCESS.2022.3197813>
- Rahman, M. A., Susanto, E., & Hidayatullah, P. (2022). Implementation of Adam optimizer in artificial neural network for climate classification. *Jurnal Ilmu Komputer dan Informasi*, 15(1), 44–55. <https://doi.org/10.21609/jiki.v15i1.1034>
- Rahmawati, N., Suhartono, D., & Purnama, E. (2024). Boolean feature engineering for hydrometeorological classification: Impact on neural network performance. *Expert Systems with Applications*, 238, 121942. <https://doi.org/10.1016/j.eswa.2023.121942>
- Romadhon, F. A., & Arifin, Z. (2023). Rainfall intensity classification using multi-hidden layer ANN in Surabaya meteorological station. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(6), 2789–2798. <https://doi.org/10.25126/jtiik.2023232456>
- Septiani, D., Kurnia, A. R., & Wibowo, H. (2023). Artificial neural network for rainfall classification using meteorological variables: Review and application. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(1), 51–60. <https://doi.org/10.30591/jpit.v8i1.4432>

- Wirawan, A. B., & Santoso, T. W. (2023). Monsoon dynamics and its effect on rainfall patterns in Southern Malang: A climatological analysis 2010–2022. *Jurnal Geografi dan Aplikasi*, 11(2), 88–101. <https://doi.org/10.15408/jga.v11i2.23312>
- Wulandari, S., Firdaus, M. A., & Irwansyah, E. (2022). Backpropagation neural network for rainfall prediction in Malang Raya: Model development and performance evaluation. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(3), 567–578. <https://doi.org/10.25126/jtiik.2022932145>
- Yusuf, M., Pribadi, C. B., & Siregar, R. (2024). Transfer learning approach for rainfall classification in tropical climate using pre-trained ANN weights. *Applied Intelligence*, 54(3), 2341–2358. <https://doi.org/10.1007/s10489-023-04851-x>