



A Comparative Analysis of Machine Learning Models for Weather Type Classification

**Rapika Dahlan^{1*}, Mutiara Maharani², Faizal Risqi Andika³,
Aisyah Salsabila Purba⁴, Siti Khusnul Khotimah⁵, Juana Barus⁶**

^{1,2,3}Department of Information Systems, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

⁴Faculty of Islamic Studies and Arabic Language, Al Azhar University, Egypt

⁵Faculty of Economics and Administrative Sciences, Gaziantep University, Turkey

⁶Faculty of Divinity, Gaziantep University, Turkey

E-Mail: ¹12350324239@students.uin-suska.ac.id, ²12350321282@students.uin-suska.ac.id,
³12350311779@students.uin-suska.ac.id, ⁴aisyahsalsabilahpurba@gmail.com,
⁵sitikhunsulkhatimah5@gmail.com, ⁶juanabarus96@gmail.com

Received Dec 30th 2025; Revised Jul 31th 2026; Accepted Aug 17th 2026; Available Online Aug 31th 2026

Corresponding Author: Rapika Dahlan

Copyright © 2026 by Authors, Published by Institute of Research and Publication Indonesia (IRPI)

Abstract

Accurate weather type classification plays an important role in supporting decision-making across various sectors, including transportation, agriculture, logistics, and disaster management. Although numerous studies have applied machine learning to weather prediction, most compare only a limited number of algorithms or evaluate them under a single experimental setting. This study addresses this gap by providing a comprehensive benchmark of five widely used classification algorithms Decision Tree, Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN) using the publicly available Weather Type Classification dataset from Kaggle. These algorithms were selected because they represent interpretable, ensemble-based, kernel-based, and neural network approaches, enabling a comprehensive comparison of different machine learning paradigms. The dataset consists of structured meteorological attributes, including temperature, humidity, wind speed, precipitation, visibility, season, and location. Model performance was evaluated using three train-test split ratios (70/30, 80/20, and 90/10) and assessed through accuracy, precision, recall, confusion matrix, and ROC curve analyses. The experimental results show that ensemble-based and neural network models outperform single-tree approaches. Random Forest, MLP, and ANN demonstrate superior and more consistent performance across all data split ratios, with ROC-AUC values approaching 0.99. These findings establish a practical benchmark for selecting appropriate machine learning models for multi-class weather classification and demonstrate the effectiveness of ensemble and neural network approaches for developing reliable weather prediction systems using publicly available datasets.

Keywords: Artificial Neural Network, Decision Tree, Random Forest, Support Vector Machine, Weather Type

1. INTRODUCTION

Weather prediction plays a vital role in various real-world applications, including aviation, agriculture, logistics, and disaster management, making accurate classification of weather conditions increasingly important [1]. The dynamic and nonlinear nature of atmospheric interactions, such as temperature, humidity, wind speed, and cloud formation, often poses challenges to traditional statistical forecasting models that struggle to capture complex relationships within meteorological variables [2]. With the growing availability of open-source datasets, including those hosted on Kaggle, machine learning has emerged as a powerful tool capable of modeling intricate weather patterns using structured and high-dimensional data [3]. These advancements have driven numerous studies to evaluate the performance of machine learning models on publicly available datasets to enhance weather classification systems [2].

Despite the increasing use of open-source meteorological datasets, many publicly available weather datasets, such as the Kaggle Weather Type Classification dataset, exhibit unique challenges, including synthetic or simulated feature distributions, class imbalance, and nonlinear relationships across variables [3]. The characteristics introduce uncertainty in how well machine learning models generalize when trained on such structured but non-real-time data [1]. Moreover, while several exploratory Kaggle notebooks demonstrate

small-scale comparisons between 2-3 machine learning algorithms, there is still no comprehensive benchmark that systematically evaluates the performance of five major models Random Forest, Decision Tree, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN) on this dataset [4]. This gap highlights the need for a focused comparative study to identify the most effective model for multi-class weather-type prediction using publicly available data [2], [5].

To address these challenges, this research compares the performance of five widely used machine learning algorithms, Random Forest, Decision Tree, SVM, MLP, and ANN, on the Weather Type Classification dataset [6]. Random Forest provides strong robustness against overfitting by aggregating multiple decision trees to improve stability and predictive accuracy [7]. Decision Tree models offer interpretability and computational efficiency that make them suitable for structured datasets [3]. SVM is recognized for its superior capability in handling high-dimensional classification problems, particularly when data exhibits nonlinear and overlapping distributions [8]. Meanwhile, MLP and ANN architectures are capable of learning deep and complex relationships within synthetic and numerical weather attributes [9].

Previous studies demonstrate that ensemble-based models such as random Forest often achieve the highest accuracy in weather classification tasks due to their ability to model multi-feature interactions effectively [10]. Additional research indicates that SVM and MLP contribute enhanced stability when applied to multi-class atmospheric datasets, especially those exhibiting synthetic or structured statistical patterns [11]. ANN models have also shown superior performance in capturing nonlinear temporal or environmental relationships, making them promising candidates for weather classification [12]. However, the literature lacks a dedicated comparative analysis involving all five models on the Weather Type Classification dataset, despite its popularity in exploratory machine learning studies [13], [3].

The novelty of this study lies in its comprehensive evaluation framework for publicly available weather classification data. Unlike previous studies that primarily compare only two or three machine learning algorithms or evaluate models under a single train-test split configuration, this research systematically benchmarks five representative classification algorithms, Decision Tree, Random Forest, SVM, MLP, and ANN, under three different data split ratios (70/30, 80/20, and 90/10). Furthermore, the study integrates multiple evaluation metrics, including accuracy, precision, recall, confusion matrix, and ROC curve analysis, to provide a more comprehensive assessment of model performance and generalization capability. This comprehensive benchmarking provides a practical reference for selecting appropriate machine learning models for structured multi-class weather classification tasks.

Given this gap, the objective of the present study is to conduct a comprehensive benchmark of five machine learning algorithms Random Forest, Decision Tree, SVM, MLP, and ANN, using the Weather Type Classification dataset [14]. This research evaluates each model based on accuracy, precision, recall to identify which algorithm performs most effectively on structured, multi-class weather data of synthetic or semi-synthetic nature [15]. The results aim to provide insight into the strengths and limitations of each algorithm, establishing a benchmark for future studies that utilize publicly available datasets for weather classification tasks [16].

2. MATERIAL AND METHOD

This study aims to evaluate and compare the performance of five machine learning algorithms Random Forest, Decision Tree, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN) in classifying weather types using the publicly available Kaggle dataset [17]. The overall workflow of this research is illustrated in Figure 1, which outlines the stages from data collection to evaluation.

2.1. Data Collection

The dataset used in this research is publicly available and has been used by various researchers for weather-related machine learning tasks, sourced from Kaggle. The dataset, published by Nikhil7280, is titled “Weather Type Classification” and contains approximately 13,000 records with 11 meteorological attributes, including temperature, humidity, wind speed, air pressure, precipitation, UV index, visibility, cloud cover, and season information. These features represent diverse atmospheric conditions and provide a comprehensive basis for multi-class weather classification.

Several studies have employed Kaggle-based weather datasets to develop and evaluate machine learning models for weather prediction. For example, recent research has shown that algorithms such as Random Forest, SVM, and ANN perform effectively when applied to structured meteorological datasets due to their ability to capture non-linear relationships among weather parameters. The availability of labeled weather categories in this dataset also makes it well-suited for supervised classification tasks, enabling researchers to analyze and compare the performance of various algorithms in identifying weather types.

Given the structured nature of the dataset and its proven relevance in machine learning research, this study further explores its potential by evaluating five different classification algorithms Decision Tree, Random Forest, SVM, MLP, and ANN [18]. This evaluation aims to determine which algorithm provides the most accurate and robust performance in predicting weather types using publicly available meteorological data [19].

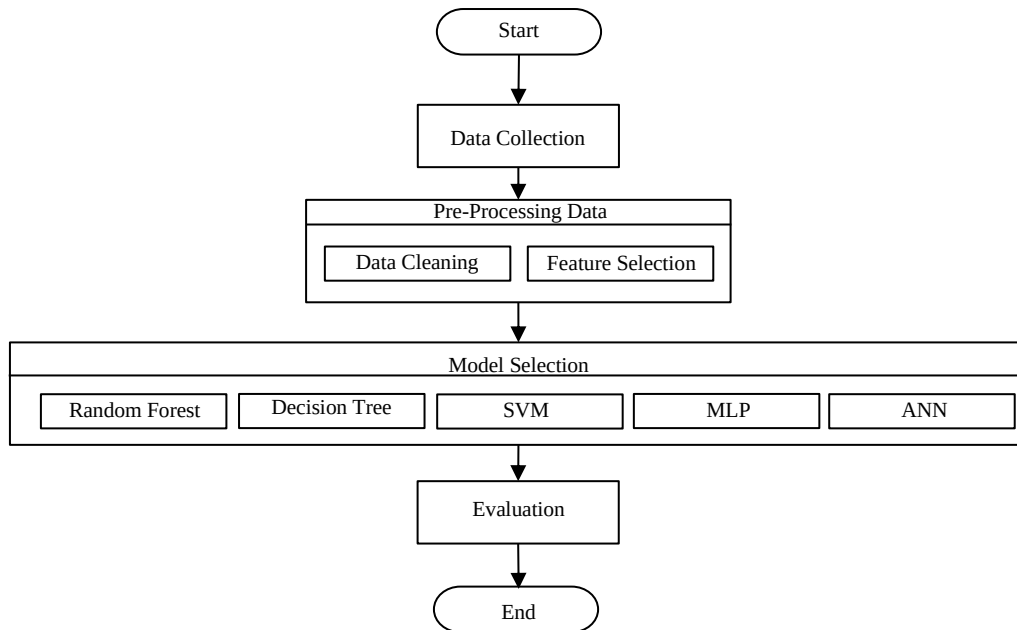


Figure 1. Research Methodology

2.2. Data Pre-processing

After the dataset is collected, the next step is preparing the data systematically before performing analysis using machine learning algorithms [20]. This stage is known as Data Preprocessing, which aims to clean, transform, and organize the dataset into a format suitable for model training. The preprocessing step involves several procedures, including handling missing values, removing irrelevant attributes, encoding categorical features, and normalizing numerical values.

In this study, categorical features such as Season and Cloud Cover were transformed using One-Hot Encoding, while numerical variables such as temperature, humidity, wind speed, and pressure were scaled using Min–Max Normalization. This step ensures that all features operate within a consistent range, preventing bias during model training. Proper preprocessing is essential to improve model performance and ensure that the classification algorithms can process the dataset effectively [10].

2.3. Classification

Classification is the task of assigning objects or entities into predefined categories based on their characteristics. In this context, the objective is to categorize weather conditions into specific weather types based on meteorological parameters [4]. Machine learning classification models work by identifying patterns within the dataset and grouping observations with similar attributes.

The process of classification also involves separating data with distinct characteristics so that the model can accurately predict each weather condition [5]. The models used in this study include Decision Tree, Random Forest, SVM, MLP, and ANN, each selected based on their ability to handle nonlinear and multivariate weather datasets [6].

2.4. Random Forest

Random Forest is an ensemble learning algorithm that constructs multiple Decision Trees and combines their outputs to improve prediction stability and accuracy [21]. Each tree is trained on different bootstrap samples, and a random subset of features is selected at each split. Prediction for a classification task is determined by majority voting as in Equation 1.

$$\hat{y} = \text{mode} \{h_1(x), h_2(x), \dots, h_T(x)\} \quad (1)$$

Where $h_t(x)$ is the prediction from the t -th tree and T is the total number of trees [22]. Random Forest reduces overfitting and performs well on high-dimensional datasets, making it suitable for weather classification.

2.5. Decision Tree

A Decision Tree is a supervised learning algorithm that performs classification by splitting data into branches based on feature values [23]. The model selects the best attribute for splitting by measuring impurity using Entropy or Gini Index. The entropy of a node is defined as Equation 2.

$$H(t) = -\sum_{k=1}^K p_k(t) \log_2 p_k(t) \quad (2)$$

Where $p_k(t)$ represents the proportion of the class k in node t [10]. The Information Gain for a feature A is computed as Equation 3.

$$IG(S, A) = H(S) - \sum_{v \in A} \frac{|S_v|}{|S|} H(S_v) \quad (3)$$

The Decision Tree selects the feature with the highest Information Gain for splitting [11]. This algorithm is widely used due to its interpretability and ability to handle both numerical and categorical data [12].

2.6. Support Vector Machine (SVM)

SVM is a powerful classification method that constructs an optimal hyperplane to separate data into distinct classes. The optimization problem for the SVM with soft margin is expressed in Equations 4-6.

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (4)$$

subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (5)$$

The decision function for prediction is:

$$f(x) = \text{sign}(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b) \quad (6)$$

where $K(x_i, x)$ is the kernel function [16]. SVM is effective for complex and nonlinear weather patterns [17].

2.7. Multilayer Perceptron (MLP)

MLP is a feed-forward neural network consisting of an input layer, hidden layers, and an output layer [24]. Each hidden layer transforms the input using Equation 7-8.

$$z_j^{(l)} = \sum_{i=1}^{n_{l-1}} w_{ji}^{(l)} a_i^{(l-1)} + b_j^{(l)} \quad (7)$$

$$a_j^{(l)} = \sigma(z_j^{(l)}) \quad (8)$$

For multi-class classification, the output layer uses Softmax using Equation 9.

$$\hat{y}_k = \frac{e^{z_k}}{\sum_{c=1}^K e^{z_c}} \quad (9)$$

The loss function used is Categorical Cross Entropy, which can be used in Equation 10.

$$\mathcal{L} = -\sum_{i=1}^N \sum_{k=1}^K y_{ik} \log(\hat{y}_{ik}) \quad (10)$$

MLP is effective for learning nonlinear relationships among weather features [18].

2.8. Artificial Neural Network (ANN)

ANN generalizes MLP by allowing deeper or more complex network structures [25]. The forward propagation process is given by Equation 11-12.

$$a^{(0)} = x \quad (11)$$

$$a^{(l)} = \sigma(W^{(l)} a^{(l-1)} + b^{(l)}) \quad (12)$$

Parameters are optimized by minimizing the loss in Equation 13.

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\theta) \quad (13)$$

ANN can model complex, dynamic weather patterns due to its ability to approximate nonlinear functions [26].

2.9. Confusion Matrix

A confusion matrix is a widely used performance evaluation method in machine learning, as it compares predicted labels with actual labels to measure classification correctness [27]. This method has been adopted extensively in modern classification research because it provides deeper diagnostic information beyond accuracy, especially for multi-class classification problems [28]. Researchers emphasize that the confusion matrix is essential in identifying true and false predictions, allowing a detailed assessment of classifier behavior [29].

The matrix consists of four components: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN), which together describe the full outcome distribution of a classifier [27]. These components are also the basis for important evaluation metrics such as accuracy, precision, recall, and F1-score, which are derived directly from the Confusion Matrix and widely used in machine learning evaluation frameworks [30].

Accuracy is calculated using Equation 14.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Precision measures how many of the predicted positive values are actually positive. It is computed as Equation 15.

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

Recall (or sensitivity) is the proportion of actual positive cases correctly predicted by the model, as given in Equation 16.

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

F1-Score combines precision and recall into a single metric and is defined as Equation 17.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (17)$$

		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	TP	FP
	NEGATIVE	FN	TN

Figure 2. Confusion Matrix Diagram [31]

Recent studies propose extensions such as probabilistic confusion matrices and hierarchical confusion matrices to enhance performance interpretation in complex classification systems [28], [31]. These advancements demonstrate how Confusion Matrix remains foundational while continuing to evolve in modern evaluation methodologies [30].

3. RESULTS AND DISCUSSION

This result indicates that the ensemble mechanism effectively reduces model variance and increases robustness against noisy meteorological features. Consequently, Random Forest is suitable for operational weather classification systems where prediction consistency is essential.

3.1. Initial Data

Before preprocessing and model development, an initial exploration of the dataset was conducted to understand its structure, feature distributions, and suitability for machine-learning classification. This analysis provides an overview of the meteorological variables used in subsequent experiments and ensures that the dataset adequately represents different weather categories.

Table 1. Weather Type Classification Dataset

Temperature	Humidity	Wind Speed	Precipitation (%)	...	Season	Visibility (km)	Location	Weather Type
14	73	9.5	82	...	Winter	3.5	inland	Rainy
39	96	8.5	71	...	Spring	10	inland	Cloudy
30	64	7	16	...	Spring	5.5	mountain	Sunny
38	83	1.5	82	...	Spring	1	coastal	Sunny
27	74	17	66	...	Winter	2.5	mountain	Rainy

The initial stage of this research involved examining the weather dataset from Kaggle, which contains approximately 13,000 records with 11 meteorological attributes, including temperature, humidity, wind speed, precipitation, visibility, season, and location. An initial inspection of the dataset was conducted to understand the basic characteristics of each feature before entering the preprocessing phase. Several samples of the initial dataset are shown in Table 1, presenting raw numerical and categorical information without modification.

Through this inspection, it was observed that the distribution of weather types Sunny, Cloudy, cloudy, rainy, and snowy, was relatively balanced, although several variables, such as precipitation and cloud cover, exhibited varying ranges across different seasons. This early analysis ensured that the dataset was suitable for further preprocessing and model training.

3.2. Preprocessing Data

The preprocessing stage was performed to transform the raw dataset into a form suitable for machine learning. Several steps were conducted, including removing irrelevant attributes, handling missing values, encoding categorical variables, and normalizing numerical features.

Table 2. Preprocessing Dataset

Temperature	Humidity	Wind Speed	Precipitation (%)	...	Weather Type Cloudy	Weather Type Rainy	Weather Type Snowy	Weather Type Sunny
14	73	9.5	82	...	FALSE	TRUE	FALSE	FALSE
39	96	8.5	71	...	TRUE	FALSE	FALSE	FALSE
30	64	7	16	...	FALSE	FALSE	FALSE	TRUE
38	83	1.5	82	...	FALSE	FALSE	FALSE	TRUE
27	74	17	66	...	FALSE	TRUE	FALSE	FALSE

Categorical features such as Season and Location were one-hot encoded, resulting in a more structured dataset, as shown in Table 2. Numerical attributes, including temperature, humidity, wind speed, and precipitation, were normalized to prevent value domination during training. This step ensured that each feature contributed fairly in the classification phase.

3.3. Performance Comparison of Machine Learning Models

3.3.1 Random Forest

The Random Forest algorithm was the first classification model assessed in this research. After training on the preprocessed dataset, the model's performance was evaluated using accuracy, precision, and recall. The results indicate that the Random Forest model achieved strong predictive performance in classifying weather types, demonstrating high accuracy, well-balanced precision and recall, and strong discriminative capability across classes.

Based on Figure 2, the Random Forest algorithm demonstrates strong performance in classifying weather types. The model achieved an accuracy of 87.42%, indicating that the majority of weather instances were correctly classified. The precision value reached 97.88%, which reflects a very high level of correctness in the positive predictions generated by the model. Meanwhile, the recall value was recorded at 87.42%, showing that the model was able to successfully identify most of the actual weather classes present in the dataset.

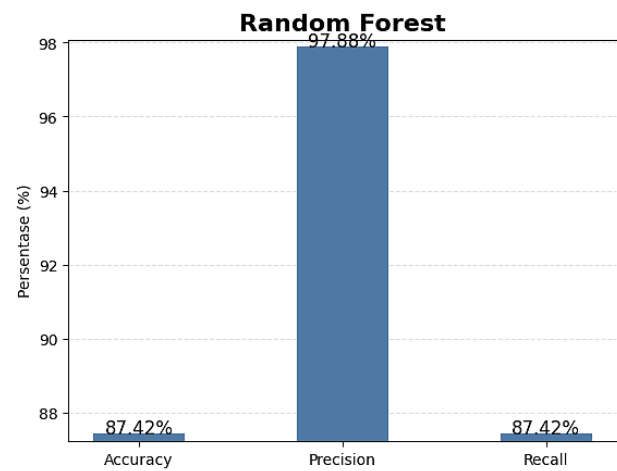


Figure 2. Random Forest Results

The high precision value suggests that the Random Forest model produces very few false positive predictions, making it highly reliable when assigning specific weather categories. Although the recall value is slightly lower than precision, it still indicates good capability in capturing relevant weather instances. Overall, the combination of high precision, balanced recall, and satisfactory accuracy confirms that the Random Forest algorithm is well-suited for multi-class weather type classification and is effective in modeling complex meteorological patterns.

3.3.2 Decision Tree

The Decision Tree classification model was evaluated in this study, and the results are presented in Figure 3. Based on Figure 3, the application of the Decision Tree algorithm in weather type classification achieved an accuracy of 87.05%, indicating a good level of classification performance. The precision value obtained was 90.05%, which is higher than the accuracy, while the recall value reached 87.05%, showing that the model is able to correctly identify most of the actual weather classes.

The classification results using the Decision Tree algorithm demonstrate that the model performs reasonably well in classifying weather data. The relatively high precision value indicates that the positive predictions generated by the model are mostly correct, resulting in a low rate of false positive predictions.

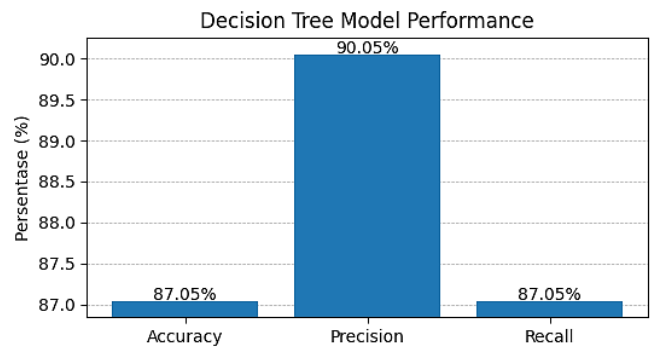


Figure 3. Decision Tree Results

However, the recall value, which is equal to the accuracy, suggests that the model still misses some relevant instances, particularly in weather categories with similar meteorological characteristics. This indicates that while the Decision Tree model is effective, its performance could be further improved through parameter tuning or the application of ensemble-based methods to enhance its generalization capability.

3.3.3 Support Vector Machine (SVM)

The Support Vector Machine algorithm was also implemented to classify weather types. Using a non-linear kernel, the SVM model achieved good performance, particularly in predicting weather types with clearer feature separation, such as Sunny and Snowy.

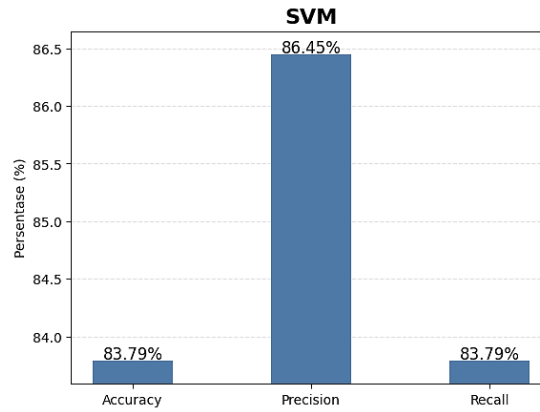


Figure 4. SVM Classification Results

Based on Figure 4, the application of the Support Vector Machine (SVM) algorithm in weather type classification achieved an accuracy of 83.79%, indicating a good level of classification performance. The precision was 86.45%, higher than the accuracy, while the recall was 83.79%, indicating that the model can detect most of the actual weather classes.

The classification results from the SVM algorithm demonstrate that the model performs reasonably well on weather data. The relatively high precision indicates that the model's positive predictions are mostly correct, while the recall suggests that it can identify most relevant instances. Although the performance of SVM is slightly lower than that of ensemble- and neural-network-based models, these results indicate that SVM remains effective for handling structured weather data with clear decision boundaries.

3.3.4 Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) classifier was evaluated in this study, and the results are presented in Figure 5. The MLP model achieved an accuracy of 90.68%, indicating strong classification performance. The precision was 91.73%, slightly higher than the accuracy, while the recall was 90.68%, demonstrating the model's effectiveness in detecting most relevant weather instances.

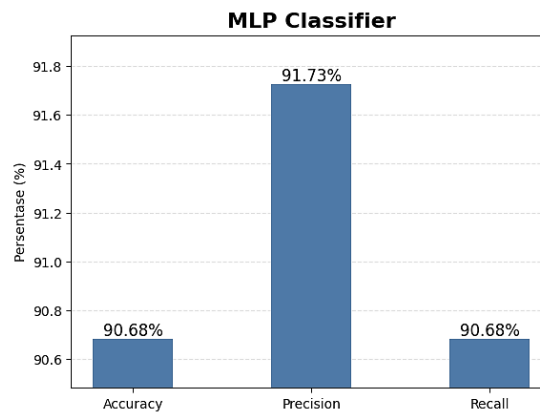


Figure 5. MLP Classifier Results

The classification results using the MLP algorithm show that the model performs very well in classifying weather data. The high recall indicates that the model accurately identifies most weather classes, while the precision, which is close to recall, suggests that the positive predictions are largely correct, with a relatively low false positive rate. Nevertheless, further performance improvements could be achieved through hyperparameter tuning or architectural optimization to enhance both precision and recall.

3.3.5 Artificial Neural Network (ANN)

The final model evaluated was the ANN, which produced strong and consistent classification performance. ANN demonstrated its ability to learn deeper relationships among weather attributes and achieved results comparable to MLP. Based on Figure 6, the application of the Artificial Neural Network (ANN) model for weather type classification achieved an accuracy of 83.79%, indicating good classification performance. The precision was 86.45%, higher than the accuracy, while the recall was 83.79%, indicating that the model can detect most actual weather instances.

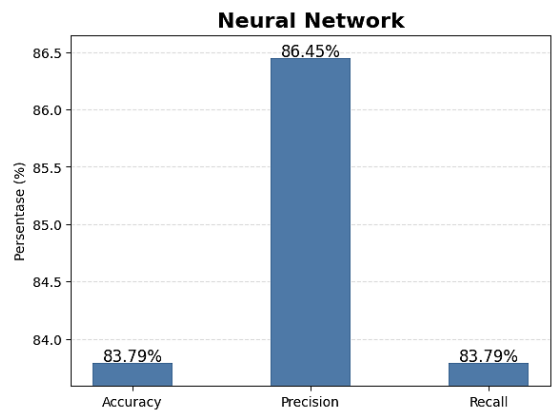


Figure 6. Neural Network Results

The classification results using the ANN model demonstrate that it performs adequately at classifying weather data. The relatively high recall indicates that the model can identify most weather classes, while the precision, which is close to recall, suggests that most of the model's positive predictions are correct. However, there is still potential to improve the model's performance through architectural optimization or parameter tuning to enhance both precision and recall.

3.4. Comparison of Results

This analysis section presents a comparison of classification accuracy obtained using five machine learning algorithms, namely Decision Tree, Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN), under three different data split ratios (70/30, 80/20, and 90/10). The evaluation aims to assess the consistency and generalization capability of each algorithm when trained with varying proportions of training and testing data. The comparison results shown in Figure 7 indicate that Random Forest, SVM, MLP, and ANN consistently achieve higher accuracy than the Decision Tree across all data split ratios.

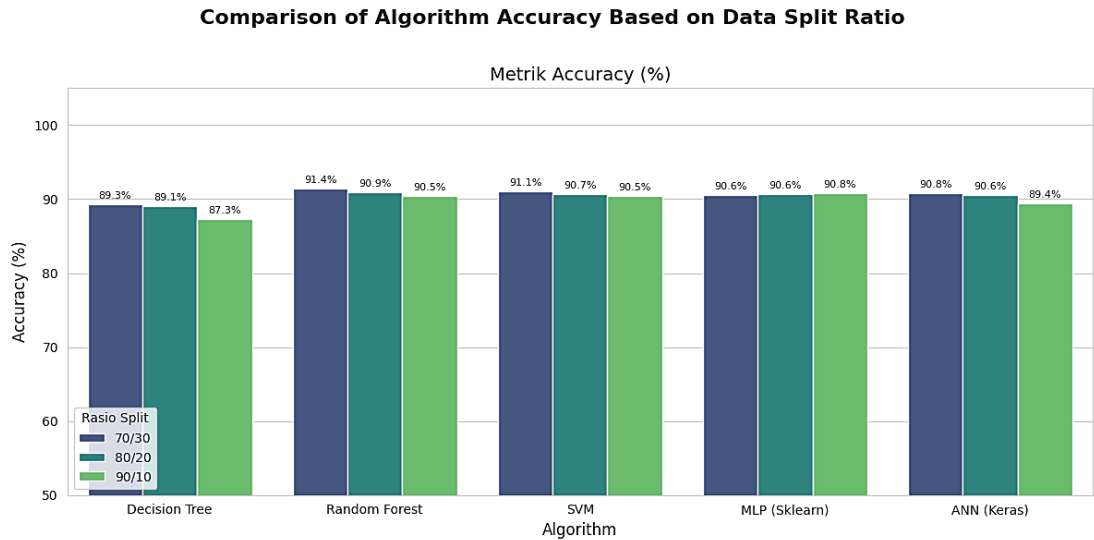


Figure 7. Algorithm Comparison Results

Among all evaluated models, MLP and ANN demonstrate the most stable performance, with accuracy values remaining above 89% for all split ratios. Random Forest also exhibits strong and consistent accuracy, particularly under the 70/30 and 80/20 splits. In contrast, the Decision Tree algorithm shows the lowest accuracy and greater performance variation, especially under the 90/10 split, indicating sensitivity to training data size. Overall, the results confirm that ensemble-based and neural network models provide superior, more reliable performance for weather-type classification compared to single-tree approaches.

3.5. Confusion Matrix Analysis of Classification Models

Among the evaluated models, the Artificial Neural Network (ANN) demonstrates the best overall classification performance based on the confusion matrix analysis. The ANN model shows consistently high

true positive values across all weather classes (Rainy, Snowy, Sunny, and Cloudy), while maintaining relatively low misclassification rates compared to the other algorithms. The diagonal dominance observed in the ANN confusion matrix indicates that the model is able to correctly distinguish between different weather categories with greater reliability and balance.

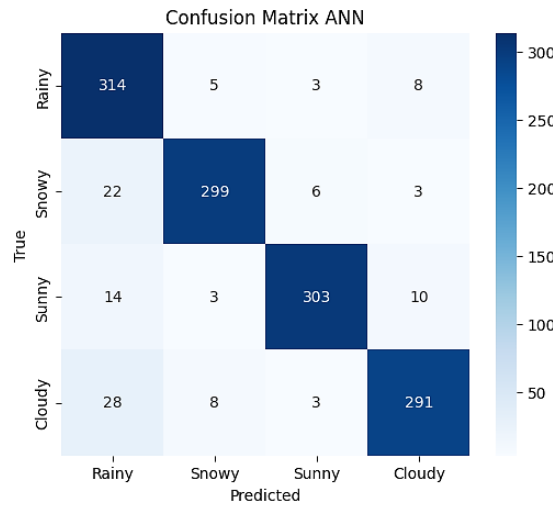


Figure 8. Best Confusion Matrix of Classification Models

Compared to Random Forest and MLP, which still exhibit noticeable confusion between Rainy and Cloudy classes, ANN produces fewer off-diagonal errors, suggesting improved class separation for meteorological conditions with overlapping features. In contrast, SVM and Decision Tree models show higher levels of misclassification across multiple classes, particularly for Rainy, Sunny, and Cloudy weather types, reflecting limitations in handling complex nonlinear patterns. The superior performance of ANN can be attributed to its ability to learn deeper and more complex feature representations through multiple hidden layers, allowing it to capture subtle nonlinear relationships within the weather dataset. Therefore, based on the confusion matrix evaluation, ANN is identified as the most effective model for multi-class weather type classification in this study.

3.6. Comparison of Classification Models Using ROC Curves

The ROC Curves of Classification Models can be view Figure 9.

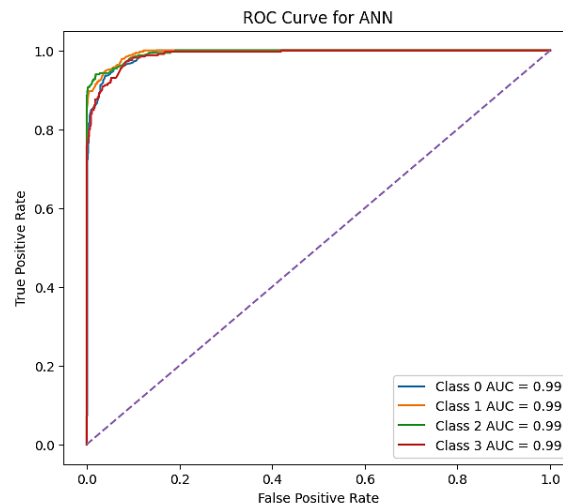


Figure 9. Best ROC Curves of Classification Models

Compared to other models, such as Decision Tree and SVM, which exhibit slightly lower AUC values and greater class-wise variation, ANN provides the most stable and uniform ROC performance. Although Random Forest and MLP also achieve high AUC values, the ANN model shows superior consistency across all classes, suggesting greater robustness to overlapping meteorological features. The superior ROC performance of ANN can be attributed to its deep learning architecture, which enables the model to learn complex nonlinear relationships and capture subtle feature interactions within the weather dataset. Therefore,

based on the ROC curve evaluation, ANN is identified as the most effective model in terms of overall classification discrimination capability in this study.

3.7. Discussion

The results of this study demonstrate that machine learning algorithms exhibit varying levels of performance in classifying weather types based on structured meteorological data. Among the five evaluated models, namely Decision Tree, Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN), ensemble-based and neural network models consistently outperform single-tree approaches across different data split ratios. This finding indicates that models capable of capturing complex nonlinear relationships are better suited to multi-class weather-type classification.

Random Forest shows strong and stable performance across all experimental settings, particularly in terms of accuracy and ROC-AUC values. Its ensemble mechanism enables the model to reduce overfitting and effectively handle feature interactions among meteorological variables such as temperature, humidity, precipitation, and cloud cover [33], [34]. This result aligns with previous studies, which reported that Random Forest performs exceptionally well in weather-related classification tasks due to its robustness and ability to manage high-dimensional data [9], [25]. The consistently high AUC values obtained in this study further confirm its strong discriminative capability across different weather classes.

Neural network-based models, namely MLP and ANN, demonstrate the highest overall accuracy and the most consistent performance across varying data split ratios. These models effectively learn nonlinear patterns in synthetic and numerical weather attributes, resulting in balanced precision and recall. Similar findings have been reported in previous studies, which highlight the effectiveness of neural networks in modeling complex atmospheric relationships. The high ROC-AUC values achieved by both MLP and ANN in this study further support their suitability for multi-class weather classification tasks.

The SVM model achieves satisfactory performance, particularly in scenarios where weather classes exhibit clearer feature separation. However, its performance is slightly lower than that of Random Forest and neural network-based models, especially as class overlap increases. This limitation is consistent with existing literature, which suggests that SVM performance may degrade in highly overlapping multi-class datasets without extensive kernel tuning [8], [35].

In contrast, the Decision Tree algorithm exhibits the lowest and most inconsistent performance across all data split ratios. The confusion matrix analysis reveals a higher rate of misclassification between weather types with similar meteorological characteristics, such as Rainy and Cloudy. This behavior indicates that single-tree models are more sensitive to data variation and prone to overfitting, which has also been reported in previous weather forecasting studies [26], [36].

Overall, the discussion confirms that ensemble- and neural network-based approaches provide superior generalization and stability for weather-type classification on publicly available datasets. The consistency of these findings with prior research strengthens the validity of this study and highlights the importance of selecting appropriate model architectures when dealing with complex, multi-class meteorological data.

4. CONCLUSION

This research aims to evaluate the effectiveness of various machine learning classification algorithms, namely Decision Tree, Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN), in classifying weather types using a publicly available Kaggle Weather Type Classification dataset. The evaluation was conducted using multiple data split ratios (70/30, 80/20, and 90/10) to ensure a comprehensive and reliable assessment of model performance. Several evaluation metrics were employed, including accuracy, precision, recall, confusion matrix, and ROC curve, to provide a thorough analysis of each model's classification capability.

The experimental results indicate that Random Forest, MLP, and ANN outperform Decision Tree and SVM. Among all evaluated algorithms, MLP and ANN achieve the most stable and consistently high accuracy across different data split ratios, while Random Forest also exhibits strong, reliable performance with excellent discriminative capability, as indicated by high AUC values. In contrast, the Decision Tree algorithm shows the lowest accuracy and greater sensitivity to data partitioning, highlighting its limitations in handling complex and overlapping meteorological patterns. These findings confirm that ensemble-based and neural network models are more effective in capturing nonlinear relationships within multi-class weather datasets.

The results of this study are consistent with previous research, which reports that Random Forest and neural network-based approaches perform well in weather classification tasks due to their robustness and ability to model complex feature interactions. The strong performance of MLP and ANN further demonstrates the potential of neural network architectures in learning intricate atmospheric patterns from structured meteorological data. This consistency with existing studies reinforces the reliability and validity of the findings presented in this research.

The practical implications of this study extend to real-world weather prediction and decision-support systems. Accurate weather type classification can support applications in agriculture, transportation, logistics,

and disaster management by providing more reliable environmental insights. By leveraging machine learning models such as Random Forest, MLP, and ANN, stakeholders can enhance predictive accuracy and improve planning and risk mitigation strategies based on weather conditions.

Despite the promising results, this study has several limitations. First, the experiments were conducted using a publicly available Kaggle dataset containing structured and partially synthetic meteorological data, which may not fully capture the complexity of real-world weather conditions. Second, the evaluation focused on classical machine learning and feed-forward neural network models without considering temporal dependencies or sequential weather patterns. These limitations may affect the generalizability of the findings when applied to operational meteorological forecasting systems.

Future research should validate the proposed benchmark using real-world meteorological observations collected from weather stations, satellite imagery, or Internet of Things (IoT)-based sensor networks. Furthermore, advanced machine learning and deep learning architectures, such as Long Short-Term Memory (LSTM), Transformer-based models, and hybrid ensemble approaches, could be investigated to better capture temporal weather dynamics and further improve classification performance. Such advancements would contribute to the development of more accurate, adaptive, and scalable weather classification systems for practical decision-support applications.

REFERENCES

- [1] L. Ji and S. Li, "A dynamic financial risk prediction system for enterprises based on gradient boosting decision tree algorithm," *Syst. Soft Comput.*, vol. 7, Dec. 2025, doi: 10.1016/j.sasc.2025.200189.
- [2] M. Alizade-Harakiyan, A. Khodaei, A. Yousefi, H. Zamani, and A. Mesbahi, "Decision tree-based machine learning algorithm for prediction of acute radiation esophagitis," *Biochem. Biophys. Reports*, vol. 42, Jun. 2025, doi: 10.1016/j.bbrep.2025.101991.
- [3] P. Mukherji, S. Rajput, and V. Mudaliar, "A novel accelerated sparse Support Vector Machine (AS-SVM) algorithm for binary classification of DNA sequences," *Franklin Open*, vol. 13, p. 100427, Dec. 2025, doi: 10.1016/j.fraope.2025.100427.
- [4] Y. Yuan, K. Wang, D. Duives, W. Daamen, and S. P. Hoogendoorn, "Machine learning-based bicycle delay estimation at signalized intersections using sparse GPS data and traffic control signals - A Dutch case study using random forest algorithm," *Artif. Intell. Transp.*, vol. 3–4, p. 100037, Nov. 2025, doi: 10.1016/j.ait.2025.100037.
- [5] M. K. Risqi, I. Wisma, D. Prastya, and M. J. Vikri, "Comparison of Decision Tree Algorithms and Support Vector Machine (SVM) In Depression Classification In Students," *Eduvest-Journal Univers. Stud.*, vol. 5, no. 4, pp. 4557–4567, 2025, [Online]. Available: <http://eduvest.greenvest.co.id>
- [6] R. F. Ramadhan and W. M. Ashari, "Performance Comparison of Random Forest and Decision Tree Algorithms for Anomaly Detection in Networks," 2024. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [7] R. S. Saputra, M. Panji, and A. Saputra, "Comparison of Random Forest and SVM Algorithms in Classification of Diabetic Retinopathy Based on Fundus Image Texture Features," *Int. J. Quant. Res. Model.*, vol. 6, no. 2, pp. 218–226, 2025.
- [8] J. Zhang, X. Qiu, X. Li, Z. Huang, M. Wu, and Y. Dong, "Support Vector Machine Weather Prediction Technology Based on the Improved Quantum Optimization Algorithm," *Comput. Intell. Neurosci.*, vol. 2021, 2021, doi: 10.1155/2021/6653659.
- [9] A. J. Hill, R. S. Schumacher, and I. L. Jirak, "A New Paradigm for Medium-Range Severe Weather Forecasts: Probabilistic Random Forest-Based Predictions," *Weather Forecast.*, vol. 38, no. 2, pp. 251–272, Feb. 2023, doi: 10.1175/WAF-D-22-0143.1.
- [10] R. Bhardwaj and V. Duhoon, "Weather forecasting using decision tree," *J. Eng. Res.*, vol. 9, 2021, doi: 10.36909/jer.ICCEMME.15635.
- [11] A. Hashemi Fath, F. Madanifar, and M. Abbasi, "Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems," *Petroleum*, vol. 6, no. 1, pp. 80–91, Mar. 2020, doi: 10.1016/j.petlm.2018.12.002.
- [12] A. Sumayli, "Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models," *Arab. J. Chem.*, vol. 16, no. 7, Jul. 2023, doi: 10.1016/j.arabjc.2023.104833.
- [13] M. Korycinski, K. A. Ciecierski, and E. Niewiadomska-Szynkiewicz, "HyTract: Advancing tractography for neurosurgical planning with a hybrid method integrating neural networks and a path search algorithm," *Neural Networks*, vol. 190, Oct. 2025, doi: 10.1016/j.neunet.2025.107624.
- [14] N. Xie, J. He, W. Wei, J. Ba, and T. Yang, "Neural Network-Driven Genetic Algorithm for Advanced Fuel Cell Performance Enhancement," *Results Eng.*, p. 108426, Nov. 2025, doi: 10.1016/j.rineng.2025.108426.
- [15] E. S. Mohamed, T. A. Naqishbandi, S. A. C. Bukhari, I. Rauf, V. Sawrikar, and A. Hussain, "A hybrid

- mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms,” *Healthc. Anal.*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100185.
- [16] T. Ansar and W. M. Ashraf, “Comparison of Kolmogorov–Arnold Networks and Multi-Layer Perceptron for modelling and optimisation analysis of energy systems,” *Energy AI*, vol. 20, May 2025, doi: 10.1016/j.egyai.2025.100473.
- [17] F. Q. Kareem, A. M. Abdulazeez, and D. A. Hasan, “Predicting Weather Forecasting State Based on Data Mining Classification Algorithms,” *Asian J. Res. Comput. Sci.*, pp. 13–24, Jun. 2021, doi: 10.9734/ajrcos/2021/v9i330222.
- [18] A. Yaqoob, M. A. Mir, and M. A. Shah, “Optimized cancer gene selection using armadillo optimization algorithm and support vector machine,” *Cancer Treat. Res. Commun.*, vol. 45, Jan. 2025, doi: 10.1016/j.ctarc.2025.101027.
- [19] A. Thakur, P. Singh, and K. Lata, “Genetic Algorithm-Driven Random Forest Model for Predicting Acute Respiratory Distress Syndrome (ARDS),” in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 1709–1718, doi: 10.1016/j.procs.2025.04.423.
- [20] S. Sanaye and A. Farvizi, “Artificial -neural -network and genetic -algorithm for optimization of helical -blade -vertical -axis -wind -turbine,” *Clean. Eng. Technol.*, vol. 29, Dec. 2025, doi: 10.1016/j.clet.2025.101088.
- [21] P. Joyarieb, V. C. Silalahi, V. Anggelica, and K. K. Ongso, “Comprehensive Approach to Weather Prediction with the Random Forest Algorithm,” 2025. [Online]. Available: <https://citedness.com/index.php/jdsihttps://citedness.com/index.php/jdsi>
- [22] H. Nurramdhani Irmanda, M. Khalid Awang, and M. Adrezo, “International Journal On Informatics Visualization journal homepage : www.joiv.org/index.php/joiv International Journal On Informatics Visualization Enhancing Weather Prediction Models through the Application of Random Forest Method and Chi-Square Feature Selection.” [Online]. Available: www.joiv.org/index.php/joiv
- [23] A. Kaur, “Weather Forecasting using Decision Tree Regression,” *International J. Sci. Res. Eng. Manag.*, vol. 06, no. 05, Jun. 2022, doi: 10.55041/ijrsrem16084.
- [24] M. Chhetri, “Manuscript Rainfall Prediction using Multilayer Perceptron (MLP),” *Int. J. Sci. Res. Eng. Manag. (IJSREM)*, p. 7, 2023, doi: 10.55041/IJSREM26194.
- [25] M. A. Indrajaya, T. S. Sollu, M. Subito, Y. A. Rahman, and E. A. Saputra, “Classification of weather conditions based on automatic weather station data using a multi-layer perceptron neural network,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 37, no. 1, pp. 540–550, Jan. 2025, doi: 10.11591/ijeecs.v37.i1.pp540-550.
- [26] J. M. Han, Y. Q. Ang, A. Malkawi, and H. W. Samuelson, “Using recurrent neural networks for localized weather prediction with combined use of public airport data and on-site measurements,” *Build. Environ.*, vol. 192, Apr. 2021, doi: 10.1016/j.buildenv.2021.107601.
- [27] S. Sathyanarayanan, “Confusion Matrix-Based Performance Evaluation Metrics,” *African J. Biomed. Res.*, pp. 4023–4031, Nov. 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [28] I. Markoulidakis and G. Markoulidakis, “Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis,” *Technologies*, vol. 12, no. 7, Jul. 2024, doi: 10.3390/technologies12070113.
- [29] A. Vanacore, M. S. Pellegrino, and A. Ciardiello, “Fair evaluation of classifier predictive performance based on binary confusion matrix,” *Comput. Stat.*, vol. 39, no. 1, pp. 363–383, Feb. 2024, doi: 10.1007/s00180-022-01301-9.
- [30] G. Zeng, “Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification,” *Mathematics*, vol. 13, no. 16, Aug. 2025, doi: 10.3390/math13162609.
- [31] A. I. Fadri, A. Zahfran, T. Irak, N. H. Firjatullah, and J. E. Herianto, “Comparison of Supervised Learning Algorithms for Predicting Airline Passenger Satisfaction,” *IJATIS Indones. J. Appl. Technol. Innov. Sci.*, vol. 2, no. 1, pp. 42–52, Mar. 2025, doi: 10.57152/ijatis.v2i1.1868.
- [32] K. Riehl, M. Neunteufel, and M. Hemberg, “Hierarchical confusion matrix for classification performance evaluation,” Jun. 2023, doi: 10.1093/jrssc/qlad057.
- [33] Y. Wu and Y. Chang, “Ransomware Detection on Linux Using Machine Learning with Random Forest Algorithm,” Jun. 07, 2024, doi: 10.36227/techrxiv.171778770.06550236/v1.
- [34] M. Pal and S. Parija, “Prediction of Heart Diseases using Random Forest,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021, doi: 10.1088/1742-6596/1817/1/012009.
- [35] O. A. Alimi, K. Ouahada, A. M. Abu-Mahfouz, and S. Rimer, “Power system events classification using genetic algorithm based feature weighting technique for support vector machine,” *Heliyon*, vol. 7, no. 1, Jan. 2021, doi: 10.1016/j.heliyon.2021.e05936.
- [36] B. T. Jijo and A. M. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, Jan. 2021, doi: 10.38094/jastt20165.