Indonesian Physical Review
Volume 09 Issue 01, January 2026
P-ISSN: 2615-1278, E-ISSN: 2614-7904

Evaluation of the C4.5 Method for Thunderstorm Classification at Sultan Hasanuddin Airport Using Radiosonde Stability Indices (2013–2024)

Muhammad Syawal[1,3*], Eko Hadi Sujiono[2], Husain[2], Adi Prasetiyo[1,3]
[1] *Physics Study Program, Postgraduate Program of Universitas Negeri Makassar, Makassar, Indonesia*
[2] *Department of Physics, Universitas Negeri Makassar, Makassar, Indonesia*
[3] *Hasanuddin Weather Observation Station, Maros, Indonesia*

Corresponding Author's E-mail: 240026301002@student.unm.ac.id

Article Info

Article info:
Received: 02-09-2025
Revised: 12-11-2025
Accepted: 24-11-2025

Keywords:
Thunderstorm; C4.5 Method; Showalter Index; Lifted Index; K Index; Precipitable Water

How To Cite:
M. Syawal, E. H. Sujiono, Husain, and A. Prasetiyo, "Evaluating a C4.5 Method for Thunderstorm Classification at Sultan Hasanuddin Airport Using Radiosonde Stability Indices (2013–2024)", Indonesian Physical Review, vol. 9, no. 1, p 107-117, 2026.

DOI:
https://doi.org/10.29303/ipr.v9i1.570.

Abstract

Thunderstorms pose a serious threat to aviation, including at Sultan Hasanuddin International Airport. Weather forecasters typically use radiosonde data, including the Showalter Index (SI), Lifted Index (LI), K Index (KI), and Precipitable Water (PW), to classify thunderstorm events. These data have varying values, which can potentially cause overlapping index distributions and subjectivity in decision-making. Therefore, the C4.5 method is needed to minimize this potential. The C4.5 method generally consists of a tree of root nodes leading to leaf nodes, derived from gain and entropy information. This research aims to classify thunderstorm occurrences using the C4.5 method and to verify the results using accuracy, recall, balanced accuracy, True Skill Statistic (TSS), and Critical Success Index (CSI). The data used in this study span the period from 2013 to 2024, with a 12-hour time interval (00 UTC and 12 UTC), encompassing SI, LI, KI, and PW data sourced from radiosonde launches, as well as thunderstorm occurrence data obtained from synop codes. The data from 2013 to 2024 was then divided into two parts, namely training data (2013-2021) and testing data (2022-2024). The classification results for 2022-2024 were dominated by the non-occurrence of thunderstorms, with 1901 occurrences, while there were only 31 thunderstorm occurrences. Overall, the C4.5 method achieves relatively good accuracy (0.785). However, recall (0.027), balanced accuracy (0.507), TSS (0.014), and CSI (0.026) are low.

Introduction

Sultan Hasanuddin International Airport is one of the busiest airports in Indonesia, particularly in the eastern part of the country. One of the many variables affecting the seamless operation of flight operations at the airport is the weather. Based on data compiled by the Meteorology, Climatology, and Geophysics Agency (BMKG) and the Makassar Air Traffic

Service Center (MATSC) between 2017 and 2024, 536 aircraft experienced disruptions during landing and takeoff at Sultan Hasanuddin International Airport due to bad weather, one of which was caused by thunderstorms. Thunderstorms are a type of weather phenomenon characterized by lightning, thunder, heavy rain, and strong winds [1]. This phenomenon can occur in various regions due to unstable atmospheric conditions, causing adverse effects on communities such as flooding, landslides, damage to buildings, power outages, transportation disruptions, and even threats to life due to heavy rain, lightning, and strong winds generated by thunderstorms [2], [3].

Thunderstorms can be predicted for the next 12 hours by weather forecasters using data from radiosonde launches such as the Showalter Index (SI), Lifted Index (LI), K Index (KI), and Precipitable Water (PW) [4], [5], [6], [7]. Negative values on the SI and LI indicate unstable atmospheric conditions that have the potential to produce thunderstorms due to massive convective activity [8], [9]. KI is a combination of temperature and dew point at the 850 hPa, 700 hPa, and 500 hPa layers, where high KI values indicate unstable atmospheric conditions that have the potential to produce thunderstorms [10]. PW is a parameter that measures the amount of water vapor in the atmosphere, where high PW values indicate the potential for cloud formation that produces thunderstorms [11].

The diversity of SI, LI, KI, and PW values in classifying thunderstorm occurrences has the potential to cause overlapping index distributions and subjectivity in decision making, thereby reducing the ability to classify thunderstorm occurrences [5], [6], [7]. Therefore, a method that can be used to process diverse data is needed, one of which is by applying the C4.5 method [12], [13]. The C4.5 method is a decision-making machine learning algorithm that takes the shape of an inverted tree [14]. Compared to other classification approaches, the C4.5 method was chosen not because it is considered to have higher generalization capabilities, but because of its easily interpretable and operationally relevant characteristics [15], [16]. In addition, the C4.5 method produces rules (gain and entropy) [14] that can standardize classification decisions and reveal interactions among atmospheric stability indices (SI, LI, KI, and PW). This interpretability is crucial in an operational context because it enables forecasters to understand the reasoning behind each classification decision.

Previous research has used the C4.5 method in various fields, particularly meteorology, to classify rainfall [12], [13], [17]. Research that specifically tests the accuracy of the C4.5 method in classifying thunderstorm occurrences using atmospheric stability index parameters such as SI, LI, KI, and PW is still limited, even though atmospheric stability indices are important indicators that are often used in classifying thunderstorm occurrences [18], [19]. Therefore, this research is expected to fill this gap in the literature by making a scientific contribution, specifically by testing the accuracy of the C4.5 method in classifying thunderstorm occurrences, while supporting improvements in the quality of extreme weather classification at the Hasanuddin Weather Observation Station.

Based on the above understanding, the author is interested in further examining the application of the C4.5 method in classifying thunderstorm occurrences at the Hasanuddin Weather Observation Station using SI, LI, KI, and PW data. The years 2022-2024 were selected as the period used for classifying thunderstorm occurrences or testing, while the years 2013-2021 were selected as the period used for compiling the C4.5 method or training. The selection of years for the training and testing data was based on previous research [20], [21], which stated that the training data should comprise at least 70% of the combined training and testing data to minimize errors in machine learning processing.

Experimental Method

The data used in this research includes SI, LI, KI, and PW data from radiosonde launches at 00 and 12 UTC, as well as hourly thunderstorm event data from synop codes. All data is sourced from observations at the Hasanuddin Weather Observation Station from 2013 to 2024. The SI, LI, KI, and PW equations [22] [23] are described as follows:

$$SI = T_{500} - T_{p_{500}} \quad (1)$$

$$LI = T_{500} - T'_{500} \quad (2)$$

$$KI = (T_{850} - T_{500}) + Td_{850} - (T_{700} - Td_{700}) \quad (3)$$

$$PW = \frac{1}{g}\int_{p1}^{p2} q\, dp \quad (4)$$

with T_{850} is the temperature of the 850 hPa layer (K), T_{700} is the temperature of the 700 hPa layer (K), T_{500} is the temperature of the 500 hPa layer (K), $T_{p_{500}}$ is the temperature of the air parcel in the 500 hPa layer (K), T'_{500} is the temperature of the air parcel rising adiabatically in the 500 hPa layer (K), Td_{850} is the dew point of the 850 hPa layer (K), Td_{700} is the dew point of the 700 hPa layer (K), g is the acceleration due to gravity (9.8 m s^{-2}), p is the pressure (Pa), and q is the specific humidity (g kg^{-1}).

After all data from 2013 to 2024 has been collected, the next steps are as follows:

1. Compare input data (SI, LI, KI, and PW) with output data (thunderstorm occurrences). SI, LI, KI, and PW data at 00 UTC will be compared with thunderstorm event data at 01-12 UTC, while SI, LI, KI, and PW data at 12 UTC will be compared with thunderstorm event data at 13-00 UTC. Selecting a 12-hour time range for thunderstorm event data is because radiosonde observations are conducted twice a day at 00 and 12 UTC, so each radiosonde data point is considered to represent atmospheric conditions for the following 12 hours.
2. Split training data (2013–2021) and testing data (2022–2024) from the combined input and output data.
3. Transform the training data and testing data with detailed input data into 3 parts (low, medium, and high) and output data into 2 parts (occurred and did not occur). The purpose of transforming training and testing data is to simplify model interpretation and reduce the influence of extreme outliers [20]. The boundaries between sections in the input data are obtained from the 33.333 percentile of the training data and the 66.667 percentile of the training data, which are described as follows:

Table 1. Boundaries Between Sections in Input Data

Input data	33.333 percentile	66.667 percentile
SI	1.073333	5.356667
LI	-1.35333	2.613333
KI	14.13333	30.32
PW	43.19667	55.67333

4. Create the C4.5 method (from the root node to the leaf node) every month in *Google Colab* (*python programming language*) using the scikit-learn library and parameters "criterion = entropy"," ccp_alpha = 0" (without pruning), "max_depth = none", and "class_weight = none". The scikit-learn library was chosen because it is stable, easy to reproduce, and has transparent parameters. The "criterion= entropy" selection was made because the entropy function mathematically represents the principle of calculating the information gain ratio in the C4.5 method, which measures the uncertainty of information for each attribute to determine the best separation. The gain value is calculated from the difference between the entropy before and after data separation as in Equations (5) and (6):
$$Entropy\ (S) = \sum_{i=1}^{n} -p_i(log_2\, p_i) \tag{5}$$
$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|}\ Entropy\ (S_i) \tag{6}$$
with S is the set of cases, S_i is the i-th partition of cases, n is the quantity of partitions of S, p_i is the proportion of S_i to S, and A is a feature.
The parameter "ccp_alpha = 0" is used to disable pruning, so that the decision tree develops fully according to the natural pattern of the data. The setting "max_depth = none" allows the C4.5 method to grow until there is no increase in information, while "class_weight = none" keeps the class distribution original so that the method is trained according to the ratio of thunderstorm occurrences and non-occurrences of thunderstorm.
5. Testing the C4.5 method monthly by entering input data into the testing data to obtain the results of thunderstorm occurrences.
6. Verifying the classification of thunderstorm occurrences using accuracy, recall, balanced accuracy, True Skill Statistic (TSS), and Critical Success Index (CSI), the equations for which are described as follows:
$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$
$$recall = \frac{TP}{TP + FN} \tag{8}$$
$$balanced\ accuracy = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{9}$$
$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \tag{10}$$
$$CSI = \frac{TP}{TP + FP + FN} \tag{11}$$
with TP is the quantity of correctly classified thunderstorm occurrences, TN is the quantity of correctly classified thunderstorm non-occurrences, FP is the quantity of incorrectly classified thunderstorm occurrences, and FN is the quantity of incorrectly classified thunderstorm non-occurrences. TP, TN, FP, and FN They are part of the confusion matrix, which will be analyzed further in the results and discussion section.

Result and Discussion

The C4.5 method used in this research spans the period from 2013 to 2021. This algorithm was compiled monthly, beginning with entropy and gain calculations in *Python programming*. After obtaining the monthly entropy and gain, the next step was to determine the root-to-leaf path, thereby forming a single unit that comprised the C4.5 method. One form of the C4.5 method in this research is presented in Figure 1 below.

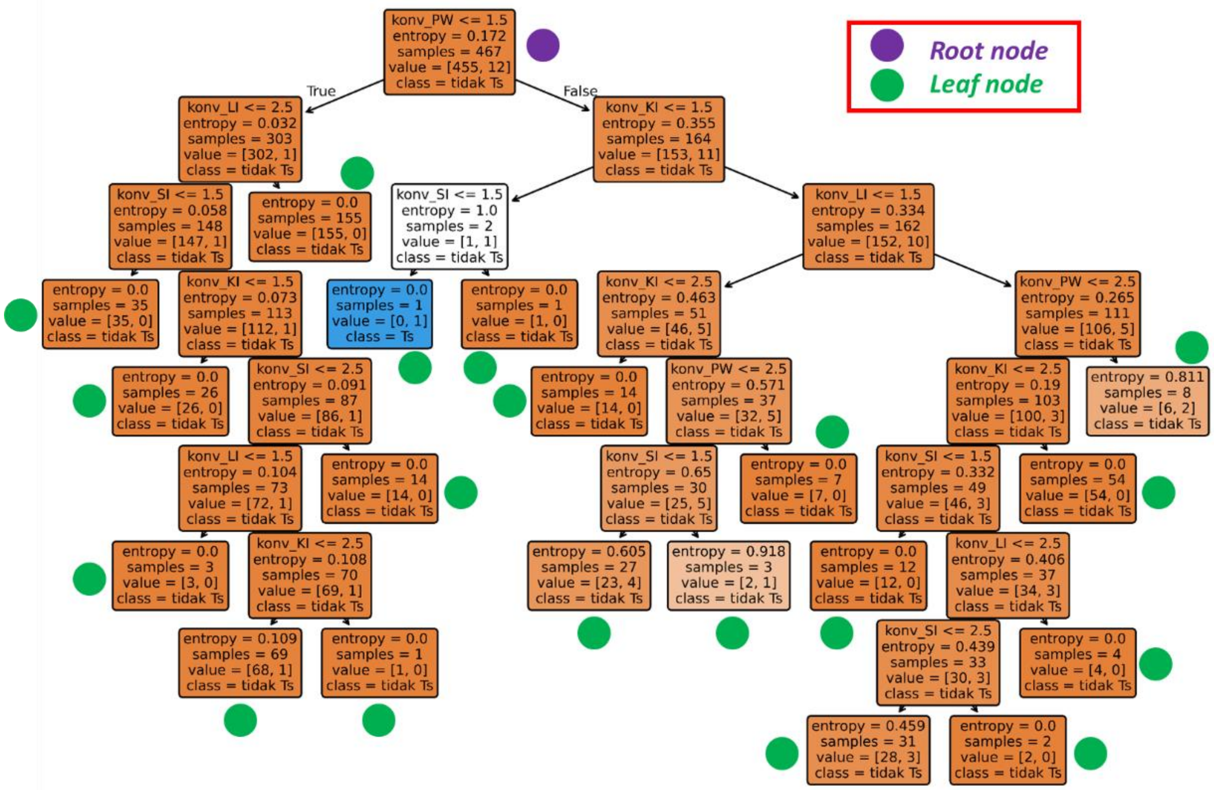

Figure 1. C4.5 method in August

In Figure 1, there is 1 root node marked by a purple circle and 19 leaf nodes marked by green circles. According to [14], the C4.5 method always has 1 root node and several leaf nodes that do not branch, so in this research, 12 root nodes and several leaf nodes were formed from January to December, sourced from monthly training data for the years 2013-2021. Details of the quantity of leaf nodes per month in this research are described in Table 2 below.

Table 2. Details of the Quantity of Leaf Nodes

Month	Ts output	Non-Ts output	Quantity of leaf nodes
Jan	1	20	21
Feb	1	21	22
Mar	2	18	20
Apr	0	21	21
May	1	23	24
Jun	0	22	22
Jul	0	14	14
Aug	1	18	19
Sep	0	20	20
Oct	3	34	37
Nov	5	26	31
Dec	2	20	22

A leaf node is the last node that contains a decision on whether the input data entered into the C4.5 method results in a thunderstorm (Ts output) or non-occurrences of thunderstorm (Non Ts output). Based on Table 2, there are 273 leaf nodes from January to December, with 16 leaf nodes with Ts output and 257 leaf nodes with non-Ts output. The maximum quantity of leaf nodes is in October with 37 leaf nodes, while the minimum quantity of leaf nodes is in July with 14 leaf nodes. April, June, July, and September are the months with the minimal quantity of leaf nodes with Ts output with 0 leaf nodes, while November is the month with the maximum quantity of leaf nodes with Ts output with 5 leaf nodes. July is the month with the minimal quantity of leaf nodes with non-Ts output, with 14 leaf nodes, while October is the month with the maximum quantity of leaf nodes with non-Ts output, with 34 leaf nodes.

The quantity of leaf nodes generated in this research was generally dominated by the non-occurrences of thunderstorm output. This occurred because the majority of the training data consisted of instances without thunderstorm output, so the C4.5 method tended to treat this class as the most informative and dominant pattern representation. The C4.5 method generally selects attributes with the highest gain value, which are the attributes that are most effective in reducing entropy or uncertainty [16]. When the dividing attributes fail to clearly separate the data between the majority and minority classes, the resulting data remains dominated by the majority class. As a result, the entropy value decreases, and the algorithm terminates early, leading to many leaf nodes with no thunderstorm output. In addition, the pruning process will cut minority nodes, such as the thunderstorm output if they are considered insignificant, so that the structure of the C4.5 method ultimately contains more leaf nodes with the non-occurrences of thunderstorm output [14].

The main advantage of the C4.5 method lies in its ability to produce decision rules that are easy to interpret and directly relevant to weather forecasting. Based on the decision tree, several high-frequency decision paths were identified that describe the logical relationships between atmospheric stability indices and thunderstorm occurrences. One of the main rules generated is that if $SI < 1.07$ and $LI < -1.35$, and $KI > 30.32$ and $PW > 55.67$, then a thunderstorm will occur. This rule physically indicates highly unstable atmospheric conditions, characterized by the air parcel's temperature increasing relative to the environmental temperature and by high humidity. This combination supports the growth of thick convective clouds and has the potential to produce thunderstorms. Conversely, the rule is that if $SI > 5.36$ and $LI > 2.61$, and $KI < 14.13$ and $PW < 43.2$, then no thunderstorm will occur. This rule indicates very stable atmospheric conditions, with the air parcel temperature cooling relative to the environment and low humidity. This combination does not support the growth of thick convective clouds and therefore lacks the potential to produce thunderstorms. This interpretation is operationally important because it provides forecasters at Sultan Hasanuddin International Airport with quick guidance for classifying the potential for thunderstorms based on radiosonde launch results.

Table 3. Details of the Confusion Matrix for 2022-2024

Month	TP	FN	FP	TN
Jan	0	41	0	130
Feb	0	35	0	129
Mar	0	35	0	147
Apr	0	50	0	127
May	0	20	5	158
Jun	0	22	0	140
Jul	0	13	0	151
Aug	0	12	0	167
Sep	0	19	0	128
Oct	6	20	8	88
Nov	5	39	4	60
Dec	0	90	3	80
Overall	**11**	**396**	**20**	**1505**

Based on the results of verifying the C4.5 method against actual thunderstorm data from 2022 to 2024, it appears that the C4.5 method more often classifies conditions as non-occurrences of thunderstorm (1901 occurrences) rather than thunderstorm occurrences (31 occurrences). The quantity of true negatives (TN) and false negatives (FN) is much greater than the quantity of true positives (TP) and false positives (FP), indicating that machine learning methods such as the C4.5 method have a strong tendency toward the majority class [24]. This is due to the characteristics of thunderstorms at Sultan Hasanuddin International Airport, which are relatively rare, with the frequency of thunderstorms being much lower than the non-occurrences of thunderstorms [3]. In machine learning, this type of data imbalance often leads the method to prioritize the dominant class, resulting in high overall accuracy even though its ability to classify thunderstorm events is limited. Additionally, these results are influenced by the characteristics of the data used. The atmospheric stability index used in this research represents general atmospheric instability and is not yet capable of capturing more specific local factors [25], [26]. Furthermore, radiosondes are launched twice a day (00 UTC and 12 UTC), so atmospheric conditions at launch do not always reflect those immediately prior to a thunderstorm. As a result, the C4.5 method is more effective at identifying stable atmospheric conditions than those conducive to thunderstorm formation.

Table 4. Verification of the C4.5 Method

Month	Accuracy	Recall	Balanced Accuracy	TSS	CSI
Jan	0.76	0	0.5	0	0
Feb	0.787	0	0.5	0	0
Mar	0.808	0	0.5	0	0
Apr	0.718	0	0.5	0	0
May	0.863	0	0.485	-0.031	0
Jun	0.864	0	0.5	0	0
Jul	0.921	0	0.5	0	0
Aug	0.933	0	0.5	0	0
Sep	0.871	0	0.5	0	0
Oct	0.77	0.231	0.574	0.147	0.176
Nov	0.602	0.114	0.526	0.051	0.104
Dec	0.462	0	0.482	-0.036	0
Overall	**0.785**	**0.027**	**0.507**	**0.014**	**0.026**

Based on the results of the C4.5 method performance verification in Table 4, the accuracy is generally 0.7–0.9, with an overall accuracy (combined data for 2022–2024) of 0.785. Although the accuracy appears quite high, this value does not reflect the method's ability to classify thunderstorm occurrences accurately. This is evident from the very low recall value (only 0.027 overall) and from the fact that most months have zero values. The low recall value indicates that the C4.5 method has not recognized most of the thunderstorm events that actually occurred at Sultan Hasanuddin International Airport. In other words, the C4.5 method more often classifies non-occurrences of thunderstorms than thunderstorm occurrences that indicate class inequality. In addition, other metrics, such as balanced accuracy (0.507), TSS (0.014), and CSI (0.026), indicate performance that is not yet capable of properly classifying thunderstorm occurrences. Overall, these results suggest that although the C4.5 method appears accurate, its ability to classify thunderstorm occurrences remains limited.

This research has several limitations that need to be considered. First, the data show a significant imbalance between thunderstorm occurrences and non-occurrences. In the creation of the C4.5 method, class imbalance between thunderstorm occurrences data and non-occurrences of thunderstorm data is not specifically addressed, for example, through class weighting ("class_weight = none"). This is intended to ensure that the C4.5 method is trained according to the original data distribution. However, this is recognized as a methodological limitation that can reduce verification values, such as recall, balanced accuracy, TSS, and CSI. Second, SI, LI, KI, and PW data are not yet capable of representing atmospheric dynamics that contribute to the formation of convective clouds, which produce thunderstorms. Therefore, additional atmospheric dynamic parameters are needed so that the C4.5 method can more comprehensively represent the process of convective cloud formation that produces thunderstorms. Third, limiting the data to every 12 hours (00 UTC and 12 UTC) also affects the C4.5 method's ability to capture rapidly changing atmospheric dynamics. Significant changes in atmospheric conditions can occur within 12 hours, potentially causing discrepancies between the atmospheric parameters used as input and thunderstorm occurrences.

Compared to previous research that determined thunderstorm occurrence using the Artificial Neural Network Perceptron (ANNP) method and stability index thresholds [3], [5], [6]This research achieves a higher combined accuracy than previous studies, with previous studies reporting combined accuracies ranging from 50% to 70%. The high accuracy of the C4.5 method compared to the ANNP method, and the stability index threshold, are due to its effectiveness in handling data with a combination of numerical and categorical attributes and its ability to form explicit, easy-to-understand decision rules. The C4.5 method can also use multiple variables simultaneously, rather than relying on a single stability index, as in the stability index threshold method. Meanwhile, the ANNP method can be less accurate, generally due to suboptimal network architecture and hyperparameter configuration.

Conclusion

Overall, the C4.5 method classified more non-occurrences of thunderstorms (1901 occurrences) than thunderstorm occurrences (31 occurrences). For the overall verification results, the C4.5 method has a relatively good accuracy (0.785), but for recall (0.027), balanced accuracy (0.507), TSS (0.014), and CSI (0.026) have low values, indicating that the C4.5 method is not yet capable of classifying thunderstorm occurrences properly. This research has several limitations, including significant class imbalance, input data that do not adequately represent thunderstorm formation, and a relatively long data time range (12 hours). Therefore, it is recommended that future research employ class weighting or resampling on the data, incorporate input data that significantly influences thunderstorm formation, and enhance the temporal resolution of the data used (to less than 12 hours).

Acknowledgment

The author would like to thank the Hasanuddin Weather Observation Station for the data used in this research.

References

[1] M. Allaby, *Encyclopedia of Weather and Climate*, Rev. ed. in Facts on File science library. New York: Facts on File, 2007.

[2] C. D. Ahrens, Ed., *Meteorology Today: An Introduction to Weather, Climate, and the Environment*, 9th ed., international student ed. Belmont, Calif.: Brooks, Cole, 2009.

[3] A. A. Rusdin, P. Palloan, S. Subaer, and A. Prasetiyo, “Uji Akurasi Ambang Batas Indeks Stabilitas Atmosfer Terhadap Pembentukan Thunderstorm dan Awan Cumulonimbus di Stasiun Meteorologi Kelas I Sultan Hasanuddin,” *JFU*, vol. 12, no. 2, pp. 269–275, Apr. 2023.

[4] S. Bhattacharya and H. C. Bhattacharyya, “A Comparative Study of Severe Thunderstorm Among Statistical and ANN Methodologies,” *Scientific Reports*, vol. 13.

[5] Indra, “Prediksi Kejadian Petir Menggunakan Artificial Neural Network Metode Perceptron dengan Indeks Radiosonde di Wilayah Kabupaten Kepulauan Tanimbar,” *Bulletin of Meteorology, Climatology, Air Quality, Geophysic, and Environment*, vol. 14, no. 2, pp. 32–38, Feb. 2024.

[6] R. Prima, F. Arifianto, Y. Donni H, and A. Avrionesti, “Determination of Air Stability Parameter Threshold Value for Cumulonimbus and Thunderstorm Cloud Events at Kualanamu Meteorological Station,” *J. Technomaterial Phys.*, vol. 5, no. 2, pp. 99–103, Aug. 2023.

[7] W. Suparta, W. S. Putro, and T. K. Darmastono, “Prediction of Thunderstorm Occurrences in Tropical Areas Using a Numerical Model,” *GT*, vol. 16, no. 1, pp. 78–86, Feb. 2021.
[8] A. K. Showalter, “A Stability Index for Thunderstorm Forecasting,” *Bulletin of the American Meteorological Society*, vol. 34, no. 6, pp. 250–252, Jun. 1953.
[9] J. G. Galway, “The Lifted Index as a Predictor of Latent Instability,” *Bulletin of the American Meteorological Society*, vol. 37, no. 10, pp. 528–529, Dec. 1956.
[10] J. J. George, *Weather Forecasting for Aeronautics*. London: Academic Press, 1960.
[11] C. Rocken, R. Ware, T. Van Hove, F. Solheim, C. Alber, and J. Johnson, “Sensing Atmospheric Water Vapor with the Global Positioning System,” *Geophysical Research Letters*, vol. 20, no. 23, pp. 2631–2634, Dec. 1993.
[12] D. Lou *et al.*, “K-Means and C4.5 Decision Tree Based Prediction of Long-Term Precipitation Variability in the Poyang Lake Basin, China,” *Atmosphere*, vol. 12, no. 7, p. 834, Jun. 2021.
[13] B. Xiang, C. Zeng, X. Dong, and J. Wang, “The Application of a Decision Tree and Stochastic Forest Model in Summer Precipitation Prediction in Chongqing,” *Atmosphere*, vol. 11, no. 5, p. 508, May 2020.
[14] J. Quinlan, *C4.5 : Programs for Machine Learning*. California: Morgan Kaufmann Publishers, 1993.
[15] D. E. Yanti, L. Framesti, and A. Desiani, “Perbandingan Algoritma C4.5 dan SVM dalam Klasifikasi Penyakit Anemia,” *JIP*, vol. 9, no. 4, pp. 427–434, Aug. 2023.
[16] M. Muslim *et al.*, *Data Mining Algoritma C4.5: Disertai Contoh Kasus dan Penerapannya dengan Program Komputer*. Semarang: Universitas Negeri Semarang, 2019.
[17] I. Risnawati *et al.*, “Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan Berdampak Banjir Daerah Menggunakan Algoritma C4.5,” *j-insan*, vol. 3, no. 2, pp. 78–84, Dec. 2023.
[18] W. Ye, Y. Yu, and J. Cao, “Region-dependent Meteorological Conditions of Thunderstorm Based on Clustering Analysis of Stability and Precipitable Water Indices,” *Intl Journal of Climatology*, vol. 44, no. 8, pp. 2650–2665, Jun. 2024.
[19] N. Sagita and T. Takemi, “Exploring the Thunderstorm Predictors in Indonesia,” *SOLA*, vol. 21, pp. 51–60, 2025.
[20] A. Prasetiyo, “Pemanfaatan Algoritma Decision Tree C4.5 dalam Memprakirakan Hujan di Stasiun Meteorologi Kelas I Sultan Hasanuddin,” *BGB*, vol. 5, no. 1, pp. 47–55, Jun. 2024.
[21] R. Trach, Y. Trach, A. Kiersnowska, A. Markiewicz, M. Lendo-Siwicka, and K. Rusakov, “A Study of Assessment and Prediction of Water Quality Index Using Fuzzy Logic and ANN Models,” *Sustainability*, vol. 14, no. 9, p. 5656, May 2022.
[22] F. L. Agroho, S. Syahreza, and D. Sugiyanto, “Analysis of the Rainfall Event in 2018-2019 Using the Air Stability Index Method at the Meteorological Station of Sultan Iskandar Muda Banda Aceh,” *J. Phys.: Conf. Ser.*, vol. 1882, no. 1, p. 012023, May 2021.
[23] S. M. Hagos *et al.*, “The Relationship between Precipitation and Precipitable Water in CMIP6 Simulations and Implications for Tropical Climatology and Change,” *Journal of Climate*, vol. 34, no. 5, pp. 1587–1600, Mar. 2021.
[24] M. Owusu-Adjei, J. Ben Hayfron-Acquah, T. Frimpong, and G. Abdul-Salaam, “Imbalanced Class Distribution and Performance Evaluation Metrics: A Systematic Review of Prediction Accuracy for Determining Model Performance in Healthcare Systems,” *PLOS Digit Health*, vol. 2, no. 11, p. e0000290, Nov. 2023.

[25] W. Ye, Y. Yu, and J. Cao, "Region-dependent Meteorological Conditions of Thunderstorm Based on Clustering Analysis of Stability and Precipitable Water Indices," *Intl Journal of Climatology*, vol. 44, no. 8, pp. 2650–2665, Jun. 2024.
[26] G. Lu, Y. Ren, S. Fu, and H. Xue, "Statistics of Isolated Deep Convection Initiation and Its Relation to Topography in the North China Area," *JGR Atmospheres*, vol. 128, no. 15, p. e2022JD037949, Aug. 2023.