TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 1, February 2026, pp. 162~174
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i1.27451

Challenges in radar-based non-supercell tornado detection using machine learning approaches

Kiki[1,3], Yonny Koesmaryono[1], Rahmat Hidayat[1], Donaldi Sukma Permana[2], Perdinan[1], Abdullah Ali[3]
[1]Applied Climatology Study Program, Department of Geophysics and Meteorology, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia
[2]Department of Climatology, Indonesian Agency of Meteorology Climatology and Geophysics (BMKG), Jakarta, Indonesia
[3]Department of Public Weather Services, Indonesian Agency of Meteorology Climatology and Geophysics (BMKG), Jakarta, Indonesia

Article Info

Article history:

Received Aug 20, 2025
Revised Nov 28 2025
Accepted Dec 8, 2025

Keywords:

Machine learning
Random forest
Single-polarized radar
Tornadoes
XGBoost

ABSTRACT

Tornado detection in Indonesia remains challenging as most areas are monitored by single-polarization weather radar, while dual-polarization systems offer superior detection capabilities. This study presents a novel approach by applying random forest (RF) and XGBoost machine learning algorithms to detect tornadoes using single-polarization radar data, addressing a critical gap in tropical tornado monitoring where dual-pol infrastructure is limited. Four tornado cases in Surabaya during 2024 were analyzed. Radar features including reflectivity, radial velocity, vorticity, and angular momentum were extracted through a multi-elevation sliding window technique. Spatial labels were assigned based on reports from the Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) with a 7.5 km radius from the event center. The dataset was balanced using synthetic minority over-sampling technique (SMOTE). Evaluation was performed using the leave-one-case-out (LOCO) scheme. Within-case evaluation showed strong performance with area under the curve (AUC) >0.94 for both models. XGBoost achieved higher probability of detection (POD 0.67-0.72) but with elevated false alarm rates (FAR up to 70%). RF demonstrated more balanced performance (POD 0.61-0.65, FAR 0.34-0.35). LOCO evaluation revealed significant POD reduction and FAR increase when tested on new cases. This indicates generalization challenges due to variability in tornado characteristics. This study demonstrates the potential of machine learning for tropical tornado early detection using readily available single-polarization radar.

Corresponding Author:

Yonny Koesmaryono
Applied Climatology Study Program, Department of Geophysics and Meteorology
Faculty of Mathematics and Natural Sciences, IPB University
Bogor, Indonesia
Email: yonny@apps.ipb.ac.id

1. INTRODUCTION

Severe convective storms present a significant challenge to meteorological monitoring and forecasting, particularly in tropical regions. These systems frequently produce significant weather phenomena, including heavy rain, thunderstorms, hail, and tornadoes. Among these, "*puting beliung*" a local term used in Indonesia for small-scale, short-lived tornadoes is a frequent occurrence, particularly during transitional seasons [1]-[3]. Tornadoes, though typically small and short-lived in Indonesia, can still cause severe damage to infrastructure and pose risks to human safety [4]. Meanwhile, in terms of economic losses, for example in Semarang, Central Java, which carries a tornado risk hazard level as high as 28.502%, tornadoes caused an

estimated economic loss of up to USD 51,000 during the period from January 2014 to December 2018 [5]. This extreme phenomenon accounts for approximately 21% of all natural disasters that occurred in Indonesia between 1815 and 2014 [6].

Given these risks, understanding the characteristics of Indonesian tornadoes is critical. Several domestic studies have attempted to analyze these events using available meteorological data. For example, a study by Siswanto and Supari [7] utilized satellite and surface observation data to identify tornado precursors, characterized by a rapid increase in relative humidity approximately one hour in advance, accompanied by a sharp temperature drop and the presence of a "horn-like" pressure anomaly. Rusmala *et al.* [8] analyzed a tornado event in Jakarta using C-band weather radar data from column maximum reflectivity (CMAX), vertical cut (VCUT), and constant altitude plan position indicator (velocity) (CAPPI (V)) at 0.5 km, 1.0 km, and 1.5 km, combined with horizontal wind (HWIND) data. The results showed that the tornado-generating convective cloud developed rapidly, with reflectivity between 35-45 dBZ and wind speeds up to 35 knots. Another study by Kiki *et al.* [3] analyzed the spatiotemporal distribution and trends of tornado occurrences in Indonesia over the past decade and reported that the primary tornado hotspots are located on the island of Java, with a positive trend in tornado frequency of approximately 12 events per year. Yudistira *et al*. [9] analyzed upper-air data and found that negative lifted index (LI) values, elevated K index (KI), total totals index (TT), and convective available potential energy (CAPE) indices, along with higher severe weather threat index (SWEAT) readings, indicated unstable atmospheric conditions conducive to convection and thunderstorm development. Overall, the results demonstrated increasing atmospheric instability and a rising potential for convective and lightning activity compared to the preceding days. Other study has attempted to evaluate community resilience to tornado events. For instance, Hidayat *et al.* [10] conducted research in Donohudan Village, Central Java, and identified a moderate level of resilience among residents. This condition was influenced by factors such as adaptability to changing circumstances and challenges, familial and social support networks, spiritual values, and a strong sense of purpose. Despite these analytical efforts, the detection and monitoring of such short-lived events remain difficult, necessitating robust early warning systems.

The primary technological constraint in Indonesia lies in the radar infrastructure, whereas weather radar is recognized as the primary tool for tornado monitoring and forecasting globally [11]. While Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG) has begun deploying dual-polarization weather radars, the network remains limited. As of 2025, BMKG operates 44 weather radars, comprising 33 single-polarization C-band radars and 11 dual-polarization systems (C-band and X-band). Consequently, most regions, including those frequently experiencing tornado events, are predominantly by single-polarization radar systems. This limitation restricts the availability of advanced polarimetric variables often used in modern storm detection, creating a need for methods that can maximize the utility of existing single-polarization data.

In the global context, advances in data science and artificial intelligence (AI) offer significant opportunities to complement numerical weather prediction and enhance real-time guidance [4]. Recent years have seen the successful application of machine learning to improve tornado detection. For dual-polarization data, Zeng *et al.* [12] introduced an extreme gradient boosting (XGBoost)-based algorithm that enhanced detection accuracy. However, progress has also been made using single-radar data, Sandmæl *et al.* [13] developed the tornado probability algorithm (TORP), a probabilistic machine learning approach utilizing single-radar data to estimate tornado occurrence probabilities. Veillette *et al.* [11] presented the tornado network (TorNet) benchmark dataset, comprising full-resolution polarimetric weather radar data to facilitate developing and evaluating machine learning algorithms for tornado detection and prediction. Various deep learning architectures were evaluated for tornado detection using weather radar data, demonstrating the potential of these models in operational settings [11]. Furthermore, deep learning architectures have shown potential in operational settings, with Zhou [14] proposed a hybrid models combining Kalman filtering, convolutional neural networks (CNNs), and bidirectional long short-term memory networks with multi-head attention mechanisms to improve tornado prediction in the United States. Sufi *et al.* [15] built a tornado compendium that encompasses both current and historical records of tornadoes in Bangladesh, in conjunction with AI-based regression analysis and the new dashboard system, which would enable any strategic decision maker to make evidence-based policy decisions regarding tornado events in Bangladesh. Complementing this, Xue *et al.* [16] introduced the multi-task identification network (MTI-Net), a detection model that utilizes a novel backbone with spatial and channel attention units. This approach has proven highly effective, reducing false alarm rates from 0.94 to 0.46 and achieving a nearly fourfold increase in the hit rate.

Despite these global advancements, there remains a notable gap in the literature concerning the application of machine learning techniques for detecting tornadoes in specifically within the Indonesian maritime continent. Most critical is the lack of research utilizing single-polarization radar data for this purpose, which constitutes the bulk of Indonesia's observational network. Addressing this gap is essential for developing detection systems that are effective in the context of Indonesia's unique meteorological conditions and infrastructure constraints.

Therefore, this study aims to evaluate the performance of machine learning algorithms, specifically random forest (RF) and XGBoost, in detecting tornado events in Indonesia using single-polarization weather radar data. Additionally, the study assesses the impact of incorporating multi-elevation radar data on detection accuracy. By addressing the identified research gap, this study contributes to developing more effective early warning systems for tornadoes in the region. The findings will provide valuable insights into the application of machine learning techniques for tropical regions and archipelagic countries, where conventional detection methods face unique challenges due to complex topography and meteorological patterns.

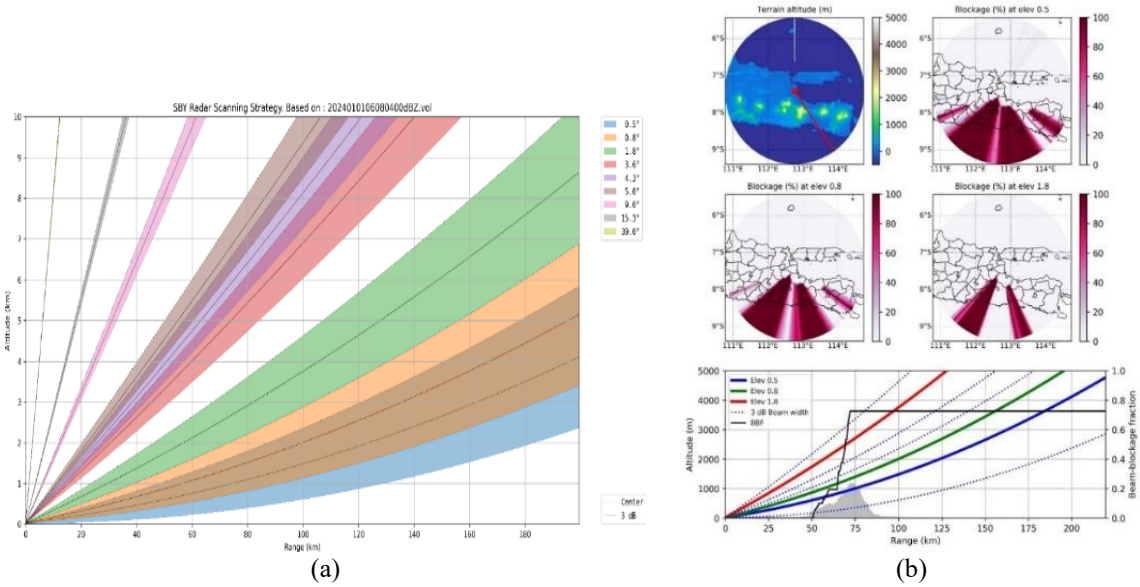
2. METHOD

2.1. Material

The analysis focused on tornado events within the coverage area of the single-polarization weather radar operated by BMKG in Surabaya (SBY). This location was chosen because it is located in East Java, which is one of the provinces with the highest frequency of tornado events in Indonesia in recent years [3]. Four documented tornado events on 1, 9, 17, and 26 January 2024 were selected based on detailed incident reports from the BMKG [17]. These reports provided accurate geographic coordinates, event timings, duration, and impact descriptions, essential for spatial labeling of the radar data. These cases were selected primarily based on data availability and the absence of significant radar beam blockage due to terrain at the event locations. Detailed summaries of the selected tornado events, including coordinates, timings, and reported impacts, are presented in Table 1. The Surabaya radar operates in 5.6 GHz frequency with a Nyquist velocity of 32 m/s, employing a staggered pulse repetition frequency (PRF) scheme of 1000 Hz with a staggering 3/4 ratio and a spatial range resolution of 250 meters. Radar beam blockage analysis results for the three lowest radar elevation angles (0.5°, 0.8°, and 1.8°) are presented in Figures 1(a) and (b), confirming minimal or negligible blockage at all event sites.

Table 1. Selected tornado case studies

Date/time	Location	Impacts
2024-01-01/ 13.00 LT	Kemlangi, Mojokerto, East Java (7.40857907 S,112.36547617 E)	16 houses damaged, several trees uprooted, two vehicles hit by fallen trees
2024-01-17/ 14.30 LT	Megaluh, Jombang, East Java (7.50154633 S,112.18614237 E)	8 houses damage
2024-01-26/ 17.00 LT	Soko, Tuban, East Java (7.09412375 S,111.93226544 E)	Dozens of houses damaged, one injury
2024-02-04/ 16.00 LT	Tarik, Sidoarjo, East Java (7.44837777 S,112.51783611 E)	200 houses damaged, one fatality, two injuries



(a) (b)

Figure 1. Coverage analysis of the Surabaya weather radar; (a) beam spreading for one volume scan and (b) beam blocking analysis for the three lowest elevations

2.2. Methods

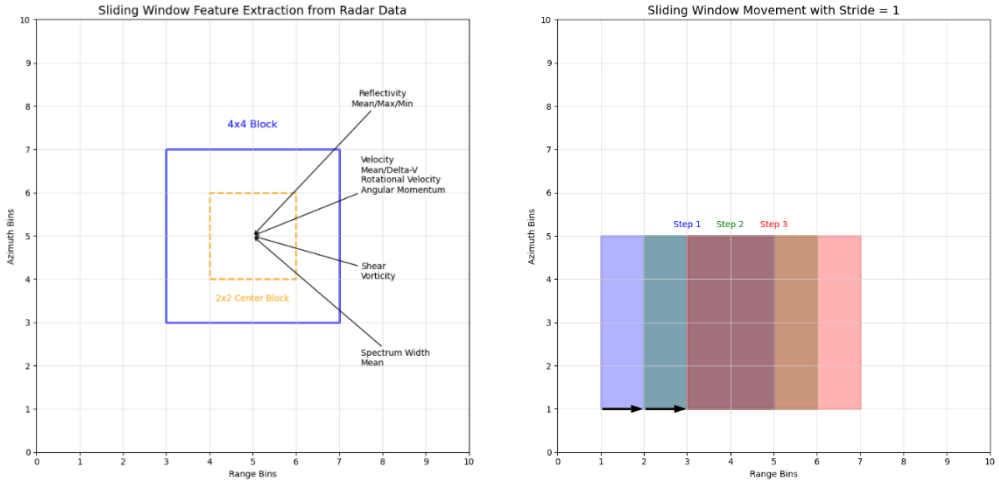
2.2.1. Radar data extraction and preprocessing

The initial phase of the methodology involves preparing the raw radar observations for computational analysis. Radar data were stored in Rainbow-5 format (".vol"). Volume scans at each elevation angle were processed to extract reflectivity, radial velocity, and spectrum width. Extracted radar data were organized into polar coordinates of azimuth angles and range bins. Subsequently, missing or erroneous data were identified and removed using numerical masking procedures to ensure data quality for subsequent analysis. The Wradlib python library is dominantly used in this study to extract radar data [18].

2.2.2. Feature extraction via the sliding window technique

Following data cleaning, spatial features were derived to capture the micro-scale structure of convective storms. Radar-derived features were computed using a sliding window approach, following similar methodologies in previous studies [12]. Each radar scan was divided into overlapping 4×4 pixel spatial blocks (~2 km^2 each, based on 250 meter resolution). Within each block, we extracted reflectivity features (mean, maximum, minimum) to characterize precipitation intensity and heterogeneity, and radial velocity-based features (mean velocity, delta-V, rotational velocity, angular momentum) to capture kinematic properties. Shear and vorticity were derived from velocity gradients to quantify rotational motion, while mean spectrum width assessed turbulence intensity.

To enhance sensitivity to small-scale vortices, a central 2×2 pixel sub-block within each 4×4 block was analyzed separately, adopted following the methodology proposed by [13], which demonstrated that this spatial extent effectively captures convective storm morphology while preserving the resolution of fine-scale vortex signatures. Localized features maximum reflectivity (c4_z_max), maximum radial velocity (c4_v_max), mean spectrum width (c4_w_avg), and central vorticity (c4_vorticity)—were extracted to focus on fine-scale rotational signatures associated with tornadoes. The sliding window operated with stride = 1, ensuring dense sampling and maximizing detection probability of localized vortex structures. Figure 2(a) illustrates the block organization, while Figure 2(b) demonstrates the sequential overlapping movement across the radar field.



(a) (b)

Figure 2. Feature extraction via sliding window scheme: (a) 4×4 sliding window and 2×2 center block from radar data and (b) illustration of overlapping sliding window movement with stride = 1 pixel across the radar scan

2.2.3. Labeling and spatial referencing

To enable supervised learning, the extracted features required precise geospatial alignment with historical ground truth records. Radar blocks were geospatially referenced by converting radar polar coordinates (range and azimuth) into geographic latitude and longitude coordinates. A radius-based labeling approach was implemented, where blocks within a 7.5 km radius from the documented tornado center were labeled as positive (label = 1). In contrast, blocks outside this radius were labeled negative (label = 0). The tornado center was determined from BMKG ground-truth reports. This approach ensured consistent spatial labeling of tornado and non-tornado blocks while reducing uncertainties from radar coverage limitations. The 7.5 km radius was selected based on a preliminary sensitivity analysis. We evaluated labeling radii of 2.5 km,

5.0 km, 7.5 km, and 10.0 km relative to the reported tornado center. The 7.5 km threshold yielded the highest probability of detection (POD) in our validation set, offering the optimal balance between encompassing the tornado's parent circulation and minimizing the inclusion of non-rotational storm areas. Figure 3 illustrates the spatial distribution of positive and negative radar blocks relative to the tornado center.

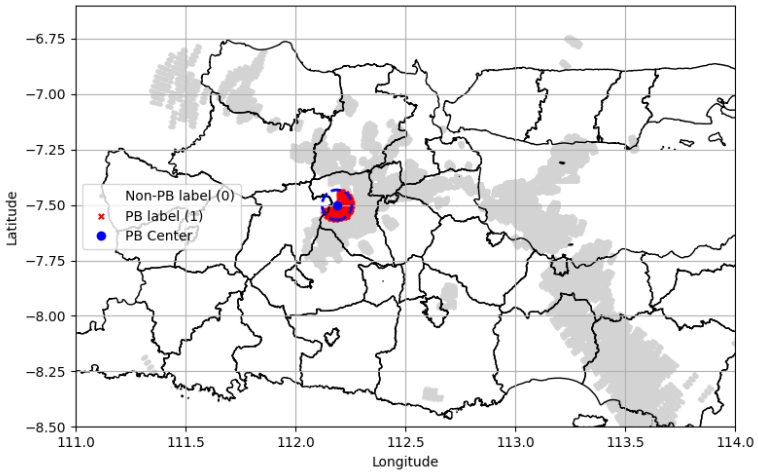


Figure 3. Spatial distribution of radar blocks after labeling, red crosses represent blocks labeled as tornado (1), gray points represent non-tornado blocks (0), and the blue circle marks the documented tornado center

2.2.4. Machine learning algorithm and balancing techniques

Having established a labeled dataset, the study proceeded to implement and evaluate specific classification algorithm. This study employed two supervised machine learning algorithms to classify radar-derived features for tornado detection: RF and XGBoost. The RF algorithm is an ensemble learning method based on aggregating multiple decision trees [19], [20]. Each tree is trained on a random bootstrap sample of the training data, and at each split, a random subset of features is considered. The final prediction is obtained by majority voting among the outputs of individual trees. Mathematically, the RF classifier can be expressed as in (1).

$$\hat{y} = mode(h_1(x), h_2(x), \ldots., h_t(x)) \quad (1)$$

where $h_T(x)$ denotes the prediction of the t-th decision tree for input feature vector x, and T is the total number of trees. XGBoost [21] is a scalable and efficient implementation of gradient-boosted decision trees (GBDT). XGBoost builds additive models in a forward stage-wise fashion, where new trees are fitted to correct the residual errors of prior trees. The objective function $\mathcal{L}$ minimized during training consists of a regularized loss, expressed as in (2).

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (2)$$

where $l(y_i,\hat{y}_i)$ denotes the loss function (e.g., logistic loss for binary classification) between the true label y and the predicted label $\hat{y}_i$, f_k is the k-th tree, and $\Omega(f)$ represents a regularization term penalizing the complexity of each tree to avoid overfitting. Both RF and XGBoost were trained using the radar-derived feature set described previously. Hyperparameters were selected empirically based on preliminary experiments: for RF, the number of trees was set to 100; for XGBoost, the number of estimators was set to 2000 with a learning rate of 0.01, maximum tree depth of 5, and a column subsampling ratio of 0.1.

Due to the severe imbalance between positive (tornado) and negative samples, a synthetic minority oversampling technique (SMOTE) [22] was applied prior to model training. SMOTE generates synthetic examples of the minority class by interpolating between existing minority class samples and their nearest neighbors. Given a minority sample x and one of its nearest neighbors x_{nn}, a synthetic sample x_{new} is generated as in (3). The initial dataset exhibited a severe class imbalance inherent to the rare nature of tornado events. The radar domain was gridded into a 1000×1000 matrix, yielding a total of 1,000,000 potential samples per

scan. Based on the 7.5 km radius labeling criteria, the minority class (positive/tornado) comprised approximately 2,800 samples, whereas the majority class (negative/non-tornado) accounted for the remaining ~997,000 samples. SMOTE generated synthetic minority samples to achieve parity with the majority class. Consequently, the post-balancing dataset consisted of approximately 997,000 positive samples (combining real and synthetic instances) and 997,000 negative samples, resulting in a total training dataset of approximately 2,000,000 samples. SMOTE improves accuracy in unbalanced datasets by generating synthetic samples that maintain the same statistical distribution as the tornado data [23].

$$x_{new} = x + \lambda(x_{nn} - x) \quad (3)$$

where λ is a random number in [0,1]. This balancing procedure was applied only to the training set to avoid information leakage into the test set.

2.2.5. Model training and validation

The compiled dataset was divided into training and testing subsets using a 70-30% split, stratified by class labels to ensure proportional representation of both tornado and non-tornado samples. Before training, feature scaling was performed using the StandardScaler to normalize feature distributions, thereby improving the convergence behavior and stability of the machine learning algorithms. The RF and XGBoost models were trained and validated on this balanced and scaled dataset. Table 2 summarizes the features extracted for this analysis.

Model performance was quantitatively assessed using multiple evaluation metrics. The area under the curve (AUC) [24] and the receiver operating characteristic (ROC) [25] curve was used to measure the model's ability to discriminate between positive and negative classes. The POD, equivalent to recall, was calculated to quantify the proportion of correctly identified tornado events. In addition, the false alarm rate (FAR) was evaluated to indicate the frequency of incorrect positive predictions. The F1-score was computed as a harmonic mean between precision and recall, providing a balanced measure of classification performance. Lastly, a confusion matrix was analyzed to offer a comprehensive view of true positive, true negative, false positive, and false negative rates. To rigorously assess the generalization capability of the models, a leave-one-case-out (LOCO) evaluation was conducted, where the model was trained on all but one case and tested on the remaining unseen event. This approach enabled the evaluation of model robustness across different tornado cases, emphasizing the challenges of generalizing machine learning models for small-scale tropical vortices with highly localized characteristics.

Table 2. Features extracted in the study

Features	Definition	Physical interpretation
Reflectivity	Z_max_1, Z_max_2, Z_max_3, Z_avg_1, Z_avg_2, Z_avg_3, Z_min_1, Z_min_2, Z_min_3, c4_z_max_1, c4_z_max_2, c4_z_max_3	Reflectivity (max/avg/min) from horizontal polarization from tilt 1, 2, and 3; c4 is central-block reflectivity on the localized core of the storm
Velocity	V_avg_1, V_avg_2, V_avg_3, rotational_velocity_avg_1, rotational_velocity_avg_2, rotational_velocity_avg_3, rotational_velocity_max_1, rotational_velocity_max_2, rotational_velocity_max_3, angular_momentum_max_1, angular_momentum_max_2, angular_momentum_max_3, angular_momentum_avg_1, angular_momentum_avg_2, angular_momentum_avg_3, delta_V_1, delta_V_2, delta_V_3, shear_min_1, shear_min_2, shear_min_3, shear_max_1, shear_max_2, shear_max_3, c4_vorticity_1, c4_vorticity_2, c4_vorticity_3	Radial velocity (max/min/avg) from horizontal polarization from tilt 1, 2, and 3; rotational velocity represents the magnitude of the inbound and outbound wind couplet; angular momentum and vorticity represents the radius of a rotating column contracts (stretching), wind speeds increase; delta_V and shear capture the lateral change in wind speed over a short distance
Spectral width	W_avg_1, W_avg_2, W_avg_3, W_max_1, W_max_2, W_max_3, W_min_1, W_min_2, W_min_3, c4_w_avg_1, c4_w_avg_2, c4_w_avg_3, c4_w_max_1, c4_w_max_2, c4_w_max_3	Spectral width (max, avg, min) from horizontal polarization from tilt 1, 2, and 3; c4 is central-block reflectivity on the localized core of the storm

In addition to the LOCO evaluation, inference was also performed on an independent tornado event that was not used during model training. This inference step served as a real-world validation to further examine the model's operational applicability and its ability to predict unseen cases based solely on the extracted radar features. The overall workflow of the tornado detection system, encompassing data preparation, feature engineering, and classification stages presented in Figure 4.

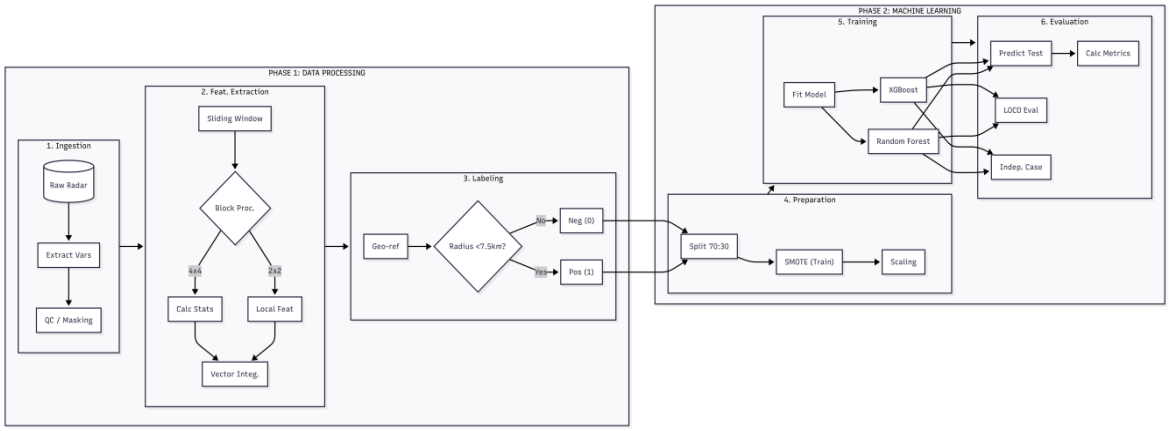


Figure 4. Flowchart of the proposed methodology

3. RESULTS AND DISCUSSION

3.1. Spatial feature distribution and feature importance

Each panel in Figures 5(a) to (d) displays four spatial maps on the left: vorticity, delta-v, angular momentum, and z_max reflectivity, all derived from the lowest radar elevation (0.5°). The blue dot indicates the center of the reported tornado. Right-side barplots show the top five most important features according to the RF and XGB models. feature selection patterns generally emphasize z_max and rotational features, consistent with spatial intensification near the event center.

The analysis of radar-derived features reveals that reflectivity (Z) remains the most robust indicator for identifying the core of tornado-producing storms in tropical settings. As visualized in the spatial distribution maps Figures 5(a) to (d), maximum reflectivity (Z_max) consistently exhibits concentrated high values in the immediate vicinity of the reported tornado centers across all four cases. This aligns with the physical understanding that strong convective updrafts, manifested as high reflectivity cores, are a prerequisite for sustaining these storm systems.

In contrast, dynamic features such as vorticity, delta-V, and angular momentum show significant variability between events. While localized clusters of high vorticity were distinct in the cases of January 1 and January 17, the signatures were much more dispersed or weaker in the other two cases. This inconsistency highlights the challenge of detecting non-supercell tornadoes compared to supercell counterparts, which typically possess coherent and persistent mesocyclones.

The feature importance evaluation using RF and XGB further corroborates these spatial observations. As shown in the bar charts in Figures 5(a) to (d), both models heavily prioritize reflectivity-based features (e.g., Z_max, Z_avg) as the primary predictors. However, the models diverge in their secondary selections; XGB tends to assign higher importance to rotational velocity and angular momentum derived from higher elevations, whereas RF focuses more on near-surface statistics. This divergence suggests that while thermodynamic intensity (reflectivity) is a universal predictor, the kinematic (rotational) signatures of Indonesian tornadoes are too variable to be captured uniformly by different algorithms.

3.2. Model performance and operational trade-offs

The quantitative evaluation focused on comparing the predictive performance of random forest (TDA-RF) and XGBoost (TDA-XGB) models across four tornado cases, using ROC curves and a suite of evaluation metrics including AUC, F1-score, POD, FAR, and confusion matrices Figures 6(a) to (d). Across all cases, both models achieved consistently high AUC values above 0.94, indicating excellent discrimination capability between tornado and non-tornado samples. Notably, TDA-RF tended to yield higher F1-scores in three out of four cases, suggesting a better balance between precision and recall. For instance, in

SBY_20240101, TDA-RF achieved an F1-score of 0.629 with a POD of 0.611 and FAR of 0.353, compared to TDA-XGB's lower F1-score of 0.595 despite a higher POD of 0.722.

However, a typical pattern emerged where TDA-XGB showed superior POD, capturing more tornado cases (true positives), albeit at higher FAR. For example, on 2024-01-17, TDA-XGB reached a POD of 0.723 but with a FAR of 0.647, while TDA-RF offered a more conservative performance with a lower FAR of 0.341 but also slightly lower POD (0.651). This trade-off highlights the critical balance between sensitivity and specificity in operational applications. Interestingly, in 2024-02-04, which had a much larger test sample size, both models achieved relatively balanced AUCs (0.946 for TDA-RF and 0.942 for TDA-XGB). However, TDA-XGB's FAR soared to 0.693, signaling potential over-sensitivity when exposed to large imbalanced datasets.

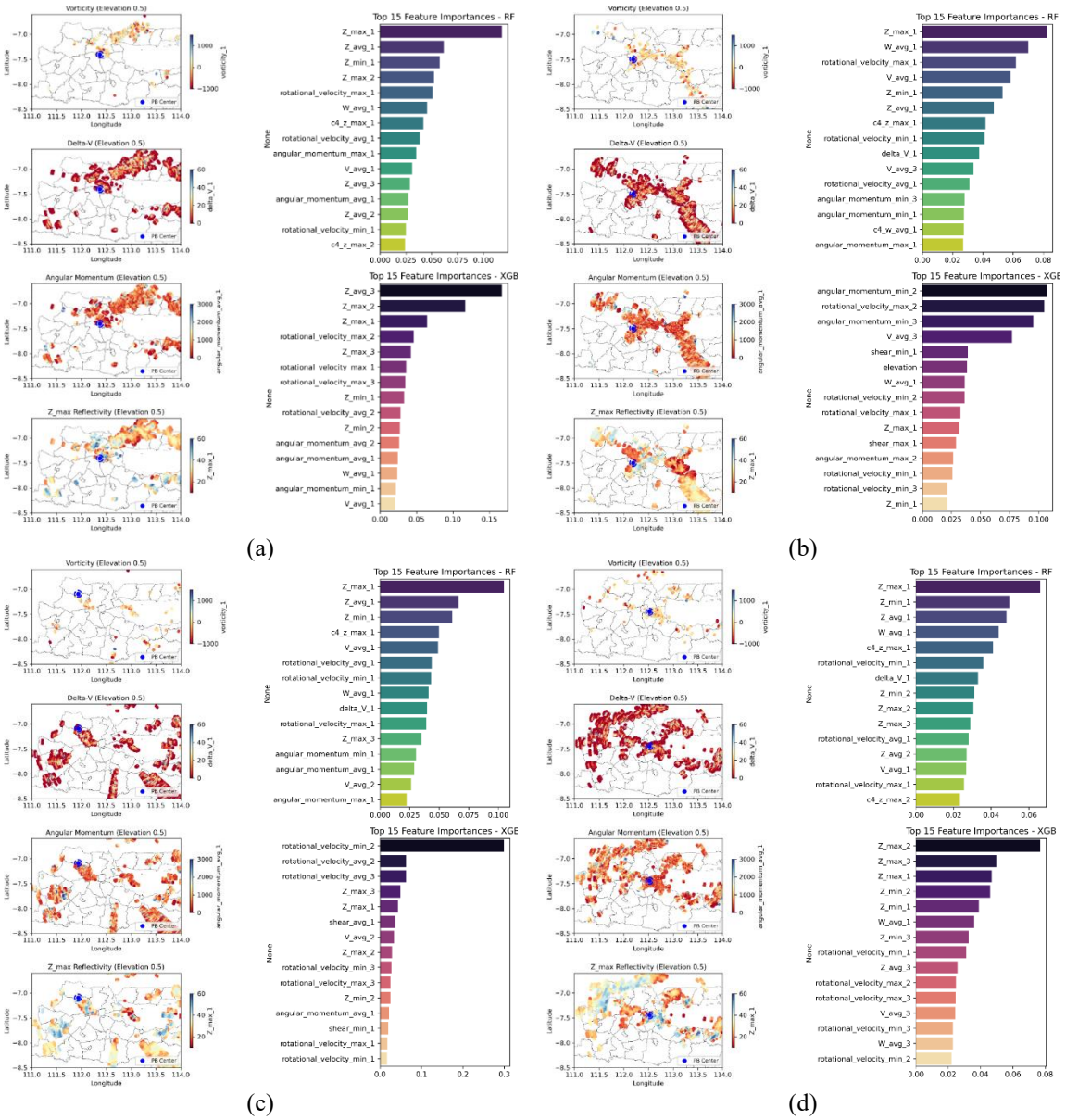


Figure 5. Spatial distribution of radar-derived features and feature importance for each tornado case: (a) 2024-01-01, (b) 2024-01-17, (c) 2024-01-26, and (d) 2024-02-02

These results underscore the challenge of achieving generalizable performance across highly localized and small-scale tornado events. While ensemble methods like RF offer stability, XGBoost may better capture rare events, albeit with a higher risk of false positives. The variability in performance across cases also

supports the need for tailored model calibration and possibly spatiotemporal stratification during model training and deployment.

From an operational perspective, this trade-off dictates distinct deployment strategies depending on the forecasting philosophy. A high-sensitivity model like TDA-XGB acts as an effective "safety net" or "early vigilance" tool, ensuring that forecasters are alerted to most potential threats. However, its high FAR implies that it cannot be used as a fully automated warning trigger, as this would likely lead to "warning fatigue" among the public. Conversely, the more conservative TDA-RF, with its lower FAR, is better suited for scenarios where maintaining public trust and reducing false positives is prioritized, though it carries the risk of missing weaker events.

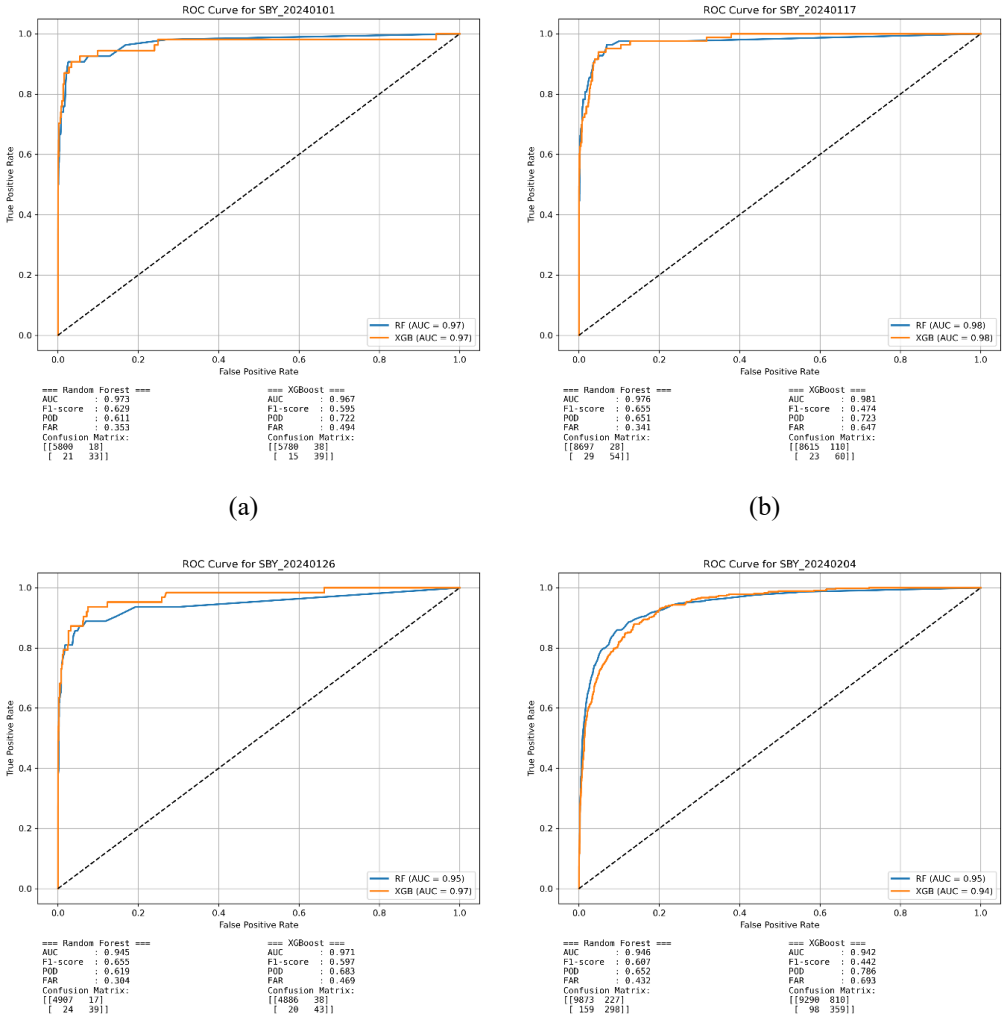


Figure 6. ROC curves and evaluation metrics for random forest (TDA-RF) and XGBoost (TDA-XGB) models using the top features of each case:(a) 2024-01-01, (b) 2024-01-17, (c) 2024-01-26, and (d) 2024-02-02

Compared to studies like Zeng *et al.* [26], which applied machine learning to detect large-scale supercell tornadoes using dual-polarization radar and environmental data, our models trained solely on single-polarization radar features demonstrated comparable AUC but generally lower F1-scores and higher FARs. This discrepancy arises from the inherently small scale, short duration, and disorganized nature of tornado events in Indonesia, which lack the coherent rotational signatures found in temperate supercells. Additionally, limited sample size and label uncertainty from field-based reports present robust model learning and generalization challenges. While RF and XGBoost showed promising results, particularly in recall, their

performance varied significantly across cases, reflecting the difficulty of building universal detectors for tropical tornadoes without broader environmental context or high-resolution multi-source data.

Both models achieved high AUC values across all cases. TDA-XGB generally offered a higher POD at the expense of an increased FAR, while TDA-RF provided more balanced performance. These results highlight the trade-off between sensitivity and specificity when detecting small-scale, non-supercell tornadoes from radar-derived features.

To further evaluate the generalization ability of the models, a LOCO evaluation was conducted in which each of the four cases was held out as the test set while the remaining three were used for training. The results of this rigorous evaluation are presented in Figure 7. LOCO results reveal a notable decline in overall detection performance compared to the within-case evaluations. While the area under the ROC curve (AUC) remained relatively stable above 0.85 for most cases both random forest (TDA-RF) and XGBoost (TDA-XGB) suffered significant drops in F1-scores and POD. TDA-XGB generally achieved higher POD across all cases, indicating stronger sensitivity to tornado occurrences; however, this came at the cost of substantially higher FAR, in some cases exceeding 70%. For instance, when SBY_20240117 was used as the unseen test set, TDA-XGB reached a POD of 0.667 but with a FAR of 0.723, highlighting the tendency of the model to over-predict positives. Meanwhile, TDA-RF showed more conservative behavior, producing lower FAR but failing to detect a significant portion of tornado instances.

These findings have significant implications for operational implementation in tropical regions like Indonesia. The current performance levels suggest that these machine learning (ML) models are best utilized as decision-support tools rather than standalone automated warning systems. Given the short lead times of tornado events, the model outputs can serve as a "first-guess" guidance field, drawing the forecaster's attention to specific storm cells that exhibit micro-scale rotational characteristics often invisible to the naked eye on standard radar displays. By integrating the ML probability maps with environmental analysis, forecasters can filter out the false alarms generated by TDA-XGB, effectively combining human expertise with machine sensitivity.

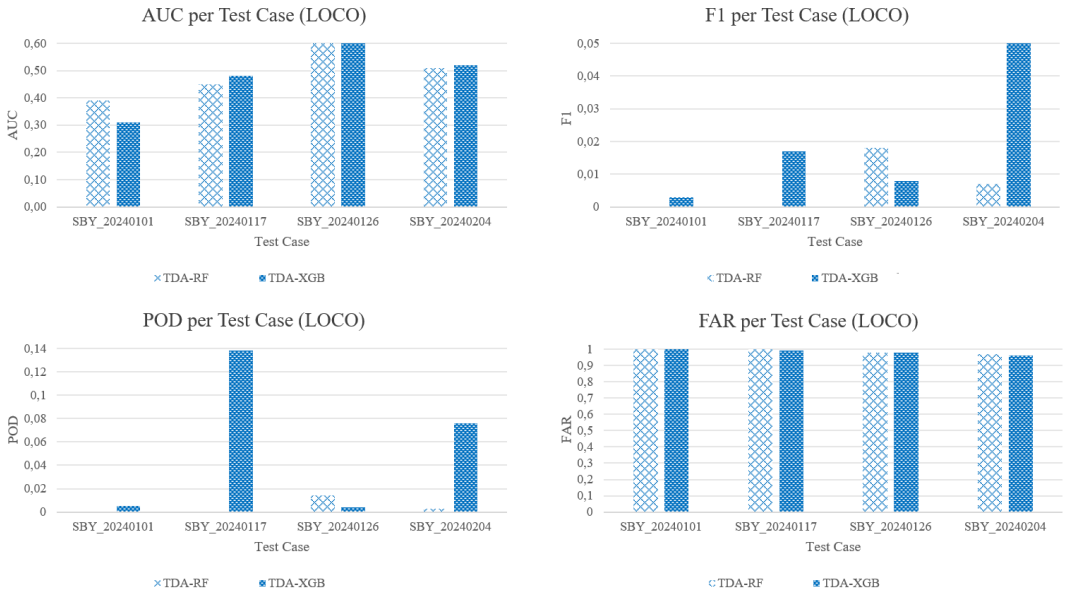


Figure 7. Model performance using LOCO evaluation for both TDA-RF and TDA-XGB

These results demonstrate the complexity of generalizing radar-based machine learning models to detect highly localized and short-lived tropical vortex phenomena. Unlike significant supercell tornadoes, which tend to exhibit consistent spatial and structural radar patterns, tornadoes in Indonesia show case-dependent variability in both structure and radar signature, limiting cross-case model performance. The LOCO evaluation highlights the need for future research on incorporating domain adaptation, temporal ensemble methods, or additional atmospheric predictors to boost generalization across operationally diverse conditions.

The LOCO evaluation results further highlight the challenges in building generalizable machine learning models for tornadoes in tropical environments like Indonesia. Unlike the within-case evaluation, where the models performed well on data drawn from the same event, LOCO results revealed performance

degradation when tested on unseen cases, particularly regarding F1-score and POD. This suggests that radar-derived feature patterns associated with tornadoes vary significantly from case to case, possibly due to differences in storm morphology, terrain influences, or local mesoscale dynamics. Compared to prior studies such as Zeng *et al.* [26] which reported stable cross-case performance using dual-polarization radar data for significant tornadoes, the findings here underscore the difficulty of applying similar methods to weak, short-lived vortices with limited spatial extent and single-polarization input. These discrepancies emphasize the need for further research on transfer learning, regional tuning, or the integration of environmental parameters (e.g., CAPE, shear) to support operational nowcasting of tornadoes in data-sparse tropical regions.

4. CONCLUSION

This study assessed the feasibility of detecting non-supercell tornadoes in Indonesia using single-polarization weather radar and machine learning. Analysis of four cases near Surabaya demonstrated that RF and XGBoost can achieve strong within-case performance (AUC > 0.94), though XGBoost produced more false alarms while RF provided more balanced results. Feature analysis highlighted the dominance of reflectivity-based predictors, with rotational features contributing less consistently across cases. The LOCO evaluation revealed major challenges in generalization, underscoring the event-specific nature of tropical tornado signatures. These results emphasize both the potential and the limitations of machine learning for radar-based tornado detection in Indonesia. To enhance operational viability as automated tornado warning triggers, future work must integrate multi-source data such as Himawari-8 satellite imagery and surface observations. Additionally, exploring hybrid spatiotemporal deep learning models (e.g., CNN-long short-term memory (LSTM)) and transfer learning from global polarimetric datasets will be essential to overcome generalization hurdles. Future progress will require larger and more diverse datasets, integration of environmental predictors, and adoption of deep learning architectures. By establishing a first baseline for radar-based ML detection in the tropics, this study contributes to advancing BMKG's early warning capabilities for localized hazardous weather in Indonesia and similar regions.

FUNDING INFORMATION

The researchers would like to express their gratitude to the Indonesia Endowment Fund for Education (LPDP) for providing financial support for this research through the 2024 research funding scheme.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kiki				✓	✓				✓	✓	✓			✓
Yonny Koesmaryono	✓					✓	✓			✓		✓	✓	
Rahmat Hidayat		✓		✓						✓		✓	✓	
Donaldi Sukma Permana		✓	✓					✓		✓	✓			
Perdinan	✓		✓		✓				✓					
Abdullah Ali			✓			✓		✓	✓					

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [initials, YK], upon reasonable request.

REFERENCES

[1] E. Yulihassstin *et al*., "High wind associated with bow echo mesovortex over Cimenyan, Indonesia," M*eteorology and Atmospheric Physics*, vol. 137, no. 6, Oct. 2025, doi: 10.1007/s00703-025-01095-7.

[2] I. M. Firdaus, T. Yamazaki, M. R. Abdillah, and E. Riawan, "Indonesia tornado database: tornado climatology of Indonesia." *EGUsphere*, May 14, 2025, doi: 10.5194/egusphere-2025-1857.

[3] K. Kiki, Y. Koesmaryono, R. Hidayat, P. Perdinan, and D. S. Permana, "Spatiotemporal Characteristics and Trend of Puting Beliung Across the Indonesian Archipelago," *Indonesian Journal of Geography*, vol. 57, no. 2, Jul. 2025, doi: 10.22146/ijg.103901.

[4] A. McGovern *et al.*, "Using artificial intelligence to improve real-time decision-making for high-impact weather," *Bulletin of the American Meteorological Society*, vol. 98, no. 10, pp. 2073–2090, Oct. 2017, doi: 10.1175/BAMS-D-16-0123.1.

[5] A. Y. Fadillah and M. R. Nurdin, "The analysis of Angin Puting Beliung risk rate by utilization of remote sensing and geographic information systems in Semarang," *International Journal for Disaster and Development Interface*, vol. 1, no. 1, Aug. 2021, doi: 10.53824/ijddi.v1i1.2.

[6] S. Lee, J. Kim, Y. N. Maharani, and E. T. P. Sunaryo, "Analysis of the risk of windstorm (angin puting beliung) in Indonesia," *Journal of the Wind Engineering Institute of Korea*, vol. 2, no. 1, pp. 21–28, 2017.

[7] S. Siswanto and S. Supari, "Identifying precursor condition for "puting beliung" event in Pangkalpinang," *Widyariset*, vol. 15, no. 3, pp. 599–610, 2012.

[8] I. Rusmala, R. Zikri, R. N. Rahman, M. I. R. Ansori, I. R. Nugraheni, and A. Ali, "Identification of small tornado event using weather radar and himawari-8 products (case study: Puting beliung event on November 22, 2018 in Jakarta)," in *The 2nd International Conference on Tropical Meteorology and Atmospheric Science*, 2021, pp. 15–20.

[9] R. D. Yudistira *et al.*, "Utilization of surface meteorological data, Himawari-8 satellite data, and radar data to analyze landspout in Sumenep, East Java, Indonesia (case study of 20 November 2017)," *IOP Conference Series: Earth and Environmental Science*, vol. 374, no. 1, Nov. 2019, doi: 10.1088/1755-1315/374/1/012038.

[10] A. Hidayat, A. Deristani, and S. Khoiriyah, "Assessing the impact and resilience of the community in the aftermath of a small-scale tornado: A field survey-based analysis," *IOP Conference Series: Earth and Environmental Science*, vol. 1317, no. 1, Mar. 2024, doi: 10.1088/1755-1315/1317/1/012010.

[11] M. S. Veillette *et al*., "A Benchmark Dataset for Tornado Detection and Prediction Using Full-Resolution Polarimetric Weather Radar Data," *Artificial Intelligence for the Earth Systems*, vol. 4, no. 1, Jan. 2025, doi: 10.1175/aies-d-24-0006.1.

[12] Q. Zeng *et al.*, "A novel tornado detection algorithm based on XGBoost," *Remote Sensing*, vol. 17, no. 1, Jan. 2025, doi: 10.3390/rs17010167.

[13] T. N. Sandmæl *et al.*, "The tornado probability algorithm: A probabilistic machine learning tornadic circulation detection algorithm," *Weather and Forecasting*, vol. 38, no. 3, pp. 445–466, Mar. 2023, doi: 10.1175/WAF-D-22-0123.1.

[14] J. Zhou, "A novel hybrid approach for tornado prediction in the United States: Kalman-convolutional BiLSTM with multi-head attention," *Prepr. arXiv.2408.02751*, Aug. 2024. [Online]. Available: http://arxiv.org/abs/2408.02751.

[15] F. Sufi, E. Alam, and M. Alsulami, "A new decision support system for analyzing factors of tornado related deaths in Bangladesh," *Sustainability*, vol. 14, no. 10, May 2022, doi: 10.3390/su14106303.

[16] J. Xie *et al*., "Multi-Task Learning for Tornado Identification Using Doppler Radar Data," *Geophysical Research Letters*, vol. 51, no. 11, Jun. 2024, doi: 10.1029/2024gl108809.

[17] Indonesian Agency of Meteorology Climatology and Geophysics "Extreme weather event information portal (PIKACU)," BMKG, Indonesia. Accessed: Jul. 21, 2025. [Online]. Available: https://pikacu.bmkg.go.id/

[18] M. Heistermann, S. Jacobi, and T. Pfaff, "Technical note: An open source library for processing weather radar data (wradlib)," *Hydrology and Earth System Sciences*, vol. 17, no. 2, pp. 863–871, Feb. 2013, doi: 10.5194/hess-17-863-2013.

[19] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: 10.1080/01431160412331269698.

[20] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random Forests," *Ensemble Machine Learning*, pp. 157–175, 2012, doi: 10.1007/978-1-4419-9326-7_5.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. June 2002, pp. 321–357, 2002, doi: 10.1613/jair.953.

[23] D. C. R. Novitasari *et al*., "Whirlwind Classification with Imbalanced Upper Air Data Handling using SMOTE Algorithm and SVM Classifier," *Journal of Physics: Conference Series*, vol. 1501, no. 1, p. 012010, Mar. 2020, doi: 10.1088/1742-6596/1501/1/012010.

[24] A. Jiménez-Valverde, "Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling," *Global Ecology and Biogeography*, vol. 21, no. 4, pp. 498–507, May 2011, doi: 10.1111/j.1466-8238.2011.00683.x.

[25] S. Narkhede, "Understanding AUC–ROC Curve," *Towards Data Science*, 2018. Accessed: Jul. 21, 2025. [Online]. Available: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

[26] Q. Zeng *et al*., "Application of Random Forest Algorithm on Tornado Detection," *Remote Sensing*, vol. 14, no. 19, p. 4909, Oct. 2022, doi: 10.3390/rs14194909.

BIOGRAPHIES OF AUTHORS

Kiki a meteorologist who graduated with a Diplome from the Academy of Meteorology and Geophysics in Jakarta, Indonesia (2005) and later completed Bachelor's in Meteorology at Bandung Institute of Technology, Indonesia (2011). She earned her Master's in Ocean-Atmosphere-Surface-Continental (OASC) interactions from École Nationale de la Météorologie, France in 2015. Currently pursuing her Doctoral studies in Applied Climatology at IPB University, Indonesia. Her focuses on extreme weather phenomena, tornadoes, tropical cyclones, and verification methods. Working at the Indonesian Agency of Meteorology, Climatology, and Geophysics since 2006 until the present at the Public Meteorology Center as an extreme weather forecaster. She can be contacted at email: 3k4siwikiki@apps.ipb.ac.id, kiki@bmkg.go.id.

Yonny Koesmaryono an expert in Biometeorology focusing on Plant Pests and Diseases, Microclimatology, and Agroclimatology. He is involved in research with the National Institute of Aeronautics and Space (LAPAN), LIPI, and IC-SEA. He completed his Bachelor of Agriculture (1980) and Master of Science in Agroclimatology (1985) at IPB University in the Agroclimatology Study Program and received his Doctorate in 1996 from Ehime University of Japan in the Agrometeorology Study Program. Since 2008 until now, he has served as Professor at Department of Geophysics and Meteorology, IPB University. He can be contacted at email: yonny@apps.ipb.ac.id.

Rahmat Hidayat a lecturer at the Department of Geophysics and Meteorology, IPB University. He earned his Bachelor's degree in Geophysics and Meteorology from the Bandung Institute of Technology (ITB) in 1999. He continued his studies at ITB, obtaining a Master's degree in Oceanography and Atmospheric Science in 2002. He further pursued his academic career in Japan, where he earned a second Master's degree in Geophysics from Tohoku University in 2006, followed by a Doctorate in Geophysics from the same institution in 2009. His areas of expertise include climatology, oceanography, and meteorology. He can be contacted at email: rahmath@apps.ipb.ac.id.

Donaldi Sukma Permana a climatologist in the Indonesian Agency of Meteorology, Climatology, and Geophysics. He holds a Ph.D. in Geological Sciences from the Ohio State University in 2015, with speciality in (past) climate. His research focuses on tropical climate variability at various timescales from diurnal to decadal over the Maritime Continent. He leads in development of regional climate model over Indonesia to support operational unit in making seasonal forecasts and also involves in CMIP6 downscaling climate change projections over Southeast Indonesia. He can be contacted at email: donaldi.permana@bmkg.go.id.

Perdinan a lecturer of Department of Geophysics and Meteorology, IPB University. He has a multidisciplinary background on Agrometeorology (B.S.), Natural Resource Economics (M.S.) and Geography and Environmental Science and Policy (Ph.D.). He was awarded Bank Central Asia Scholarship (1999-2002), Australian Development Scholarship (2007-2007), and Fulbright Presidential (2008-2011) to pursue his B.S., M.S., and Ph.D. degree, respectively. In Indonesia, he joined with the Bogor Agricultural University, as a teaching staff and researcher with specialization in Applied Climatology. He can be contacted at email: perdinan@apps.ipb.ac.id.

Abdullah Ali a meteorologist at the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG), he graduated from the State College of Meteorology, Climatology, and Geophysics (STMKG) in 2017. He later completed his Master's studies in Physics at the University of Indonesia in 2022. He has been working at BMKG since 2018, with research interests focusing on weather radar, meteorology, and artificial intelligence. He can be contacted at email: abdullah.ali@bmkg.go.id.