PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 01, No.2, Setember 2026
Pages 333-344

Check for updates

Class Imbalance-Aware SVM Optimization to Improve Daily Rainfall Event Prediction

Ary Pulung Baskoro[1*], Supartono[2], I Made Jiwa Astika[3]

[1,2,3]Department of Marine Engineering, Universitas Hang Tuah,
Jl. Arif Rahman Hakim No. 150, Surabaya 60111, East Java, Indonesia

E-mail Correspondence Author: *ary.pulung@gmail.com

Abstract

Daily rainfall events can disrupt port operations. This study evaluates class imbalance-aware optimization of an SVM-RBF rainfall-event classifier using Madura Strait sea surface temperature anomalies, the Oceanic Niño Index, and cyclic seasonal features for 2015-2024. The 2015-2022 development set (n = 2,913) comprised 1,917 non-rain (65.8%) and 996 rain (34.2%) observations, while the independent 2023-2024 test set (n = 731) comprised 518 non-rain (70.9%) and 213 rain (29.1%) observations. Data from 2015-2021 were used for initial fitting and 2022 for probability-threshold selection before refitting on the full development period. Five configurations were compared: baseline SVM-RBF, class weighting, threshold tuning, class weighting with threshold tuning, and random oversampling with threshold tuning. Under the predefined F1-first ranking criterion, the class-weighted model achieved F1-score 0.640, recall 0.676, balanced accuracy 0.748, and MCC 0.482, compared with 0.581, 0.531, 0.704, and 0.435 for the baseline. Threshold-tuned models achieved higher recall but more false alarms. Thus, class weighting is the preferred default for this dataset under the stated selection criterion.

Keywords: *Class imbalance; decision threshold; port operations; rainfall events; support vector machine.*

: https://doi.org/10.30598/parameterv5i2pp333-344

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright © 2026 The Author(s).
Research Ariticle ***Open Acces***
Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter

Submitted: May 2026
Accepted: September 2026

1. INTRODUCTION

Daily rainfall events are important operational weather variables in port environments because rainfall can disrupt loading and unloading activities, reduce occupational safety, hinder vessel movements, and affect scheduling efficiency. At Tanjung Perak Port, rainfall prediction is therefore not merely a climatological exercise but also an operational requirement for supporting rain-related risk mitigation. Reliable identification of rainfall occurrence is particularly important because a missed rain event may affect several port activities even when the total amount of rainfall is relatively small. Given the port's coastal setting adjacent to the Madura Strait, the predictor design represents three complementary scales: nearby marine thermal variability through the Madura Strait SST anomaly, large-scale ENSO variability through ONI, and recurring annual seasonality through cyclic month terms.

Machine-learning methods have increasingly been applied to weather and hydrometeorological prediction problems. Support Vector Machine (SVM), in particular, provides a nonlinear classification framework that has been used in weather prediction, short-term climate prediction, rainfall prediction, and hydrometeorological classification applications [1], [2], [3], [4], [5], [6]. In the baseline configuration considered in this study, an SVM with a Radial Basis Function (RBF) kernel uses Madura Strait sea surface temperature (SST) anomalies, the Oceanic Niño Index (ONI), and cyclic seasonal features to classify daily rainfall events at Tanjung Perak. However, rainfall and non-rainfall observations are not equally represented in the dataset. Consequently, model performance may be influenced not only by the nonlinear relationship between predictors and rainfall occurrence, but also by the imbalance between the two target classes. These studies establish the usefulness of SVM and related machine-learning approaches across weather, flood, and rainfall prediction tasks. The present study differs by holding both the predictor set and the core SVM-RBF architecture constant while isolating the effect of class-imbalance treatment through weighting, probability-threshold adjustment, and random oversampling. Class imbalance is a common challenge in binary classification because a learning algorithm may become biased toward the majority class when both classes are treated with equal misclassification costs [7], [8]. Under such conditions, overall accuracy alone may provide an incomplete assessment of model performance because high accuracy can coexist with poor recognition of the less frequent event class. Event-sensitive measures such as precision, recall, F1-score, and balanced accuracy are therefore more informative for evaluating rainfall-event detection. In addition, precision-recall behavior is particularly relevant for imbalanced data because changes in the classification decision threshold can alter the trade-off between missed events and false alarms [9].

One way to address this limitation is to optimize the classifier without necessarily expanding the predictor space. Class weighting modifies the relative penalty assigned to errors from different classes, whereas resampling modifies the class representation in the training data [7], [8]. Synthetic oversampling approaches such as SMOTE have also been widely studied in imbalanced learning [10], [11], although the present experiment uses random oversampling rather than SMOTE. Decision-threshold adjustment provides another mechanism for changing the balance between sensitivity and false alarms after model training. These strategies provide a suitable framework for determining whether the existing SVM-RBF classifier can be improved specifically through treatment of class imbalance while keeping the underlying predictors unchanged.

The novelty of this study lies in its controlled class imbalance-aware optimization of daily rainfall-event prediction for the operational environment of Tanjung Perak Port. Rather than improving performance by introducing additional predictor variables or replacing the baseline classifier, this study retains the same SST anomaly, ONI, and cyclic seasonal predictor space and systematically compares class weighting, probability-threshold tuning, and random training-data oversampling within the SVM-RBF framework. The optimization uses a chronological design with 2015-2021 for initial training, 2022 for internal validation and threshold selection, and 2023-2024 as an independent test period after refitting on the full 2015-2022 development data. This design allows performance changes to be evaluated primarily as a consequence of class-imbalance treatment while preserving temporal independence in the final assessment. Accordingly, this study aims to identify the optimization strategy that provides the most appropriate balance between rainfall-event detection and false-alarm burden for operational rainfall-risk applications at Tanjung Perak Port.

2. METHOD

a. Study Area and Dataset

This study uses daily rainfall observations at Tanjung Perak Port, Surabaya, Indonesia, together with marine and climate predictors representing the surrounding environmental conditions. Daily rainfall (curah_hujan, mm/day) was observed at BMKG's Meteorological Maritime Station Perak II (approximately 7.2052°S, 112.7354°E) from January 1, 2015, to December 31, 2024. Sea surface temperature was obtained from the Copernicus Marine SST_GLO_SST_L4_REP_OBSERVATIONS_010_024 product (analysed_sst), a gap-free daily 0.05° × 0.05° product. SST was spatially averaged over the Madura Strait domain 7.875°S-7.025°S and 112.575°E-114.525°E, and the domain-mean anomaly was retained as the local marine predictor. The Oceanic Niño Index (ONI) was obtained from the NOAA Climate Prediction Center. ONI is a three-month running mean of ERSST.v5 sea surface temperature anomalies in the Niño-3.4 region; the ONI series was temporally aligned with the corresponding daily observations. The remaining predictors are cyclic seasonal variables represented by the sine and cosine of month. No additional atmospheric predictors are introduced because the objective is to isolate the effect of class imbalance-aware optimization on the existing SVM-RBF framework.

b. Target Variable and Class Distribution

The target variable is formulated as a binary rainfall-event label. An observation is assigned to the rain class (y = 1) when daily rainfall is greater than 0 mm and to the non-rain class (y = 0) when daily rainfall equals 0 mm. The >0 mm threshold is deliberately used to represent meteorological occurrence (rain observed versus no rain) rather than an operational-impact threshold. Operationally relevant rainfall thresholds can differ by activity and consequence, whereas the present study evaluates whether the classifier detects any observed daily rainfall event. The dataset contains more non-rain than rain observations, producing a moderate class imbalance that can bias a conventional classifier toward the majority class.

c. Data Preprocessing and Chronological Splitting

Prior to model development, the daily Madura Strait SST series is expressed as an anomaly relative to the 2015-2022 monthly climatological mean and aligned temporally with ONI and rainfall observations. Seasonal periodicity is represented by $\sin(2\pi m/12)$ and $\cos(2\pi m/12)$, where m is the month number from 1 to 12. Predictor variables are

standardized using StandardScaler within the model pipeline, so the scaler is fitted only on the data used for model fitting at each stage. A chronological rather than random split is applied to preserve temporal order and reduce information leakage. In Stage 1, observations from 2015-2021 are used for model fitting and 2022 is retained as an internal validation period for probability-threshold selection. In Stage 2, the selected configuration is refitted on the complete 2015-2022 development period and evaluated on the untouched 2023-2024 independent test set. Oversampling, when used, is applied only to fitting data and never to validation or test observations. As a distributional diagnostic, monthly rainfall-event proportions in the 2015-2022 development period and 2023-2024 test period are compared using Pearson and Spearman correlations, together with seasonal event proportions for DJF, MAM, JJA, and SON.

d. Baseline Model and Optimization Strategy

The baseline classifier is a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel **[1]**, **[12]**, **[13]**, **[14]**. For n training samples, x_i denotes the p-dimensional predictor vector for sample i and $y_i \in \{-1, +1\}$ denotes its binary class label, for i = 1, …, n. The soft-margin SVM estimates a separating hyperplane by minimizing

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \left\{ \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{n} \xi_i \right\} \quad (1)$$

subject to the constraints shown below, where w is the normal vector of the separating hyperplane, b is the intercept, C > 0 is the regularization parameter, $\xi_i \geq 0$ is the slack variable for sample i, and $\varphi(x_i)$ maps the predictors into a higher-dimensional feature space.

$$y_i[\mathbf{w}^T \phi(\mathbf{x}_i) + b] \geq 1 - \xi_i, \qquad \xi_i \geq 0$$

Because relationships between environmental predictors and rainfall occurrence may be nonlinear, the RBF kernel is used **[13]**:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2), \qquad \gamma > 0 \quad (2)$$

All configurations use an RBF kernel with C = 1.0 and gamma = 'scale'. These are fixed scikit-learn settings (C = 1.0 and the variance-dependent 'scale' rule for gamma), not hyperparameters selected by optimization in this experiment; they are held constant so that performance differences can be attributed to the class-imbalance treatments. Probability estimates are enabled to support probability-based threshold optimization, and a fixed random state of 42 is used where applicable. In the baseline formulation, the same C penalizes errors from both classes. Under class imbalance, this symmetric treatment can favor the majority class **[15]**, **[16]**. Class weighting modifies the optimization by assigning class-specific penalties **[17]**, **[18]**:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \left\{ \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{n} \omega_{y_i} \xi_i \right\} \quad (3)$$

In **Equation (3)**, ω_{yi} denotes the weight assigned to the observed class y_i of sample i. For M1 and M3, class weights are generated automatically using the balanced rule $\omega_k = N/(K N_k)$, where N is the number of fitting observations, K = 2 is the number of classes, and N_k is the number of fitting observations belonging to class k; the effective class-specific penalty is $C_k = C\omega_k$. Consequently, the less frequent rain class receives a larger misclassification penalty. For M4, random minority-class oversampling with replacement is performed until the rain class equals the non-rain class in size. Because random oversampling duplicates observed minority samples, it can increase memorization and overfitting risk. This risk is limited here by applying oversampling

only to fitting partitions, while validation and independent test data remain untouched. Oversampling uses random_state = 42. Five configurations are evaluated under the same predictor space: M0, baseline SVM-RBF; M1, class-weighted SVM-RBF; M2, baseline SVM-RBF with probability-threshold tuning; M3, class-weighted SVM-RBF with threshold tuning; and M4, SVM-RBF with random oversampling and threshold tuning. For M2-M4, rainfall-class probabilities from the 2022 validation period are evaluated over thresholds from 0.05 to 0.95 at increments of 0.005. The threshold that maximizes validation F1-score is retained. After threshold selection, the corresponding model configuration is refitted on 2015-2022 and the selected threshold is applied unchanged to the 2023-2024 independent test probabilities.

e. Evaluation Metrics and Model Selection

Because overall accuracy may be dominated by the majority class in an imbalanced dataset, model comparison emphasizes event-sensitive metrics [19]. Let TP denote correctly detected rainfall events, FP non-rain observations incorrectly classified as rain, FN missed rainfall events, and TN correctly classified non-rain observations. Precision, recall, F1-score, and balanced accuracy are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2\left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right) \tag{6}$$

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{7}$$

Balanced accuracy gives equal importance to sensitivity and specificity [20]. The Area Under the Receiver Operating Characteristic Curve (AUC) is additionally reported as a threshold-independent measure of ranking ability. Matthews Correlation Coefficient (MCC) is also reported because it summarizes all four confusion-matrix cells and is robust for imbalanced binary classification; MCC = (TP×TN - FP×FN) / √[(TP+FP)(TP+FN)(TN+FP)(TN+FN)]. Accuracy is retained as a complementary measure. Candidate models are ranked first by independent-test F1-score, followed by balanced accuracy, precision, and AUC; MCC is used as an additional robustness indicator rather than a replacement for the predefined ranking rule. To quantify uncertainty in the selected M1-versus-M0 differences, 5,000 paired stratified bootstrap resamples of the independent test set (resampling within rain/non-rain strata; random seed 2026) are used to obtain 95% percentile confidence intervals for accuracy, recall, F1-score, balanced accuracy, and MCC. An exact McNemar test is additionally applied to paired classification correctness.

3. RESULTS AND DISCUSSION

a. Class Distribution in Development and Test Sets

The 2015-2022 development dataset contains 1,917 non-rain observations and 996 rain observations, corresponding to approximately 65.8% and 34.2% of the data. The 2023-2024 independent test set contains 518 non-rain and 213 rain observations, or approximately 70.9% and 29.1%, respectively. Thus, the rainfall class is consistently underrepresented in both periods. Although the imbalance is not extreme, it is sufficiently pronounced to affect model behavior, consistent with the broader class-imbalance literature [7], [8], [15], [16]. The monthly rainfall-event pattern remains

strongly consistent between the development and test periods (Pearson r = 0.940, p < 0.001; Spearman ϱ = 0.909, p < 0.001). Seasonal event proportions also preserve the same broad wet-to-dry ordering: DJF 60.9% versus 60.2%, MAM 43.0% versus 37.5%, JJA 11.2% versus 7.6%, and SON 22.1% versus 11.5% for development and test periods, respectively. The lower test-period proportion in SON indicates some temporal distribution shift despite preservation of the overall seasonal pattern.

Table 1. Class distribution in the development and test sets

Dataset Split	Non-rain Class (0)	Rain Class (1)	Total
Development (2015-2022)	1917	996	2913
Test (2023-2024)	518	213	731

b. Performance Comparison Between Baseline and Optimized Models

Table 2 compares the five SVM-RBF configurations on the independent 2023-2024 test period. The baseline model (M0) achieves accuracy of 0.777, precision of 0.642, recall of 0.531, F1-score of 0.581, AUC of 0.830, and balanced accuracy of 0.704. Class weighting produces the clearest improvement under the predefined F1-first criterion: M1 increases recall to 0.676, F1-score to 0.640, and balanced accuracy to 0.748 while maintaining accuracy at 0.778. Precision decreases moderately to 0.608. MCC also increases from 0.435 for M0 to 0.482 for M1; the remaining MCC values are 0.260 (M2), 0.339 (M3), and 0.460 (M4). This pattern is consistent with cost-sensitive SVM learning, in which a greater penalty on minority-class errors improves event sensitivity while introducing some additional false positives **[15]**, **[16]**, **[17]**, **[18]**.

Table 2. Performance comparison of baseline and optimized models on the 2023-2024 test data

Model	Accuracy	Precision	Recall	F1-score	AUC	Balanced Accuracy
M0 Baseline SVM-RBF	0.777	0.642	0.531	0.581	0.830	0.704
M1 Class Weighted	0.778	0.608	0.676	0.640	0.830	0.748
M2 Threshold Tuned	0.469	0.350	0.962	0.514	0.830	0.614
M3 Class Weighted + Tuned	0.535	0.383	0.972	0.549	0.830	0.664
M4 Random Oversampling + Tuned	0.676	0.471	0.930	0.626	0.836	0.751

The threshold-tuned models show a different error structure. M2 and M3 raise recall to 0.962 and 0.972, respectively, but precision falls to 0.350 and 0.383, and overall accuracy falls to 0.469 and 0.535. This illustrates the precision-recall trade-off characteristic of imbalanced classification **[9]**. M4, which combines random oversampling and threshold tuning, yields the highest AUC (0.836) and balanced accuracy (0.751) with recall of 0.930, but its precision (0.471) and F1-score (0.626) remain below M1. Thus, M4 is better interpreted as a high-sensitivity alternative rather than the preferred general-purpose configuration. Resampling studies similarly show that increasing minority representation can improve sensitivity without necessarily producing the best overall error balance **[10]**, **[11]**.

c. Threshold Behavior and Precision–Recall Trade-off

Probability-threshold optimization was conducted only on the 2022 validation data. For M2, the highest validation F1-score occurs at a threshold of 0.180, with precision of 0.633, recall of 0.704, and F1-score of 0.667. M3 selects 0.100, producing precision of 0.542, recall of 0.914, and F1-score of 0.680. M4 selects 0.225, with precision of 0.594, recall of 0.802, and F1-score of 0.682. However, these validation gains do not transfer equally to the independent test period: M2 and M3 retain very high recall but experience substantial losses in precision and F1-score. The result indicates that a probability threshold optimized in one period may not remain equally suitable under later conditions, so operational thresholds may require periodic reassessment rather than being treated as fixed properties of the classifier. The seasonal analysis above, including the lower test-period event proportion in SON, illustrates how changes in event frequency can alter probability calibration; changes in large-scale climate state represented by ONI may likewise contribute to temporal shifts in score distributions. Accordingly, threshold recalibration should use recent data that are representative of the current seasonal and climate regime. This interpretation is consistent with the importance of precision-recall analysis under class imbalance **[9]**.

d. Best Model Selection and Confusion Matrix Comparison

Using the predefined ranking procedure, M1 is selected as the preferred optimized configuration for this dataset because it achieves the highest independent-test F1-score of 0.640. Although M4 records slightly higher balanced accuracy and AUC, its lower precision and F1-score indicate a less favorable balance under the stated F1-first criterion. The confusion matrices show that M0 detects 113 rainfall events and misses 100, whereas M1 detects 144 and misses 69. Thus, class weighting produces 31 additional correctly detected rainfall events. False positives increase from 63 to 93, corresponding to 30 additional false alarms.

Table 3. Confusion matrix comparison between the baseline and selected optimized model

Model	Actual 0, Predicted 0	Actual 0, Predicted 1	Actual 1, Predicted 0	Actual 1, Predicted 1
Baseline SVM-RBF	455	63	100	113
M1 Class Weighted	425	93	69	144

Without an explicit cost-loss analysis, these 30 additional false alarms cannot be categorically described as operationally acceptable. M1 is therefore interpreted as statistically preferred under the predefined F1-first evaluation criterion, not as demonstrated operationally superior. The trade-off may be preferable when the operational consequence of missing rainfall is greater than the cost of an additional false alarm, but quantifying that preference requires a dedicated operational cost-loss framework.

e. Final Comparison and In-depth Discussion

Table 4. Final comparison between the baseline and selected optimized model

Metric	Baseline SVM-RBF	M1 Class Weighted	Change
Accuracy	0.777	0.778	+0.001
Precision	0.642	0.608	-0.034
Recall	0.531	0.676	+0.146
F1-score	0.581	0.640	+0.059
AUC	0.830	0.830	+0.000
Balanced Accuracy	0.704	0.748	+0.044
MCC	0.435	0.482	+0.047

Compared with the baseline, M1 increases recall by 0.146, F1-score by 0.059, and balanced accuracy by 0.044 while leaving overall accuracy effectively unchanged. Precision decreases by 0.034, whereas AUC is nearly unchanged at the reported three-decimal precision. These results support previous findings that aggregate accuracy alone can mask minority-class performance and that class-sensitive learning can improve minority-event recognition **[7]**, **[8]**, **[15]**, **[16]**. In particular, the behavior of M1 is consistent with weighted-SVM approaches that increase the cost of minority-class errors and thereby improve sensitivity, although some loss of precision may occur **[16]**, **[17]**, **[18]**. Paired uncertainty analysis supports this interpretation. In 5,000 paired stratified bootstrap resamples, the M1-M0 differences were +0.146 for recall (95% CI: 0.099 to 0.197), +0.059 for F1-score (0.021 to 0.100), and +0.044 for balanced accuracy (0.018 to 0.070), with all three confidence intervals excluding zero. In contrast, the accuracy difference was +0.001 (95% CI: -0.019 to 0.021) and the MCC difference was +0.047 (-0.002 to 0.098). The exact McNemar test on paired correctness outcomes found 31 cases correct only for M0 and 32 correct only for M1 ($p = 1.000$), providing no evidence of a change in overall classification correctness. The measurable advantage of M1 is therefore concentrated in event-sensitive performance rather than aggregate accuracy. The comparison among M1-M4 further shows that imbalance treatments are not interchangeable. Random oversampling changes the training-sample composition, class weighting changes the SVM loss function, and threshold tuning changes the probability operating point after fitting. Here, class weighting yields the highest F1-score and a relatively stable error balance, whereas threshold tuning prioritizes sensitivity much more aggressively. This extends previous rainfall, weather, and hydrometeorological machine-learning applications **[2]**, **[3]**, **[4]**, **[5]**, **[6]** by demonstrating that meaningful improvement can arise from explicit treatment of class distribution even when the predictor set and core algorithm are held constant.

The nearly unchanged AUC between M0 and M1 is also informative. Because AUC evaluates ranking ability across thresholds, similar AUC values indicate that class weighting does not fundamentally alter the discriminative information contained in the retained predictors. Its main benefit appears in the classification operating point, where the available information is converted into a better balance between rainfall detection and non-rain classification. This explains why recall, F1-score, and balanced accuracy improve even though AUC changes little. Several limitations remain. The study intentionally retains a limited predictor set and does not test whether additional atmospheric variables would provide further improvement. All experiments are conducted for a single port environment, so transferability to other locations remains untested. The deterioration of the threshold-tuned configurations from the 2022

validation period to the 2023-2024 test period indicates temporal sensitivity in selected probability thresholds, and the seasonal comparison shows that event frequency can shift even when the broad annual pattern is preserved. Random oversampling also duplicates minority samples and may increase overfitting risk, although this study confines oversampling to fitting data and evaluates it on untouched validation and test periods. Finally, the study evaluates statistical classification performance but does not quantify the economic or operational costs of false alarms and missed rainfall events. Future work should therefore consider additional atmospheric predictors, external multi-site validation, season- and climate-aware threshold recalibration, and an explicit cost-loss decision framework. Overall, class imbalance-aware optimization improves daily rainfall-event detection for the present Tanjung Perak dataset within the existing SVM-RBF framework. Under the predefined F1-first selection criterion, class weighting is the preferred default configuration because it achieves the highest F1-score and MCC while substantially improving recall and balanced accuracy without materially changing overall accuracy. Threshold-tuned and random-oversampling configurations remain useful high-sensitivity alternatives when the analytical priority is minimizing missed rainfall events, but operational deployment requires explicit cost-loss assessment and local recalibration.

4. CONCLUSION

For the present Tanjung Perak dataset, class imbalance-aware optimization improves daily rainfall-event prediction without expanding the original predictor set. Using Madura Strait SST anomalies, the Oceanic Niño Index, and cyclic seasonal features, five SVM-RBF configurations were evaluated under a chronological development and independent-testing framework. Under the predefined F1-first ranking criterion, the class-weighted SVM-RBF model (M1) achieved the highest F1-score (0.640) and MCC (0.482), with recall of 0.676, balanced accuracy of 0.748, and accuracy of 0.778 on the 2023-2024 independent test period. Relative to the baseline, recall increased by 0.146 and F1-score by 0.059 while overall accuracy remained nearly unchanged. The results also show that imbalance treatments produce different operating behaviors. Probability-threshold tuning greatly increases rainfall-event sensitivity but generates substantially more false alarms, while random oversampling combined with threshold tuning achieves high recall and the highest balanced accuracy but lower precision and F1-score than M1. Bootstrap confidence intervals support improvements in recall, F1-score, and balanced accuracy for M1 relative to the baseline, whereas the accuracy confidence interval includes zero and the exact McNemar test shows no difference in overall paired correctness. Consequently, M1 is preferred statistically under the stated criterion rather than claimed to be operationally superior. These findings are specific to the single-site Tanjung Perak dataset and should not be generalized to other ports without external validation. Within this setting, performance can be improved not only by adding predictors or replacing the classifier, but also by adjusting the learning objective to account for unequal class frequencies. Future research should evaluate additional atmospheric predictors, multi-location transferability, season- and climate-aware recalibration of probability thresholds, and cost-sensitive decision frameworks to strengthen operational applicability.

Acknowledgments

The authors wish to express their gratitude to the Meteorological Maritime Station Perak II for providing the daily rainfall data used in this study. Appreciation is also extended to the Juanda Aviation Meteorology Station, Sidoarjo, for institutional support and to Universitas Hang Tuah, Surabaya, for academic facilities and scholarly guidance. The authors also thank colleagues and mentors whose discussions contributed to the development of this work.

Funding Information

Authors state no funding involved.

Author Contributions Statement

Ary Pulung Baskoro: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - original draft. Supartono: Supervision, Conceptualization, Validation, Writing - review & editing. I Made Jiwa Astika: Supervision, Validation, Writing - review & editing. All authors discussed the results and contributed to the final manuscript.

Conflict of Interest Statement

Authors state no conflict of interest.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[2] G. Gunawan, W. Andriani, and A. A. Akbar, "Application of machine learning for short-term climate prediction in Indonesia," *Jurnal Mantik*, vol. 8, no. 1, pp. 828–837, 2024, doi: 10.35335/mantik.v8i1.5215.

[3] M. F. R. Mahendra, N. L. Azizah, and Sumarno, "Implementasi machine learning untuk memprediksi cuaca menggunakan Support Vector Machine," *Jurnal Ilmiah Komputasi*, vol. 23, no. 1, pp. 45–50, 2024.

[4] S. E. Purwati and Y. Pristyanto, "Model Random Forest and Support Vector Machine for flood classification in Indonesia," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 4, pp. 2261–2268, 2024.

[5] W. P. Putra, R. Robiyanto, A. Puspaningrum, A. Rifai, R. Burjulius, and A. Sumarudin, "Energy-efficient rainfall prediction using Support Vector Machine on edge AI platforms," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 3–2, pp. 1686–1692, 2024, doi: 10.62527/joiv.8.3-2.3158.

[6] O. A. Wani *et al.*, "Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas," *Sci. Rep.*, vol. 14, p. 27876, 2024, doi: 10.1038/s41598-024-77687-x.

[7] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, p. Article 31, 2016, doi: 10.1145/2907070.

[8] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[9] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot is more informative than the ROC Plot when evaluating binary classifiers on imbalanced datasets," *PLoS One*, vol. 10, no. 3, p. e0118432, 2015.

[10] 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, pp. 321–357, 2002.

[11] A. Fernández, S. Garcia, F. Herrera, and N. V Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, 2018.
[12] V. N. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.
[13] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
[14] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[15] H. Guo, Y. Li, J. Shang, M. Gu, Y. Huang, and B. Gong, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, 2017, doi: 10.1016/j.eswa.2016.12.035.
[16] R. Batuwita and V. Palade, "Class imbalance learning methods for support vector machines," in *Imbalanced Learning: Foundations, Algorithms, and Applications*, H. He and Y. Ma, Eds., Hoboken, NJ, USA: Wiley, 2013, pp. 83–99.
[17] Y. Tang, Y. Q. Zhang, N. V Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 281–288, 2009.
[18] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review," *International Journal of Advances in Soft Computing and its Applications*, vol. 7, no. 3, pp. 176–204, 2015.
[19] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020.
[20] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proceedings of the 20th International Conference on Pattern Recognition*, Istanbul, Turkey, 2010, pp. 3121–3124.