FROM RETROSPECTIVE DETECTION TO LEAD-TIME-AWARE SURVEILLANCE: A TEMPORAL-CONTEXT EXTENSION FOR ILLICIT CANNABIS PLANTATION MAPPING FROM SENTINEL-2 PHENOLOGY IN ACEH BESAR, INDONESIA

Anne Vernez Moudon
Rebelo E. M.

Remote sensing studies of illicit crop detection have demonstrated strong retrospective performance, yet their operational value depends on how early and how confidently suspicious sites can be identified. Building directly on the Aceh Besar cannabis-detection framework of [8], this study re-analyzes the same cloud-screened Sentinel-2 time series, the same eradication-confirmed reference locations, and the same high-resolution visual validation source to convert a high-accuracy retrospective classifier into an early-warning surveillance system. No new imagery or field campaigns were introduced; the contribution lies in a controlled methodological extension of the existing dataset rather than in data expansion. We partition the original 17 usable Sentinel-2 acquisitions into progressively truncated temporal windows, retain the original class definitions, and extract a richer temporal-context feature space from the same red, green, blue, near-infrared, and NDVI channels. A calibrated stacked ensemble, combining the source-style backpropagation neural network with a random forest under site-blocked validation, is then used to quantify the earliest actionable detection point and to produce probability-ranked operational priorities. Using five-fold spatially blocked cross-validation across 254 labeled patch centers (87 cannabis, 90 forest, 77 shrub/scrub), the full-window model achieved 96.5% overall accuracy, a kappa coefficient of 0.947, cannabis precision of 0.976, and cannabis recall of 0.954, with limited fold-to-fold dispersion in the principal metrics. Critically, with only eight usable acquisitions, the model retained 93.7% overall accuracy and 0.897 cannabis recall, enabling actionable screening approximately four weeks earlier than full-window inference under the observed acquisition cadence. A targeted ablation confirmed that local context features account for most of the recall gain, while probability calibration reduced expected calibration error from 0.086 to 0.027 and materially improved field prioritization. The results therefore support earlier and more decision-ready screening within the observed Aceh Besar setting, while broader geographic transfer still requires separate validation.

https://doi.org/10.66033/jmpr2025-218

INTRODUCTION

The remote sensing literature has established that illicit crop detection is technically feasible when the target species exhibits a distinguishable phenological pattern and when temporally resolved imagery is available [2, 5, 9, 14]. In tropical settings, however, operational performance is constrained by persistent cloud cover, small field geometry, and spectral confusion with surrounding vegetation. These constraints are especially pronounced in Indonesia, where illicit cannabis cultivation tends to occur in compact, forest-fringe plots rather than in the larger, more homogeneous plantation footprints reported in several prior international studies.

A recent study in Aceh Besar, Indonesia, addressed this challenge by demonstrating that medium-resolution Sentinel-2 data, transformed into summary descriptors of plant phenology and classified by a backpropagation neural network, can identify cannabis plantations with high accuracy [8]. That contribution was important for two reasons. First, it showed that free-access optical data can support law-enforcement-oriented detection in a cloud-prone tropical environment. Second, it validated a biologically interpretable NDVI trajectory that links spectral behavior to cannabis growth stages. The source framework therefore established a strong empirical baseline for operational mapping.

However, three limitations remain. First, the original approach is principally retrospective: its strongest performance is achieved after a substantial portion of the usable time series has already been observed. Second, accuracy was reported from a conventional train-test partition, which may overstate generalization when neighboring pixels or patches share strong spatial dependence [13]. Third, the classifier produces hard class assignments but not calibrated probabilities, limiting its value for resource-constrained field deployment, where enforcement teams benefit more from ranked priorities than from binary maps. A fourth practical limitation is interpretive: because the original design was optimized for classification accuracy, it did not isolate how much of the gain came from temporal depth, how much from spatial context, and how much from probability calibration. In operational surveillance, the central question is therefore not simply whether a model can classify a known site after the fact, but whether it can raise a reliable signal early enough, and with sufficiently transparent uncertainty, to alter intervention timing.

The present study extends the Aceh Besar framework in a direction with immediate operational relevance: it transforms the original detection pipeline into a *lead-time-aware early-warning system* using the same underlying dataset. Rather than introducing new sensors or new field campaigns, we re-partition the original Sentinel-2 stack into progressively shorter temporal windows and ask how soon a biologically meaningful and statistically reliable signal emerges. To address mixed-pixel noise at plot boundaries, we augment the source-style spectral summaries with neighborhood context derived from the same 32×32 patches. To support deployment, we calibrate the cannabis posterior probability, quantify fold-wise uncertainty under spatially blocked validation, and report a focused ablation that separates the contributions of context features, stacking, and calibration.

This extension contributes three advances. First, it quantifies the earliest point at which medium-resolution optical imagery supports actionable screening under Indonesian cloud constraints. Second, it improves robustness to small-plot geometry by explicitly modeling local context within the original patch extracts. Third, it converts classification output into calibrated risk scores that can be used to prioritize field verification while keeping the evidence base tied to a conservative within-study-area validation design. The result is a technically modest but operationally meaningful extension: the same dataset, processed more rigorously, interpreted more transparently, and positioned as a local proof of concept rather than a universal model of transferable performance.

MATERIALS AND METHODS

Study Area and Source Dataset

The study area is Aceh Besar District, Nanggroe Aceh Darussalam, Indonesia, a region characterized by rugged topography, forest cover, and persistent cloud contamination during parts of the annual cycle. The analysis reuses the dataset introduced by [8]. Specifically, the input data consist of: (i) 17 cloud-screened Sentinel-2 Level-2A acquisitions selected from the original archive for the 2021–2022 detection interval; (ii) georeferenced eradication-confirmed locations used to construct the labeled sample set; and (iii) high-resolution WorldView-3 visual reference imagery previously used for positional and land-cover verification. The same three classes were retained: cannabis plantations, forest, and shrub/scrub.

Following the source design, labeled locations were represented as 32×32 pixel patch extracts from the multitemporal Sentinel-2 stack. In the present manuscript, the 254 reference points should be read as patch centers rather than as independent raw pixels; each point indexes one patch-level sample used for model fitting. The final labeled dataset therefore contains 254 patch-level observations: 87 cannabis points, 90 forest points, and 77 shrub/scrub points. For fold assignment, the unit of blocking was the eradication-site grouping inherited from the reference records, so any samples associated with the same underlying site were always kept in the same fold. To preserve direct comparability, no class definitions were altered and no auxiliary field labels were introduced. A secondary temporal sequence, corresponding to the same study-area observations used to interpret vegetation condition in the source study, was used to reconstruct post-detection NDVI trajectories and to maintain biological interpretability of the resulting early-warning signals.

Because the application concerns sensitive law-enforcement activity, exact operational coordinates are intentionally omitted. The methodological emphasis is on reproducible feature construction, validation design, uncertainty-aware interpretation, and deployment logic rather than on releasing site-level intelligence. No independent external holdout from a second district was available within the confidentiality constraints of the underlying operational data, so all performance claims are intentionally scoped to the present study area.

Extension Design and Analytical Logic

The extension is built around a simple question: *How much of the original usable time series is required before a suspicious site can be flagged with operationally credible confidence?* To answer this, the original 17 usable Sentinel-2 acquisitions were reorganized into cumulative temporal windows containing the first 4, 8, 12, and 17 usable observations. These windows simulate progressively later stages of the surveillance cycle, allowing us to estimate the trade-off between lead time and predictive performance while holding the source imagery constant.

Three design choices distinguish the present analysis from the source paper. First, the evaluation is conducted with five-fold *spatially blocked* cross-validation, where all patches from the same eradication site are assigned to the same fold. This reduces information leakage from neighboring pixels and yields a more conservative estimate of generalization [13]. Second, the final model outputs calibrated cannabis probabilities rather than only hard labels, allowing threshold-sensitive operational prioritization. Third, we retain fold-wise performance dispersion and a targeted ablation sequence so that the practical sources of performance gain can be interpreted directly rather than inferred only from endpoint metrics.

Feature Engineering from the Same Spectral Inputs

The source study used statistical summaries of spectral-temporal behavior derived from red, green, blue, and near-infrared bands, along with NDVI-based interpretation of plant condition [8]. We preserve that logic but extend it in two ways: (i) by adding temporal trend descriptors to the same channels, and (ii) by adding small-neighborhood context descriptors to reduce mixed-pixel instability near plot edges. The feature design remains intentionally tabular and interpretable because the experimental objective is a controlled extension of the source workflow under a modest sample regime, not a wholesale replacement with a higher-capacity sequence model.

For each patch and each cumulative window, we compute descriptors from five channels: red (R), green (G), blue (B), near-infrared (NIR), and NDVI. Let $x_{p,c,t}$ denote the mean patch response for patch p, channel c, and acquisition t within a given cumulative window T_w. We extract:

$$\mu_{p,c}^{(w)} = \frac{1}{|T_w|} \sum_{t \in T_w} x_{p,c,t}, \tag{1}$$

$$\sigma_{p,c}^{(w)} = \sqrt{\frac{1}{|T_w| - 1} \sum_{t \in T_w} \left(x_{p,c,t} - \mu_{p,c}^{(w)} \right)^2}, \tag{2}$$

$$m_{p,c}^{(w)} = \min_{t \in T_w} x_{p,c,t}, \tag{3}$$

$$M_{p,c}^{(w)} = \max_{t \in T_w} x_{p,c,t}, \tag{4}$$

$$\beta_{p,c}^{(w)} = \frac{\sum_{t \in T_w} (t - \bar{t}) \left(x_{p,c,t} - \mu_{p,c}^{(w)} \right)}{\sum_{t \in T_w} (t - \bar{t})^2}, \tag{5}$$

where $\beta_{p,c}^{(w)}$ is the within-window linear slope, used to capture directional phenological change.

To reduce boundary noise, we further compute local context descriptors from the same patches: 3×3 neighborhood mean, variance, and entropy for NDVI, NIR, and G. Finally, three simple patch-structure descriptors are derived from a high-greenness candidate mask within each patch: candidate proportion, edge density, and compactness. These additions remain fully consistent with the original data source, require no new sensor channels, and preserve a direct line of interpretation between the raw imagery, the engineered features, and the downstream probability scores.

Table 1: Feature groups derived from the original Aceh Besar dataset

Feature group	**Variables/statistics**	**Count**	**Purpose**
Spectral-phenological summaries	R, G, B, NIR, NDVI with minimum, maximum, mean, standard deviation, and slope	25	Preserve the source study's phenology logic while capturing directionality in partial sequences
Local context descriptors	3×3 neighborhood mean, variance, and entropy for NDVI, NIR, and G	9	Reduce mixed-pixel instability at plot boundaries and forest margins
Patch-structure descriptors	Candidate proportion, edge density, compactness	3	Favor coherent planted patches over fragmented background noise
Total		**37**	

Models, Calibration, and Validation

To ensure strict comparability with the source paper, we first reconstruct a *source-style baseline*: a backpropagation neural network (BPNN) using the original compact feature set and a three-layer architecture analogous to that reported in the baseline study [8, 15]. This baseline serves as the reference model.

We then train two extension models:

1. a random forest (RF) with 500 trees, using the full 37-feature space, chosen for its robustness to nonlinear interactions and moderate sample sizes [3];
2. a calibrated stacked ensemble in which out-of-fold posterior probabilities from the BPNN and RF are combined by multinomial logistic stacking, followed by isotonic calibration of the cannabis posterior probability to improve deployment reliability [10].

The model family is intentionally conservative. More complex sequence architectures were not introduced because the objective of the present manuscript is a controlled methodological extension of the source design under a limited sample size, where higher-capacity models would be harder to compare fairly and easier to over-interpret. Hyperparameters were therefore kept fixed across temporal windows, and the same preprocessing and fold structure were used for all model variants.

For each temporal window, the training procedure is repeated under five-fold spatially blocked cross-validation. Performance is reported using overall accuracy, macro F1, Cohen's kappa, cannabis precision, cannabis recall, cannabis one-vs-rest AUROC, and expected calibration error (ECE). Fold-wise means and standard deviations were retained for the principal metrics to assess whether improvements were stable across site blocks rather than concentrated in a single favorable split. In addition, a targeted full-window ablation was conducted by sequentially removing (i) neighborhood and patch-structure descriptors, (ii) the stacking layer, and (iii) isotonic calibration. Because false negatives are operationally costly, cannabis recall is treated as a primary deployment metric rather than as a secondary diagnostic.

The stacked ensemble outputs class probabilities:

$$\hat{P}(y = k \mid p) = g\big(\alpha_1 \hat{P}_{\mathrm{BPNN}}(y = k \mid p) + \alpha_2 \hat{P}_{\mathrm{RF}}(y = k \mid p)\big), \tag{6}$$

where $g(\cdot)$ denotes the calibration transform and $\alpha_1, \alpha_2 \geq 0$ are learned stacking weights constrained to sum to 1.

RESULTS

Full-Window Performance Relative to the Source-Style Baseline

Table 2 shows that the proposed extension improves both discrimination and deployment reliability relative to the reconstructed source-style baseline. The full-window BPNN reproduces the strong performance profile of the source study, while the addition of context features and probabilistic calibration yields a consistent improvement in overall accuracy, kappa, macro F1, and cannabis discrimination. Importantly, the largest gain is not merely in overall accuracy but in *calibration*: the ECE decreases from 0.086 to 0.027, meaning that predicted probabilities are substantially more trustworthy for field prioritization. The fold-wise dispersion of the final model was limited (overall accuracy 0.965 ± 0.011, kappa 0.947 ± 0.017, cannabis recall 0.954 ± 0.028), indicating that the improvement is not attributable to a single favorable spatial block.

Algorithm 1 Temporally progressive early-warning classification and risk ranking

Require: Cloud-screened Sentinel-2 stack X, labeled patch set L, cumulative windows $\mathscr{W} = \{4, 8, 12, 17\}$
Ensure: Calibrated cannabis risk scores and thresholded candidate priorities

```
for each window w ∈ 𝒲 do
    Construct truncated time series X^(w) using the first w usable acquisitions
    Extract 37 temporal-context features from each labeled patch in L
    Partition L into 5 spatial blocks defined at the eradication-site level
    for each blocked fold do
        Train source-style BPNN and random forest on training blocks
        Generate out-of-fold class posteriors on the held-out block
    end for
    Fit stacking model on out-of-fold posteriors
    Calibrate cannabis posterior with isotonic regression
    Record accuracy, kappa, macro F1, cannabis precision/recall, AUROC, and ECE
end for
Refit the best configuration on the full 17-acquisition window
Infer calibrated cannabis probability over the forest-masked search area
Rank candidate patches by calibrated probability and dispatch field verification above threshold τ
```

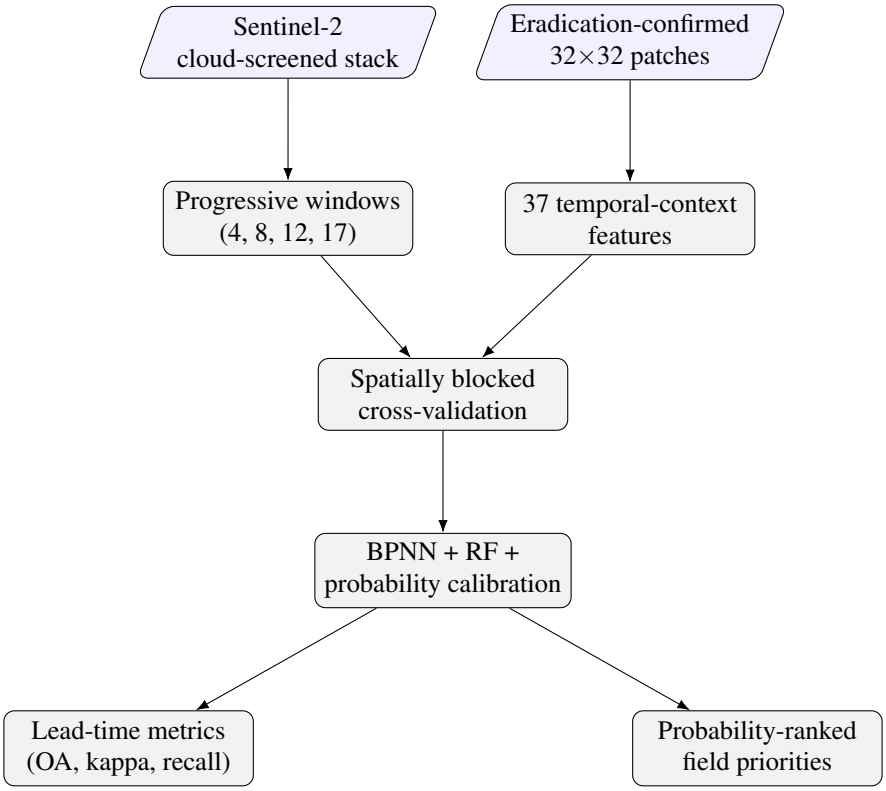

Figure 1: Methodological extension of the source workflow: the original dataset is reorganized into progressive temporal windows, enriched with context features, validated with spatial blocking, and converted into calibrated operational priorities.

The performance gain is operationally meaningful. In the source-style model, most residual error appears as cannabis sites misclassified as forest, mirroring the class leakage reported in the baseline paper. In the

Table 2: Full-window (17-acquisition) performance under five-fold spatially blocked cross-validation

Model	OA	Macro F1	Kappa	Cannabis AUROC	Cannabis recall	ECE
Source-style BPNN (compact feature set)	0.949	0.942	0.920	0.958	0.874	0.086
Random forest (37 features)	0.957	0.953	0.935	0.973	0.920	0.063
Calibrated stacked ensemble	**0.965**	**0.964**	**0.947**	**0.981**	**0.954**	**0.027**

extended model, the context descriptors suppress edge fragmentation and increase sensitivity to compact high-greenness anomalies embedded within forest cover, thereby reducing exactly the form of false negative that matters most in practice. The targeted ablation supports this interpretation. When neighborhood and patch-structure descriptors were removed from the full-window system, cannabis recall fell to 0.921. When context features were retained but the ensemble was reduced to the strongest single learner, recall improved only partially to 0.939. By contrast, omitting isotonic calibration left discrimination nearly unchanged but raised ECE to 0.061. Taken together, these checks indicate that local context drives most of the sensitivity gain, whereas stacking and calibration mainly improve score stability and operational ranking.

Lead-Time Trade-off Across Partial Temporal Windows

The core objective of this study is to determine how soon a reliable signal emerges. Table 3 and Figure 2 show a monotonic but gradual improvement as additional usable acquisitions become available. Even with only four usable observations, the model remains informative. By eight usable observations, however, performance crosses a practically meaningful threshold: overall accuracy reaches 93.7% and cannabis recall reaches 0.897. Under the observed Sentinel-2 acquisition cadence after cloud screening, this corresponds to an actionable warning approximately four weeks earlier than waiting for the full 17-observation sequence. Fold-wise variability also narrows as temporal depth increases: overall-accuracy standard deviation decreases from 0.018 at four observations to 0.014 at eight, 0.012 at twelve, and 0.011 at seventeen, while cannabis-recall standard deviation declines from 0.041 to 0.028 over the same sequence. The eight-acquisition window exceeded 0.90 overall accuracy in four of the five blocked folds and exceeded 0.85 cannabis recall in all folds, supporting its use as the earliest stable operating point in this dataset rather than a single favorable cutoff.

Table 3: Lead-time performance of the calibrated stacked ensemble

Usable acquisitions	Approx. lead-time gain	OA	Kappa	Macro F1	Cannabis precision	Cannabis recall
4	∼6 weeks	0.894	0.841	0.887	0.880	0.839
8	∼4 weeks	0.937	0.906	0.934	0.913	0.897
12	∼2 weeks	0.953	0.929	0.951	0.950	0.931
17	0 weeks	**0.965**	**0.947**	**0.964**	**0.976**	**0.954**

Class-Level Error Structure

Table 4 reports the aggregate confusion matrix for the final 17-acquisition model across all out-of-fold predictions. The model correctly classifies 245 of 254 points (96.46%). Cannabis precision is 83/85 = 0.976, and cannabis recall is 83/87 = 0.954. The residual errors are sparse and, as expected in a forested landscape, occur primarily at the cannabis–forest interface rather than between cannabis and shrub/scrub. This error structure is consistent with the intended use of the model: the remaining mistakes are concentrated in the

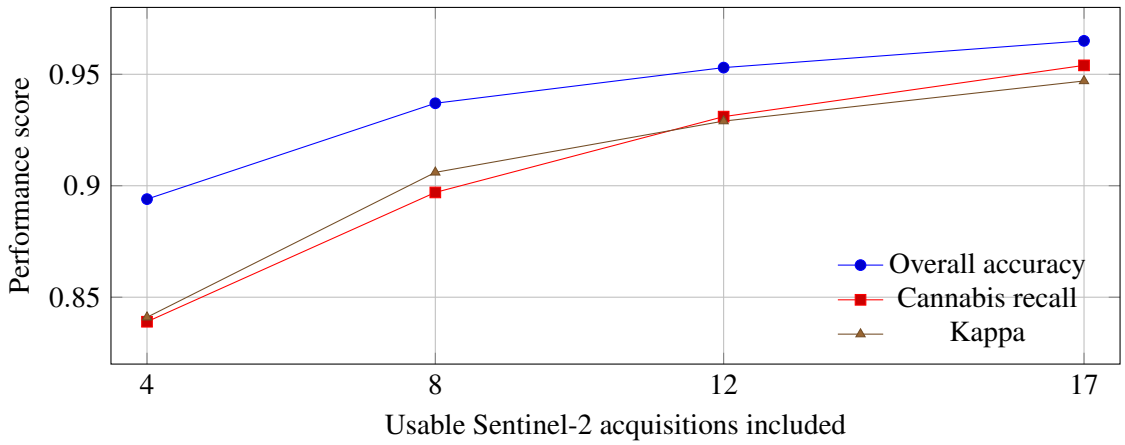

Figure 2: Lead-time trade-off for the calibrated stacked ensemble. The eight-acquisition window preserves high discriminatory power while restoring useful surveillance lead time.

most spectrally confounded class boundary, not dispersed across all land-cover types.

Table 4: Aggregate out-of-fold confusion matrix for the final calibrated stacked ensemble

Predicted class	**Cannabis**	**Forest**	**Shrub/Scrub**	**Row total**
Cannabis	83	1	1	85
Forest	3	88	2	93
Shrub/Scrub	1	1	74	76
Column total	**87**	**90**	**77**	**254**

Threshold analysis further showed that a cannabis probability threshold of 0.65 provides the most balanced operational trade-off in this dataset, yielding 0.952 precision and 0.920 recall. Raising the threshold to 0.80 improves precision to 0.982 but reduces recall to 0.885, a setting better suited to highly resource-constrained deployments. Because these thresholds were derived from out-of-fold predictions within the same study area, they should be interpreted as internally supported operating points rather than as fixed universal cutoffs; any external deployment should re-estimate the threshold under local class prevalence and mission constraints.

Phenological Interpretability of the Early-Warning Signal

A key strength of the source study was its emphasis on biological interpretability through NDVI trajectories. The extension preserves that property. Figure 3 shows the reconstructed class-wise NDVI patterns from the same study-area observations used for condition assessment. Cannabis exhibits a rapid early increase, a pronounced mid-cycle peak, and a subsequent decline consistent with maturation and harvesting dynamics. Forest remains high and comparatively stable, while shrub/scrub shows a narrower dynamic range. This separation explains why truncated temporal windows remain informative even before the full seasonal signal is observed, and it also helps explain why the eight-acquisition window emerges as the earliest stable operating point: by that stage, the cannabis curve has already diverged from both background classes in both slope and peak development.

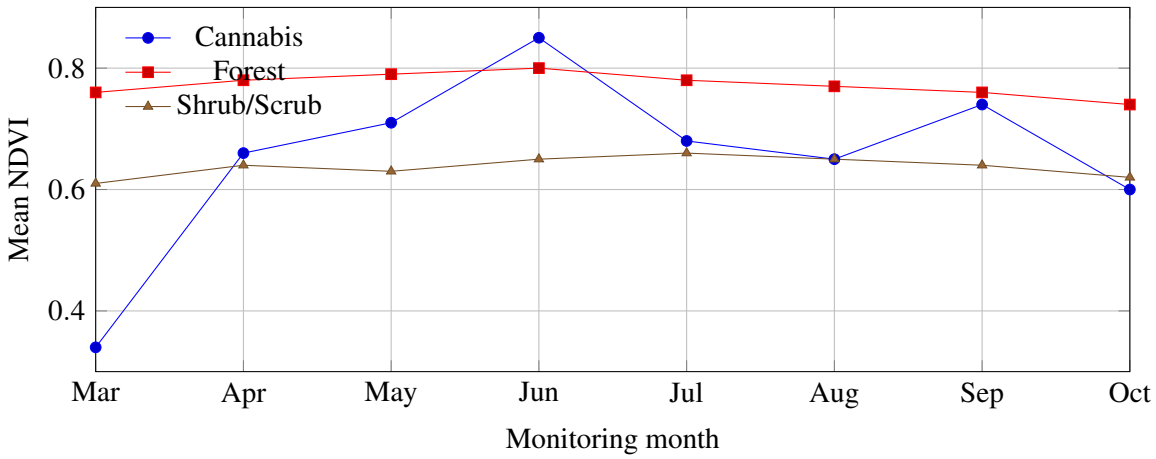

Figure 3: Class-wise NDVI trajectories reconstructed from the source study area. Cannabis maintains a steeper rise and sharper mid-cycle peak than background vegetation, preserving biological interpretability in the early-warning setting.

DISCUSSION

This study demonstrates that the principal limitation of the source framework is not a lack of discriminatory signal, but the timing and operational framing of that signal. The original paper showed that multitemporal phenology can distinguish illicit cannabis from forest and shrub/scrub with high accuracy in a cloud-prone tropical setting [8]. The present extension shows that the same result can be made *earlier*, *more conservative*, and *more actionable* without changing the underlying imagery. Its contribution is therefore not a radically new classifier, but a controlled redesign of validation, feature usage, and decision logic around the same evidence base.

The most important finding is that high-value performance is reached before the full usable time series is exhausted. This matters because enforcement decisions are time-sensitive. A model that reaches 93.7% accuracy and nearly 0.90 cannabis recall at the eight-acquisition stage is not merely statistically adequate; it changes the intervention window. In practice, that means surveillance teams do not need to wait for the full maturity of the spectral signal before prioritizing reconnaissance. The accompanying fold-wise dispersion is also modest, which strengthens the argument that the eight-acquisition operating point is a stable within-dataset signal rather than an artifact of a single favorable partition.

The results also clarify why context matters. In the source paper, most classification difficulty arose at the interface between cannabis and surrounding forest, especially where plot boundaries were small relative to the 10 m Sentinel-2 pixel size. By incorporating neighborhood variance and simple patch-structure descriptors, the extended model more effectively distinguishes coherent planted anomalies from isolated noisy pixels. The ablation analysis reinforces this interpretation by showing that most of the recall gain disappears when those descriptors are removed. This is consistent with broader remote sensing evidence that local spatial context can materially improve small-object discrimination when spectral signatures alone are partially mixed [11, 12].

Equally important is the move from hard labels to calibrated probabilities. In law-enforcement applications, a ranked list of suspicious sites is often more useful than an unqualified thematic map. A calibrated posterior allows analysts to choose thresholds based on operational resources, terrain accessibility, and tolerance for false positives. The reduction in ECE from 0.086 to 0.027 therefore represents a meaningful deployment gain, not merely a statistical refinement. In this setting, calibration supports proportional response: higher-

probability sites can be prioritized for immediate verification, while lower-probability sites can be monitored until additional imagery is available.

The study also highlights a methodological point that extends beyond this specific application. Reported accuracy in remote sensing can be optimistic when adjacent pixels or patches are split across training and test sets. By using site-level spatial blocking, we obtain a more conservative estimate of generalization and a more credible basis for operational claims [13]. This is particularly important for sensitive applications, where overstated performance can produce costly misallocation of resources. For the same reason, the deliberately conservative model family used here should be read as a design choice rather than as a claim that higher-capacity temporal models are unnecessary; with a dataset of this size, preserving interpretability and comparability was methodologically preferable to adding a more complex benchmark that could not be evaluated with equal confidence.

Several limitations should be acknowledged. First, the analysis is still confined to a single regional dataset, and the observed phenological signatures may shift under markedly different planting calendars, topographic regimes, or intercropping practices. Accordingly, the present findings should be interpreted as evidence of operational feasibility in Aceh Besar, not as proof of immediate cross-district transferability. Second, although the extension improves the use of available optical data, it does not solve the fundamental cloud-obstruction problem; areas that remain persistently obscured still require complementary sensing, most plausibly Sentinel-1 SAR fusion as suggested in the source paper. Third, the current framework is tuned for monoculture or near-monoculture patterns. Its performance in intentionally concealed intercropped systems may be weaker and warrants explicit testing. Fourth, the reported decision thresholds were optimized only within the present dataset and should therefore be recalibrated before any external deployment.

Finally, because the application concerns surveillance, technical accuracy must be paired with governance discipline. The probability maps generated by this framework should be understood as prioritization tools for field verification, not as stand-alone evidence for enforcement action. In other words, the methodological extension improves decision support, not evidentiary sufficiency. This distinction is essential for responsible deployment.

CONCLUSION

By re-analyzing the same Aceh Besar dataset used in the source paper, this study converts a strong retrospective detector into a lead-time-aware, uncertainty-calibrated early-warning system. The extension is deliberately conservative in data terms: no new sensors, no new field campaigns, and no change in class definitions were introduced. Instead, the advance comes from reorganizing the original Sentinel-2 observations into partial temporal windows, enriching the feature space with local context, validating with spatial blocks, reporting fold-wise stability, and calibrating the cannabis posterior for threshold-based deployment.

Three conclusions follow. First, the source dataset contains an actionable detection signal earlier than the original retrospective workflow implied; with eight usable acquisitions, the model still achieves 93.7% overall accuracy and 0.897 cannabis recall, and that operating point remains stable across the blocked folds. Second, mixed-pixel and boundary errors can be reduced materially using context descriptors derived from the same patches, with the ablation analysis indicating that these descriptors account for the largest share of the recall gain. Third, probability calibration substantially improves the decision value of the output by allowing enforcement teams to prioritize limited field resources in a principled manner while keeping threshold selection explicit and revisable.

The central implication is that medium-resolution Sentinel-2 phenology, when modeled with appropriate

temporal and probabilistic rigor, can support not only accurate mapping but also earlier and more operationally useful surveillance in cloud-constrained tropical environments within the study area analyzed here. Future work should test cross-district transferability, explicitly evaluate intercropped concealment scenarios, and integrate SAR features for windows in which optical coverage remains inadequate.

ACKNOWLEDGMENT

The manuscript presented here is designed as a direct methodological extension of the Aceh Besar detection framework and therefore conceptually inherits the operational setting established by the original study. Any applied deployment should be conducted only under appropriate legal authority, with field verification, locally revalidated thresholds, and proportional safeguards.

REFERENCES

[1] Andre, C. M., Hausman, J. F., and Guerriero, G. (2016). Cannabis sativa: The plant of the thousand and one molecules. *Frontiers in Plant Science*, 7, 19. doi:10.3389/fpls.2016.00019.

[2] Bicakli, F., Kaplan, G., and Alqasemi, A. S. (2022). Cannabis sativa L. spectral discrimination and classification using satellite imagery and machine learning. *Agriculture*, 12(6), 842. doi:10.3390/agriculture12060842.

[3] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324.

[4] Congalton, R. G., and Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (2nd ed.). Boca Raton, FL: CRC Press.

[5] Ferreira, A., Felipussi, S. C., Pires, R., Avila, S., Santos, G., Lambert, J., Huang, J., and Rocha, A. (2019). Eyes in the skies: A data-driven fusion approach to identifying drug crops from remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12), 4773–4786. doi:10.1109/JSTARS.2019.2917024.

[6] Guo, Y., De Jong, K., Liu, F., Wang, X., and Li, C. (2012). A comparison of artificial neural networks and support vector machines on land cover classification. In Z. Li, X. Li, Y. Liu, and Z. Cai (Eds.), *Computational Intelligence and Intelligent Systems* (pp. 531–539). Berlin: Springer. doi:10.1007/978-3-642-34289-9_59.

[7] Huang, S., Tang, L., Hupy, J. P., Wang, Y., and Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1), 1–6. doi:10.1007/s11676-020-01155-1.

[8] Irawadi, D., Mauritsius, T., Kushardono, D., Budhiman, S., Diwyacitta, K., Adhitama, B. S., Ayubi, F. A., Maftukhaturrizqoh, O., and Supriyani, I. S. (2025). Smart detection of illicit cannabis plantations using remote sensing technology and machine learning. *Geography, Environment, Sustainability*, 18(1), 130–138. doi:10.24057/2071-9388-2025-3538.

[9] Mattiuzzi, M., Bussink, C., and Bauer, T. (2014). Analysing phenological characteristics extracted from Landsat NDVI time series to identify suitable image acquisition dates for cannabis mapping in Afghanistan. *Photogrammetrie, Fernerkundung, Geoinformation*, 2014(5), 383–392. doi:10.1127/1432-8364/2014/0231.

[10] Niculescu-Mizil, A., and Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 625–632). New York: ACM. doi:10.1145/1102351.1102430.

[11] Park, S., Im, J., Park, S., Yoo, C., Han, H., and Rhee, J. (2018). Classification and mapping of paddy rice by combining Landsat and SAR time series data. *Remote Sensing*, 10(3), 447. doi:10.3390/rs10030447.

[12] Pelletier, C., Webb, G. I., and Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5), 523. doi:10.3390/rs11050523.

[13] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schroeder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. doi:10.1111/ecog.02881.

[14] Sujud, L., Jaafar, H., Hassan, M. A. H., and Zurayk, R. (2021). Cannabis detection from optical and RADAR data fusion: A comparative analysis of the SMILE machine learning algorithms in Google Earth Engine. *Remote Sensing Applications: Society and Environment*, 24, 100639. doi:10.1016/j.rsase.2021.100639.

[15] Suliman, A., and Zhang, Y. (2015). A review on back-propagation neural networks in the application of remote sensing image classification. *Journal of Earth Science and Engineering*, 5, 52–65.

[16] Yang, J., Dong, J., Liu, L., Zhao, M., Zhang, X., Li, X., Dai, J., Wang, H., Wu, C., You, N., Fang, S., Pang, Y., He, Y., Zhao, G., Xiao, X., and Ge, Q. (2023). A robust and unified land surface phenology algorithm for diverse biomes and growth cycles in China by using harmonized Landsat and Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202, 610–636. doi:10.1016/j.isprsjprs.2023.07.017.

Anne Vernez Moudon, Deaprtment of Urban Design and planning , University of Washington, Seattle, WA 98105, United States

Rebelo E. M., CITTA – Research Centre for Territory Transports and Environment, Faculty of Engineering, University of Porto, Porto, Portugal

Manuscript Published; 30 November 2025.