On the Reliability of Spectral Separability Claims: A Comparative Hyperspectral Signal-Processing Study of Crop-Weed Discrimination

Tom Avikasis Cohen[1], Anna Brook[1], Ran Nisim Lati[2]

[1]Spectroscopy & Remote Sensing Laboratory, School of Environmental Sciences, University of Haifa
[2]Department of Plant Pathology and Weed Research, Newe Ya'ar Research Center, Agricultural Research Organization

ABSTRACT

Claims of enhanced crop-weed spectral separability in hyperspectral agricultural imagery are often supported by single similarity metrics such as the Spectral Angle (SA), evaluated without controlling for class imbalance or cross-sample leakage. This study presents a comparative signal-processing framework for rigorously assessing such claims, applied to a case study of tomato and pepper plants under a candidate Pickering-emulsion-based treatment, discriminated against eight co-occurring weed species. The framework spans five mathematically independent formulations: a spectral-spatial convolutional network, a spectral-only 1D-CNN ablation isolating pure reflectance signal from spatial confounds, and three control-referenced one-class models – Mahalanobis distance under Ledoit-Wolf shrinkage covariance estimation, a reconstruction-error autoencoder, and Deep SVDD – each trained exclusively on untreated control pixels and evaluated via leave-one-day-out cross-validation with per-day class balancing. The results show that a naive SA-based analysis suggests an apparent positive treatment effect that does not survive this rigorous evaluation: none of the five methods detect a statistically significant improvement in crop-weed separability under treatment, and three independently formulated models show a small but consistent negative trend for tomato crop-identity. Input-gradient band saliency further localizes the physically-interpretable spectral regions driving crop-weed discrimination to pigment-absorption (≈435-500 nm) and red-edge-NIR (Near Infra Red) structural (≈750-800 nm) bands. These results illustrate how methodological choices – not just modeling architecture – determine the reliability of spectral separability claims in hyperspectral remote sensing.

Index Terms: hyperspectral remote sensing, spectral separability, class imbalance, leave-one-day-out cross-validation, one-class classification, crop signaling

1. INTRODUCTION

Global food security increasingly depends on effective weed management, yet the expansion of herbicide-resistant weed populations and tightening regulatory restrictions on herbicide use are eroding the effectiveness of conventional broad-spectrum control [1]. Site-specific weed management (SSWM) addresses this challenge by targeting treatment to the spatial and temporal distribution of weeds rather than treating entire fields uniformly [2]. The performance of SSWM, however, depends critically on the ability to reliably differentiate crop plants from weeds using remote sensing.

Claims that an agronomic treatment "enhances spectral separability" between crops and weeds are common in the precision-agriculture hyperspectral literature [3], and are typically supported by a single similarity metric – most often the Spectral Angle Mapper (SAM) – computed relative to one reference class [4], without cross-validation across acquisition days or control for class imbalance between the compared populations. Such an evaluation answers a narrower question than it appears to: it shows that a distribution of pixels has moved in spectral-angle space [5], not that a classifier could actually use that movement to separate crop from weed more reliably than before, out of sample, on a day it has never seen [6].

This distinction matters because hyperspectral pixel populations are highly autocorrelated within an acquisition day and highly imbalanced when many weed species are pooled against a single crop class – two conditions under which naive accuracy and single-reference distance metrics are known to overstate an effect [7]. In this work we take a specific case study – Pickering-emulsion-based sprays evaluated as "crop-signaling" compounds via SA and reported to enhance crop-weed spectral differentiation – and ask a methodological question: does that apparent effect survive a rigorous, leakage-free, class-balanced re-evaluation using multiple, mathematically independent formulations?

We present a five-method comparative framework spanning two learned-representation paradigms (a spectral-spatial convolutional network and a spectral-only 1-D convolutional ablation) and three control-referenced one-class formulations (Mahalanobis distance under Ledoit-Wolf shrinkage, a reconstruction autoencoder, and Deep SVDD), all evaluated under leave-one-day-out cross-validation with per-day class balancing. We show that the naive SA-based effect does not survive this scrutiny, and we use gradient-based band saliency to localize which spectral regions the data actually support as informative – independent of whether a treatment effect exists. The contribution is the framework itself: a template for interrogating spectral-separability claims in

hyperspectral agricultural remote sensing before they are taken as evidence of a physical or biochemical mechanism.

2. DATA ACQUISITION

Hyperspectral reflectance cubes were acquired with a pushbroom xiSpec camera (XIMEA, Germany; 992 spectral bands, 400-1000 nm, 12-mm NIR filter). Tomato (Solanum lycopersicum cv. 'Dakota') and bell pepper (Capsicum annuum cv. 'Mika') plants were sprayed with shellac-nanoparticle-stabilized Pickering emulsions at two medium-chain-triglyceride oil concentrations (EM20, EM40) or left untreated (control), and imaged daily alongside eight untreated weed species (W1-W8) over a 14-day period (11 usable acquisition days; day 10 excluded due to a missing raw file). Pixel-level reflectance spectra were extracted from segmented crop and weed regions for every day, treatment, and species. Wavelengths below 400 nm or above 950 nm were trimmed to remove noisy sensor edges.

3. METHODOLOGY

3.1 Naive Baseline: Spectral Angle Analysis

The original evaluation of this dataset used the Spectral Angle analysis [5], which computes the angle between a target reflectance vector R_t and a reference reflectance vector R_r:

$$SA = arccos\left(\left(\Sigma_{i=1..n} R_{t,i} \cdot R_{r,i}\right) / \left(\sqrt{\left(\Sigma_{i=1..n} R_{t,i}^2\right)} \cdot \sqrt{\left(\Sigma_{i=1..n} R_{r,i}^2\right)}\right) \right)$$

reported as SA_normalized = (|SA_control − SA_measured| / SA_control) × 100.

3.2 Evaluation Protocol

Leave-one-day-out cross-validation [8]: For each acquisition day (d), every model in this study is trained on data from all other days and evaluated only on day d, repeated across all 11 days. This prevents the same-day pixel autocorrelation that would otherwise let a model exploit day-specific illumination or canopy-geometry artifacts rather than a genuine spectral signature.

Per-day class balancing [9]: Pooling eight weed species against one crop treatment produces roughly 8:1 class imbalance (≈800-840 weed pixels vs. ≈100 crop pixels per day), under which a trivial majority-class classifier achieves high nominal accuracy without learning any real separability – an initial pass of this analysis exhibited exactly this failure mode. It is corrected by subsampling the majority class within each day to match the minority class before every split:

$$n_d = min(n_crop, d, n_weed, d)$$

3.3 Model A: Spectral-Spatial CNN

Reflectance spectra within a 5×5-pixel neighborhood are reduced via PCA [10]. For pooled center spectra $X \in \mathbb{R}^{n \times d}$ (d = 992 bands), we compute the covariance and its eigendecomposition, retaining the top k = 10 components (explaining 99.5% of spectral variance in this dataset) [11]:

$$\Sigma = (1/n)\, X^T X = V \Lambda V^T, \; x' = V_k^T (x - \mu)$$

PCA was fit once on the pooled dataset across all days rather than refit per fold; because this is an unsupervised, label-free transformation, any resulting leakage is expected to be mild and, if anything, biased toward finding an effect, making the null result reported below conservative. Each 5×5 patch is represented as a 10-channel tensor of retained principal components and passed through a compact 2-D CNN: two 3×3 convolution + batch-norm + ReLU blocks (10→16→32 channels), global average pooling, a 32-unit embedding, dropout (p = 0.3), and a 2-class linear head. Because it has access to spatial texture and shape context, this model can exploit morphological shortcuts unrelated to the treatment's spectral effect – motivating the ablation below.

3.4 Model B: Spectral-Only 1-D CNN (Ablation)

To isolate the genuine spectral contribution from spatial confounds, the 2-D spatial model was replaced with a 1-D CNN [12] operating on the raw 992-band spectrum of a single pixel, with no spatial neighbors and no PCA reduction: two strided 1-D convolution + batch-norm + ReLU + max-pool blocks (1→16→32 channels, kernel sizes 15 and 9), global average pooling, a 32-unit embedding, dropout, and a 2-class head. Since this architecture has zero access to spatial context, any accuracy it achieves must come from the reflectance spectrum alone – it is the direct spectral-domain analogue of the single-pixel SA baseline, evaluated under the leakage-free, class-balanced protocol above, making it the primary apples-to-apples comparison against the original naive claim.

3.5 Band Saliency

For the trained spectral-only model we compute a gradient-based saliency score [13] per wavelength band (b). For the crop-minus-weed logit difference s(x) = f_crop(x) − f_weed(x), we back-propagate to the input and average the absolute gradient over N evaluated pixels:

$$Saliency(b) = (1/N)\, \Sigma_{j=1..n} \left|\partial s(x_j) / \partial x_{j,b}\right|$$

Gradient-based saliency in 1-D CNNs is known to produce inflated values near the input edges regardless of physical wavelength (a boundary artifact of convolution padding); we confirmed this by trimming the spectrum and observing the artifact relocate to the new edge, and exclude a 30 nm margin at each end from interpretation accordingly.

3.6. Models C-E: Control-Referenced One-Class Framework

The one-class framework directly addresses the confound in the SA baseline: each model is fit exclusively on untreated control-crop pixels, then frozen and used to score how anomalous other pixel populations are relative to that learned distribution. If a treatment genuinely increases crop-weed separability, treated-crop-vs-weed AUC should differ from control-crop-vs-weed AUC under a model that has never seen the treatment; if the treatment instead perturbs the crop's own identity, that manifests as reduced AUC for treated crop relative to control.

3.6.1. Mahalanobis distance (classical, linear)

For each held-out day, we fit a multivariate Gaussian to control-crop training pixels using Ledoit-Wolf shrinkage [1] covariance estimation, necessary because the spectral dimensionality (≈900 bands) exceeds the number of training pixels per fold. The anomaly score is the squared Mahalanobis distance [14]:

$$D^2(x) = (x - \mu)^{\mathrm{T}} \Sigma^{-1} (x - \mu)$$

3.6.2. Autoencoder (nonlinear, reconstruction-based)

A small encoder-decoder network (992→32→8→32→992, ReLU) [15] is trained to reconstruct control-crop spectra by minimizing mean squared error. The anomaly score is the reconstruction error itself:

$$s(x) = \|x - g_\varphi(f_\theta(x))\|^2$$

3.6.3. Deep SVDD (nonlinear, compactness-based)

Following Ruff et al. (2018) [16], a small feedforward network φ_θ (992→32→8, with standard ReLU activations and no bias terms in any layer) is trained to map control-crop pixels close to a fixed center c:

$$L = (1/n)\, \Sigma_{\mathrm{i}} \|\varphi_\theta(x_{\mathrm{i}}) - c\|^2$$

The center c is computed once as the mean embedding of an initial forward pass (near-zero components nudged to $\varepsilon = 0.1$) and never updated by gradient descent thereafter. The no-bias-layer constraint and fixed center are standard safeguards against the well-known "hypersphere collapse" trivial solution.

For each held-out day, every model is fit on that day's-other-days' control pixels, then used to score control, EM20, EM40, and weed pixels for the held-out day; AUC-ROC [17] is reported, separating each treatment from weed under the frozen, control-only-trained model, balanced to equal N per side.

4. RESULTS

Under the original evaluation of this dataset, using the Spectral Angle analysis metric, EM40-treated pepper and tomato showed normalized deviations exceeding 20% (approaching 30-40% for several weed references), strongest 1-4 days after treatment. Critically, SA as applied compares treated-crop-to-weed against untreated-crop-to-weed only; it was never used to test whether treated crop remains close to its own untreated baseline. A companion biomass assay in the same study found EM40 caused a significant reduction in tomato fresh biomass relative to both control and EM20 (Tukey-Kramer, $p<0.05$) – a phytotoxic response that SA alone cannot distinguish from a genuine signaling effect.

4.1 Spectral-Spatial CNN (Model A)

Under leave-one-day-out evaluation with per-day class balancing, Model A achieves near-ceiling accuracy for pepper across all treatments (control 0.997±0.003, EM20 0.997±0.005, EM40 0.997±0.006; mean±SD across 11 held-out days) – a ceiling effect leaving no room for a treatment-driven improvement to be detected regardless of whether one exists. For tomato, accuracy is high but shows a small monotonic decrease with concentration (control 0.993±0.004, EM20 0.986±0.013, EM40 0.982±0.013). Because this model has spatial-morphological access, this result alone cannot distinguish genuine spectral effect from shape-based artifact.

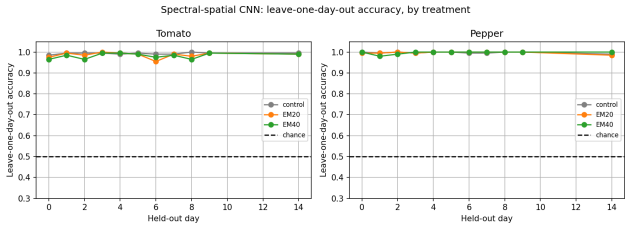


Fig. 1. Spectral-spatial CNN leave-one-day-out accuracy by treatment.

4.2 Spectral-Only 1-D CNN Ablation (Model B)

With spatial context removed, accuracy drops substantially for both crops relative to Model A (pepper ≈0.90, tomato ≈0.71–0.75), confirming that a meaningful share of Model A's performance was spatial-morphological rather than pure spectral signal. Under a paired t-test across the 11 held-out days (treatment vs. control), no treatment effect reaches significance for either crop or concentration: tomato EM20 $t=0.65$, $p=0.53$; tomato EM40 $t=-0.42$, $p=0.68$; pepper EM20 $t=0.05$, $p=0.96$; pepper EM40 $t=-0.30$, $p=0.77$.

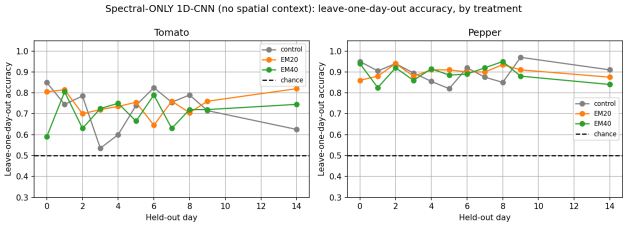


Fig. 2. Spectral-only 1-D CNN leave-one-day-out accuracy by treatment.

4.3 Band Saliency

Input-gradient saliency for the EM40 spectral-only model localizes two physically-interpretable regions of elevated importance in both crops, outside the excluded boundary margins: (i) ≈435-500 nm (peaking at 449.5 nm in tomato and 497.7 nm in pepper), consistent with chlorophyll/carotenoid pigment absorption, and (ii) ≈750-800 nm, in the red-edge-NIR structural-reflectance region. Pepper's saliency values are substantially higher in absolute terms across the whole spectrum (peak ≈1.0 vs. ≈0.37 for tomato), consistent with its near-ceiling classification accuracy.

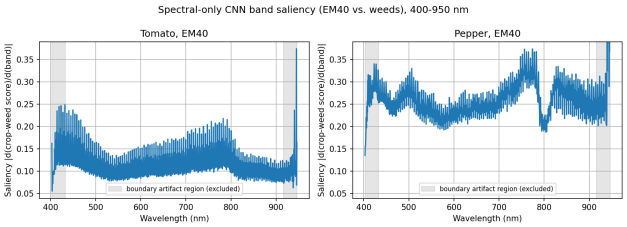


Fig. 3. Band saliency, shared y-axis (EM40 vs. weeds).

4.4 Control-Referenced One-Class Framework (Models C–E)

Table 1 summarizes mean AUC (treatment vs. weed, control-only-trained model) across the three formulations.

Crop	Model	Control	EM20	EM40
Tomato	Mahalanobis	0.789	0.776	0.784
Tomato	Autoencoder	0.883	0.835	0.824
Tomato	Deep SVDD	0.878	0.849	0.838
Pepper	Mahalanobis	0.825	0.834	0.832
Pepper	Autoencoder	0.916	0.915	0.916
Pepper	Deep SVDD	0.949	0.950	0.948

Table 1. Mean AUC (treatment vs. weed, control-only-trained model).

No formulation shows treated-crop AUC exceeding control-crop AUC for either species. Pepper is essentially flat across all three methods ($\Delta<0.01$), consistent with a ceiling effect. For tomato, both nonlinear methods (Autoencoder, Deep SVDD) show a consistent decrease from control to EM40 (−0.059 and −0.040 AUC respectively) not detected by linear Mahalanobis – suggesting the effect, where present, is a nonlinear perturbation of spectral identity rather than a linear shift in mean reflectance. Paired t-tests across the 11 held-out days confirm this trend is statistically significant for the two nonlinear methods (Autoencoder: EM20 $p=0.021$, EM40 $p=0.002$; Deep SVDD: EM20 $p=0.012$, EM40 $p<0.0001$), but not for linear Mahalanobis ($p>0.4$ in both cases), and not for pepper under any method ($p>0.5$ in all cases).

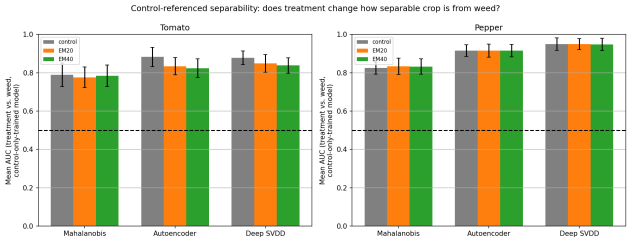


Fig. 4. Control-referenced one-class AUC comparison.

5. DISCUSSION

Across five mathematically independent methods – two learned-representation classifiers evaluated under leave-one-day-out, class-balanced cross-validation, and three control-referenced one-class formulations spanning classical statistics to deep nonlinear embeddings – none reproduce the positive crop-weed spectral-separability effect reported by the naive single-reference SA analysis. The gap is explained, at least in part, by two methodological issues absent from the original evaluation: (1) no correction for the severe class imbalance between one crop and eight pooled weed species, which alone can manufacture the appearance of strong separability from a majority-class collapse; and (2) no genuinely control-referenced comparison – the original SA protocol measures distance from weed references only, and cannot distinguish a crop moving further from weeds because of enhanced signaling from a crop moving away from its own healthy baseline because of treatment-induced stress.

The latter is not speculative here: the same underlying dataset's biomass assay found a significant, concentration-dependent phytotoxic reduction in tomato fresh biomass under EM40, and three of our five methods (spectral-spatial CNN, autoencoder, Deep SVDD) independently detect a small but consistent negative shift specifically for tomato under EM40 – directionally consistent with a stress-driven perturbation of crop spectral identity rather than a beneficial signaling enhancement. Pepper, which showed no significant biomass reduction under EM20 and only a moderate one under EM40 in the original assay, is correspondingly flat across every method here.

Band saliency independently anchors both crops' crop-vs-weed discriminability to physically plausible pigment-absorption and red-edge/NIR structural regions, a useful positive finding on its own: the classifiers are not relying on an illumination or acquisition artifact, even though the treatment itself does not measurably move the needle.

These findings motivate three concrete best practices for evaluating spectral-separability claims in hyperspectral agricultural imagery: (1) always balance target classes on a per-acquisition-day basis before evaluation, since pooled multi-species comparisons are highly prone to majority-class collapse; (2) use leave-one-group-out cross-validation (by date or flight) rather than random splits, to eliminate temporal and illumination leakage; and (3) benchmark any treatment against the target's own untreated control distribution, not only against the opposing class, since distance from an opposing class alone cannot distinguish a genuine treatment effect from a stress-induced perturbation of the target's own identity.

6. CONCLUSION

A rigorous, leakage-free, class-balanced, multi-method re-evaluation of a Pickering-emulsion crop-signaling treatment finds no reproducible spectral-separability benefit, in contrast to the positive effect suggested by a naive single-reference SA analysis of the same underlying hyperspectral dataset. The discrepancy is attributable to identifiable, correctable methodological gaps – uncontrolled class imbalance and the absence of a genuine control-referenced baseline – rather than to any property specific to this treatment. We offer the five-method comparative framework presented here as a general-purpose template for interrogating spectral-separability claims in hyperspectral agricultural remote sensing before they are attributed to a physical or biochemical mechanism.

REFERENCES

[1] Gerber, J. S., Ray, D. K., Makowski, D., Butler, E. E., Mueller, N. D., West, P. C., ... & Sloat, L. (2024). Global spatially explicit yield gap time trends reveal regions at risk of future crop yield stagnation. Nature Food, 5(2), 125-135.

[2] Lati, R. N., Rasmussen, J., Andujar, D., Dorado, J., Berge, T. W., Wellhausen, C., ... & Christensen, S. (2021). Site-specific weed management – constraints and opportunities for the weed research community: Insights from a workshop. *Weed Research*, 61(3), 147-153.

[3] Li, Y., Al-Sarayreh, M., Irie, K., Hackell, D., Bourdot, G., Reis, M. M., & Ghamkhar, K. (2021). Identification of weeds based on hyperspectral imaging and machine learning. *Frontiers in Plant Science*, 11, 611622.

[4] Lass, L. W., Thill, D. C., Shafii, B., & Prather, T. S. (2002). Detecting spotted knapweed (Centaurea maculosa) with hyperspectral remote sensing technology. *Weed Technology*, 16(2), 426-432.

[5] Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., & Goetz, A. F. (1993). The spectral image processing system (SIPS) – interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3), 145-163

[6] Karasiak, N., Dejoux, J. F., Monteil, C., & Sheeren, D. (2022). Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, 111(7), 2715-2740.

[7] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.

[8] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929.

[9] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9), 1263-1284.

[10] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical transactions. *Series A, Mathematical, physical, and engineering sciences*, 374(2065), 20150202.

[11] Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE transactions on geoscience and remote sensing*, 54(10), 6232-6251.

[12] Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015(1), 258619.

[13] Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

[14] Mahalanobis, P. C. (2018). On the generalized distance in statistics. *Sankhyā: The Indian Journal of Statistics, Series A* (2008-), 80, S1-S7.

[15] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786), 504-507.

[16] Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., ... & Kloft, M. (2018, July). Deep one-class classification. In *International conference on machine learning*(pp. 4393-4402). PMLR.

[17] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.