Proceedings of Machine Learning Research **318** (2026)
The 39th Canadian Conference on Artificial Intelligence

Canadian AI 2026
25–29 May 2026, Vancouver, BC

Next-Generation AI Vegetation Analytics: Low-Cost PSRI Translation from RGB and NDVI for Precision Crop Monitoring

Jacob Serafin [‡,*], Yuvraj Singh Gill[†], Lokesh Kumar K M[‡], Dylan C. P. Lewis[‡], Gurjit Singh Randhawa[‡], Aitazaz A. Farooque[†,◇]

[‡] School of Computer Science, University of Guelph, ON, Canada
[†] Faculty of Sustainable Design Engineering, University of Prince Edward Island, PE, Canada
[◇] Canadian Centre for Climate Change and Adaptation, PE, Canada

Abstract

Plant senescence monitoring is important for crop health assessment and precision agriculture, but senescence-sensitive indices such as the Plant Senescence Reflectance Index (PSRI) are not always readily available in low-cost imaging workflows. This study presents an inexpensive artificial intelligence approach to translate RGB and Normalized Difference Vegetation Index (NDVI) imagery into PSRI maps using supervised image-to-image regression. Using tiled samples from the open-sourced Canadian Cropland Dataset, three models (UNet, pix2pix GAN, and R2AttUNet) and a baseline model (linear regressor) were trained and evaluated under a consistent pipeline using mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). The three models all outperformed the baseline across all scores; of those, UNet achieved the best performance on both validation and test sets, producing the lowest MAE (0.0248) and the highest PSNR (19.85) and SSIM (0.9009) on the test set, while GAN showed competitive but weaker results and R2AttUNet underperformed. The close validation and test agreement indicates stable generalization under the current split. These results demonstrate the feasibility of estimating PSRI from inexpensive RGB+NDVI inputs and support the use of lightweight convolutional models for low-cost vegetation monitoring.

Keywords: Plant Senescence Reflectance Index (PSRI), Convolutional Neural Networks, Image-to-Image Translation, Deep Learning, Multimodal Data Fusion

1. Introduction

Plant senescence is a biologically decisive phase of crop development, involving coordinated changes in chlorophyll degradation, canopy structure, and water status that directly affect late-season photosynthesis, yield formation, and crop quality [1]. Despite its agronomic importance, operational vegetation monitoring still relies heavily on greenness-oriented indices, most notably the Normalized Difference Vegetation Index (NDVI), which are highly effective for vegetation presence and biomass estimation but comparatively less sensitive to early senescence physiology and prone to saturation in dense canopies [2, 3]. The result is a persistent mismatch between the biological process of interest and the spectral proxies most commonly available in low-cost monitoring pipelines.

The Plant Senescence Reflectance Index (PSRI) was developed specifically to capture senescence-associated pigment transitions, especially shifts in the chlorophyll-to-carotenoid balance, using a reflectance relationship that combines visible and near-infrared (NIR) wavelengths [4]. However, PSRI is rarely available in low-cost sensing workflows because it depends on specific spectral evidence that inexpensive RGB systems do not directly measure [5]. This creates a key practical barrier, the index most relevant for senescence is often the least accessible for large-scale deployment.

Leaf and canopy reflectance are shaped by visible-band pigment absorption and NIR scattering, meaning that vegetation monitoring is not simply a "colour" problem but a coupled optical response to biochemical and structural states [6]. Senescence further complicates this

[*] Corresponding Author: serafinj@uoguelph.ca

relationship because it is not a single monotonic decline in greenness; instead, it is a non-linear transition involving pigment ratio shifts, structural breakdown, and canopy geometry effects that can mimic or obscure stress signatures [5]. Consequently, any attempt to infer senescence-sensitive indices from cheaper inputs must preserve physiologically meaningful spectral relationships rather than learning superficial colour-texture shortcuts.

Recent advances in artificial intelligence suggest that spectral or index "translation" may provide a path forward. Peer-reviewed studies have demonstrated that deep learning and image-to-image translation can predict NDVI or Normalized Difference Red Edge (NDRE) from RGB imagery, and can even synthesize NIR-like channels under constrained learning objectives, potentially reducing the need for costly multispectral sensors [7–9]. At the same time, the literature is clear that these models are vulnerable to domain shift (e.g., sensor, season, illumination, site), and visually plausible outputs do not necessarily preserve physically valid spectral structure for downstream index computation [10, 11]. Thus, while AI-based index synthesis is technically feasible, robust physiological translation remains an open problem.

RGB-based vegetation indices remain attractive because they are inexpensive, scalable, and compatible with commodity cameras, and foundational agricultural vision studies have shown that visible-band colour transforms can effectively separate vegetation from soil and residue [12, 13]. Yet RGB-only signals are pigment-ambiguous and lack direct NIR information, while NDVI mainly contributes structural and biomass sensitivity. This makes the combination of RGB & NDVI especially compelling for PSRI inference: RGB contributes visible spectral texture, NDVI contributes NIR-informed canopy structure, and PSRI serves as the senescence-targeted output.

This study addresses a literature gap by using the Canadian Cropland Dataset to develop a cost-effective AI approach for translating RGB and NDVI imagery into PSRI for vegetative monitoring. Although RGB to NDVI, and index-constrained translation have been reported, a rigorous demonstration of RGB + NDVI to PSRI translation with index-level fidelity and domain-aware validation remains missing [7, 8]. We frame PSRI estimation not as a cosmetic image synthesis task, but as a constrained inference problem in which preserving senescence-relevant spectral meaning is central to model design, evaluation, and deployment.

2. Materials and methods

2.1. Dataset & Preprocessing

This study uses the open-sourced Canadian Cropland Dataset [14]. From this, canola and corn samples, restricted to RGB, NDVI, and PSRI bands are used from the years 2017-2020. For each sample, the RGB, NDVI, PSRI tiles are paired by key and loaded as aligned rasters. The RGB tiles are represented as three-channel rasters and normalized to [0,1] by scaling the 8-bit PNG intensities. The NDVI and PSRI tiles are loaded as a single-channel rasters and are loaded using an 8-bit grayscale pipeline. Each tile is read in as a single-channel image, converted to a uint8 array with pixel intensities in [0,255] and then cast to floating point and normalized to [0,1] by dividing by 255. The normalized NDVI raster is concatenated with the normalized RGB channels to form the model input, preserving pixelwise alignment across bands.

The training split contains 16,969 samples, comprising 7,953 canola and 9,016 corn tiles. The validation dataset consists of 3635 samples, with 1696 canola tiles and 1939 corn tiles. The test split contains 3,640 samples, consisting of 1,717 canola and 1,923 corn tiles. Data augmentation includes random paired horizontal and vertical flips.

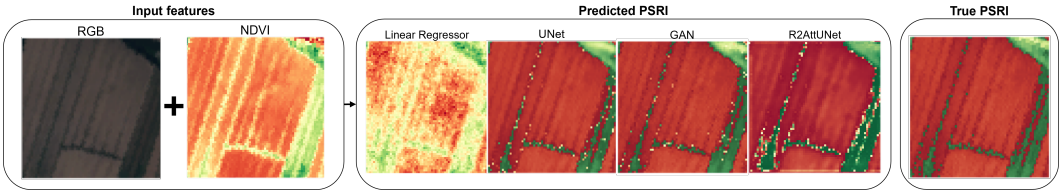

Figure 1. Input features, PSRI predictions, and true PSRI for an example data point.

2.2. **Model Selection**

This study addresses a supervised image-to-image regression problem, where two aligned raster layers are used to predict a third raster layer at the same spatial resolution. Let $x_{\text{RGB}} \in \mathbb{R}^{3\times H\times W}$ be the RGB input and $x_{\text{NDVI}} \in \mathbb{R}^{1\times H\times W}$ be the NDVI input, and let $\hat{y} \in \mathbb{R}^{1\times H\times W}$ be the PSRI prediction. The goal is to learn a function f_θ such that:

$$\hat{y} = f_\theta\big([x_{\text{RGB}},\, x_{\text{NDVI}}]\big) \tag{2.1}$$

where $[x_{\text{RGB}},\, x_{\text{NDVI}}]$ denotes channel concatenation. Thus, the input for the models is $x = [x_{\text{RGB}},\, x_{\text{NDVI}}] \in \mathbb{R}^{4\times H\times W}$, where H = W = 65 for all samples.

Four architectures are considered: a baseline linear regressor, UNet, pix2pix GAN and R2AttUNet. The baseline linear regressor serves as the primary image-to-image regression model. The UNet model uses an encoder-decoder structure with skip connections to map the stacked input rasters to the target layer under a reconstruction objective. The pix2pix model builds upon that by introducing adversarial supervision via a PatchGAN discriminator that distinguishes real target tiles from generated targets conditioned on the same inputs, while the generator is trained with a weighted combination of adversarial loss and L1 loss to evaluate potential gains in structural realism and edge fidelity. Finally, R2AttUNet is evaluated as a higher-capacity variant that incorporates recurrent residual convolutional blocks to iteratively refine features at each scale and attention gates on skip connections to suppress irrelevant background responses and emphasize task-relevant regions, with the aim of improving reconstruction quality for fine-grained spatial patterns in the target raster.

2.3. **Performance Metrics**

The models are trained for 50 epochs with a batch size of 32. To mitigate overfitting, an early stopping callback was implemented. The networks are optimized using the Adam optimizer, with an initial learning rate of 10^{-3}. To ensure stable convergence, a learning rate scheduler is utilized to decay this learning rate dynamically based on the validation loss. Mean Absolute Error (L1 Loss), given in equation B.1, is used as the primary loss function due to its robustness to outliers and its tendency to preserve edges better than Mean Squared Error (L2 Loss) in image reconstruction settings [15]. Additionally, the models are evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) alongside the L1 Loss. PSNR is reported as a standard reconstruction fidelity metric that complements MAE by expressing the mean squared error on a logarithmic scale and is given by equation C.1. SSIM is used to quantify the structural similarity between prediction $\hat{y}$ and target y. SSIM is computed over a sliding window and averaged over the image and is calculated by equation D.1. SSIM is reported in [0,1], where higher values indicate better structural agreement.

3. Results and Discussion

Test and validation scores are given in Table 1. All three models outperform the linear regressor baseline by a large margin. The UNet architecture achieves the lowest MAE (0.0250 val / 0.0248 test), highest PSNR (19.71 val / 19.85 test), and highest SSIM (0.901 on both splits). The fact that validation and test metrics are extremely close, and the test is only marginally better, suggests stable generalization and no obvious overfitting. The metrics suggest the GAN's adversarial term did not provide a net benefit under the current loss weighting/training regime. R2AttUNet is worse, suggesting that the architecture's added complexity did not translate into improved performance and may have made training harder under the current setup.

Table 1. Reported MAE, PSNR, and SSIM on validation and test sets for the best checkpoint selected by minimum validation MAE.

Model	**Validation Set**			**Test Set**		
	MAE	**PSNR**	**SSIM**	**MAE**	**PSNR**	**SSIM**
UNet	**0.0250**	**19.71**	**0.9012**	**0.0248**	**19.85**	**0.9009**
GAN	0.0262	18.67	0.8942	0.0260	18.67	0.8942
R2AttUNet	0.083	19.7	0.719	0.0824	19.7905	0.7209
Linear Regressor	0.4461	5.03	0.1487	0.4409	5.06	0.1477

The qualitative comparison shown in Fig. 1 aligns closely with the quantitative performance reported in Table 1, highlighting clear differences in spatial reconstruction across the evaluated architectures. The UNet model demonstrates the strongest agreement with the ground-truth PSRI maps, preserving crop-row structure, boundary transitions, and overall reflectance patterns with minimal artifacts. Both GAN and R2AttUNet clearly outperform the linear regressor.

Overall, both visual and quantitative analyses indicate that a simpler UNet architecture provides a more stable and structurally accurate PSRI translation, while increased architectural complexity does not necessarily yield improved performance for this regression task.

4. Conclusion

This study investigated an inexpensive machine learning pipeline for translating aligned RGB and NDVI raster inputs into PSRI maps for vegetation monitoring, formulated as a supervised image-to-image regression task on 65×65 tiles from the open-sourced Canadian Cropland Dataset. The proposed framework concatenated normalized RGB and NDVI channels and evaluated three architectures (UNet, pix2pix GAN, and R2AttUNet) against a baseline learner (linear regressor) under a consistent training setup using L1 loss, early stopping, and validation-based checkpoint selection. UNet achieved the best overall performance, far outperforming the baseline, with the lowest MAE (0.0248) and the highest PSNR (19.85) and SSIM (0.9009) on the test set, indicating the most accurate and structurally consistent PSRI reconstruction among the tested models. The pix2pix GAN produced competitive but weaker results than UNet, followed by the R2AttUNet. All three models strongly outperformed the baseline. These findings indicate that, for RGB+NDVI to PSRI translation, a well-regularized convolutional regression model can outperform more complex adversarial or attention-recurrent variants, at least under the present data and training conditions. Future experiments with calibration-aware preprocessing and multi-task formulations may further improve robustness and interpretability for real-world vegetation monitoring applications.

Acknowledgments

This research was funded by the Natural Sciences and Engineering Research Council of Canada Discovery Grant RGPIN-2022-03547 to G.S.R. The authors acknowledge the use of AI tools for proofreading and grammar checking of author-written text.

Appendix A. Discriminator and Generator Loss

The discriminator is trained using binary cross-entropy with logits.

$$\mathcal{L}_D = \frac{1}{2}\Big(\mathrm{BCE}(D(x,y),1) + \mathrm{BCE}(D(x,G(x)),0)\Big) \tag{A.1}$$

where x is the input image, y is the ground-truth target, $G(x)$ is the generator output, $D(\cdot,\cdot)$ is the discriminator, and BCE is binary cross-entropy. The discriminator is trained to classify real pairs (x,y) as 1 and generated pairs $(x,G(x))$ as 0.

The generator is trained to both fool the discriminator and remain close to the target:

$$\mathcal{L}_G = \mathrm{BCE}\big(D(x,G(x)),1\big) + \lambda\,\|G(x)-y\|_1 \tag{A.2}$$

where λ is a weighting coefficient and $\|G(x)-y\|_1$ is the L_1 reconstruction loss. The first term encourages realistic outputs, while the second encourages numerical similarity to the target.

Appendix B. Mean Absolute Error

Mean Absolute Error (MAE) is given by:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{B.1}$$

where N is the total number of evaluated elements, y_i is the ground-truth value at index i, and $\hat{y}_i$ is the predicted value at the same index. The absolute value $|\hat{y}_i - y_i|$ measures the magnitude of the prediction error at element i, and MAE is the average absolute error over all N elements.

Appendix C. Peak Signal-to-Noise Ratio

PSNR is computed from Mean Squared Error (MSE) as:

$$\mathrm{PSNR}(\hat{y},y) = 10\log_{10}\left(\frac{1}{\mathrm{MSE}(\hat{y},y)}\right), \qquad \mathrm{MSE}(\hat{y},y) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{C.1}$$

where $\hat{y}$ is the predicted image, y is the ground-truth image, and $\mathrm{MSE}(\hat{y},y)$ is the mean squared error between them. Here, N is the total number of evaluated elements, $\hat{y}_i$ is the predicted value at index i, and y_i is the corresponding ground-truth value. The squared term $(\hat{y}_i - y_i)^2$ penalizes larger errors more strongly than MAE. PSNR is expressed in decibels.

Appendix D. Structural Similarity Index Measure

SSIM is given by:

$$\mathrm{SSIM}(y,\hat{y}) = \frac{\big(2\mu_y\,\mu_{\hat{y}} + C_1\big)\big(2\sigma_{y\hat{y}} + C_2\big)}{\big(\mu_y^2 + \mu_{\hat{y}}^2 + C_1\big)\big(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2\big)} \tag{D.1}$$

where y is the ground-truth image and $\hat{y}$ is the predicted image. The quantity μ_y is the mean intensity of y, and $\mu_{\hat{y}}$ is the mean intensity of $\hat{y}$. The terms σ_y^2 and $\sigma_{\hat{y}}^2$ are the variances of y and $\hat{y}$, respectively, while $\sigma_{y\hat{y}}$ is the covariance between y and $\hat{y}$. Constants

C_1 and C_2 are small positive values introduced to stabilize the expression and avoid division by very small denominators.

References

[1] L. D. Noodén, J. J. Guiamét, and I. John. "Whole plant senescence". In: *Plant cell death processes*. Elsevier, 2004, pp. 227–244.

[2] C. J. Tucker. "Red and photographic infrared linear combinations for monitoring vegetation". In: *Remote sensing of Environment* 8.2 (1979), pp. 127–150.

[3] A. A. Gitelson. "Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation". In: *Journal of plant physiology* 161.2 (2004), pp. 165–173.

[4] M. N. Merzlyak, A. A. Gitelson, O. B. Chivkunova, and V. Y. Rakitin. "Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening". In: *Physiologia plantarum* 106.1 (1999), pp. 135–141.

[5] J. Anderegg, K. Yu, H. Aasen, A. Walter, F. Liebisch, and A. Hund. "Spectral vegetation indices to track senescence dynamics in diverse wheat germplasm". In: *Frontiers in plant science* 10 (2020), p. 1749.

[6] E. B. Knipling. "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation". In: *Remote sensing of environment* 1.3 (1970), pp. 155–159.

[7] C. Davidson, V. Jaganathan, A. N. Sivakumar, J. M. P. Czarnecki, and G. Chowdhary. "NDVI/NDRE prediction from standard RGB aerial imagery using deep learning". In: *Computers and Electronics in Agriculture* 203 (2022), p. 107396.

[8] A. A. Farooque, H. Afzaal, R. Benlamri, S. Al-Naemi, E. MacDonald, F. Abbas, K. MacLeod, and H. Ali. "Red-green-blue to normalized difference vegetation index translation: a robust and inexpensive approach for vegetation monitoring using machine vision and generative adversarial networks". In: *Precision Agriculture* 24.3 (2023), pp. 1097–1115.

[9] A. Picon, A. Bereciartua-Perez, I. Eguskiza, J. Romero-Rodriguez, C. J. Jimenez-Ruiz, T. Eggers, C. Klukas, and R. Navarra-Mestre. "Deep convolutional neural network for damaged vegetation segmentation from RGB images based on virtual NIR-channel estimation". In: *Artificial intelligence in agriculture* 6 (2022), pp. 199–210.

[10] J. Doornbos, Ö. Babur, and J. Valente. "Evaluating generalization of methods for artificially generating NDVI from UAV RGB imagery in vineyards". In: *Remote Sensing* 17.3 (2025), p. 512.

[11] M. Krestenitis, K. Ioannidis, S. Vrochidis, and I. Kompatsiaris. "Visual to near-infrared image translation for precision agriculture operations using GANs and aerial images". In: *Computers and Electronics in Agriculture* 237 (2025), p. 110720.

[12] G. E. Meyer and J. C. Neto. "Verification of color vegetation indices for automated crop imaging applications". In: *Computers and electronics in agriculture* 63.2 (2008), pp. 282–293.

[13] A. A. Gitelson, Y. J. Kaufman, R. Stark, and D. Rundquist. "Novel algorithms for remote estimation of vegetation fraction". In: *Remote sensing of Environment* 80.1 (2002), pp. 76–87.

[14] A. A. Boatswain Jacques, A. Diallo, and E. Lord. *The Canadian Cropland Dataset: A New Land Cover Dataset for Multitemporal Deep Learning Classification in Agriculture*. 2023. arXiv: 2306.00114 [cs.CV].

[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. "Image-to-Image Translation with Conditional Adversarial Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. ISSN: 1063-6919. July 2017, pp. 5967–5976. DOI: 10.1109/CVPR.2017.632.