Deep Generative Reconstruction of Near Infrared Band for Crop Health Monitoring under Variable Illumination Conditions

Alejandro Vidal-Domínguez
Universidad Católica del Norte
Robert Guamán-Rivera
Universidad Católica del Norte
Tito Arévalo-Ramírez
Pontificia Universidad Catolica de Chile
Elizabeth Lam-Esquenazi
Universidad Católica del Norte
Ranjit Das
Universidad Católica del Norte
Oswaldo Menéndez-Granizo
oswaldo.menendez@ucn.cl
Universidad Católica del Norte

Article

Keywords:

Posted Date: April 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8897011/v1

Additional Declarations: No competing interests reported.

Deep Generative Reconstruction of Near Infrared Band for Crop Health Monitoring under Variable Illumination Conditions

Alejandro Vidal-Domínguez[1], Robert Guamán-Rivera[1], Tito Arévalo-Ramírez[2], Elizabeth Lam-Esquenazi[3], Ranjit Das[1], and Oswaldo Menéndez-Granizo[1,*]

[1]Universidad Católica del Norte, Departamento de Ingeniería de Sistemas y Computación, Antofagasta 1249004, Chile
[2]Department of Electrical Engineering, Department of Mechanical Engineering and Metallurgy, School of Engineering, Faculty of Engineering, Pontificia Universidad Catolica de Chile, Santiago, Chile
[3]Universidad Católica del Norte, Departamento de Ingeniería Química y de Medio Ambiente, Antofagasta 1249004, Chile
[*]oswaldo.menendez@ucn.cl

ABSTRACT

Monitoring crop health is a pivotal pillar of sustainable agricultural production, requiring non-invasive sensing methods to prevent biotic and abiotic stress. Although multispectral and hyperspectral imaging has enabled advances in crop phenomics, its large-scale adoption remains limited by high costs, restricted spatial resolution, and strong sensitivity to lighting conditions. This work analyzes how varying illumination affects generative artificial intelligence models designed to reconstruct the near-infrared (NIR) band for agricultural and forestry applications. Two deep learning architectures were developed and evaluated: Convolutional Variational Autoencoders (CNN-VAEs) and Conditional Generative Adversarial Networks (cGANs). Each model was trained on aerial datasets collected over two forest sites under four shadow scenarios (without, minimal, short, and moderate) to assess their robustness across heterogeneous lighting conditions. The performance model was analyzed using SSIM, PSNR, and NRMSE to enable a standardized comparison between architectures. The experimental findings revealed that the multi-dataset VAE achieved the highest performance, reaching 0.96 SSIM, 35.06 dB PSNR, and an NRMSE of 0.018 under moderate shadows. These results demonstrate that VAE-based architectures provide more stable and reliable multispectral reconstructions than GANs under diverse illumination conditions.

Introduction

Over recent decades, climate change has progressed faster than anticipated, increasing the frequency and intensity of extreme weather events such as heat waves, heavy rainfall, droughts, and frosts[1]. These phenomena severely affect forest and agricultural systems, threatening global crop productivity through processes such as soil erosion, nutrient depletion, and increased yield uncertainty[2,3]. At the same time, agricultural production is projected to increase by 60–100% by 2050 to meet the nutritional demands of a growing population[4]. This dual challenge—adapting to a changing climate while simultaneously increasing productivity in terms of yield and quality—demands new strategies for the development of resilient and sustainable management practices in forest and agricultural systems. In this context, early detection of vegetation stress symptoms enables timely decision-making and can significantly reduce production losses[5].

Vegetation stress traits are commonly assessed using classical, labor-intensive, and time-consuming field-based methods, as well as remote sensing approaches. The latter have gained increasing attention due to their non-invasive nature and their ability to relate vegetation spectral reflectance to plant physiological status[6]. Vegetation spectral reflectance describes the interaction between plant canopies and different wavelengths of the electromagnetic spectrum[7]. In particular, remote sensing techniques exploit hyperspectral and multispectral information for large-scale vegetation monitoring, enabling the extraction of valuable insights into canopy structure, physiological condition, and early signs of stress or disease[6,8]. Nevertheless, vegetation assessment using hyperspectral or multispectral systems is not always feasible due to sensor availability and technical constraints. For instance, sensors capable of capturing detailed spectral responses are often deployed on space-borne platforms, which impose spatial and temporal limitations[9,10]. Conversely, sensors suitable for deployment on unmanned aerial platforms are typically band-limited, thereby restricting the spectral resolution available for vegetation analysis. Most of these sensors record vegetation information only in the blue (B), green (G), red (R), RedEdge, and near-infrared (NIR) bands[11].

Due to these limitations in sensor availability and spectral coverage, several studies have explored the inference of missing spectral information from available bands[12,13]. In particular, previous work has modeled inter-band reflectance correlations using deep learning (DL) approaches. Convolutional neural networks (CNNs) have been shown to effectively learn non-linear mappings between RGB inputs and spectral responses, achieving high reconstruction accuracy in controlled environments and within the visible spectrum[14–16]. To improve spatial consistency and preserve fine-scale details, encoder–decoder U-Net architectures with skip connections have been widely adopted, resulting in enhanced reconstruction fidelity[17]. More recently, attention-based models and Vision Transformers have demonstrated the ability to capture long-range spectral–spatial dependencies and achieve competitive performance on benchmark datasets and real-field scenarios[18]. However, several studies have reported that reconstruction accuracy degrades at longer wavelengths, with the near-infrared (NIR) band exhibiting higher spectral distortion and reduced robustness, as correlations with RGB inputs weaken and become more sensitive to radiometric inconsistencies, such as illumination variability and sensor calibration effects[19,20].

In parallel, generative models have been investigated to address uncertainty and incomplete spectral observations. Variational Autoencoders (VAEs) have emerged as probabilistic frameworks capable of learning compact and structured latent representations of heterogeneous environments, enabling effective dimensionality reduction, noise suppression, and robust feature extraction[21–23]. Recent studies have reported promising results for VAE-based models in agricultural imaging tasks and multispectral reconstruction pipelines, particularly when combined with convolutional backbones or hybrid architectures[24–26]. Generative Adversarial Networks (GANs) have also been employed to improve perceptual quality; however, previous studies have highlighted training instability and systematic spectral bias under varying illumination conditions, often resulting in darker or distorted NIR and red-edge reconstructions[19,27].

Despite the significant progress achieved by recent deep learning (DL)–based models, several open challenges remain in RGB-to-multispectral reconstruction. In particular, the analysis and reconstruction of spectral reflectance in outdoor environments are hindered by radiometric distortions caused by heterogeneous illumination conditions, shadows, and sensor limitations. Illumination variability directly affects passive optical imaging, upon which most hyperspectral and multispectral systems rely, thereby compromising spectral consistency and limiting the generalization capability of reconstruction models[20,28,29]. Nevertheless, only a limited number of studies explicitly aim to develop reconstruction models that preserve radiometric fidelity, which is essential for reliable multispectral feature extraction and vegetation monitoring under real-world field conditions. For instance, conditional generative adversarial network (cGAN)–based approaches have demonstrated that incorporating samples acquired under diverse lighting conditions can partially improve reconstruction performance for green, red-edge, and near-infrared (NIR) bands[19]. However, these methods often exhibit systematic radiometric distortions, such as darkened reconstructions, revealing limited robustness to illumination variability. More recent models based on residual networks, attention mechanisms, and advanced generative frameworks have reported improved reconstruction accuracy[30–32]. Nonetheless, these approaches typically assume stable illumination conditions or rely on controlled acquisition settings, which restricts their applicability in real-world outdoor scenarios. Consequently, robust multispectral reconstruction from RGB inputs under heterogeneous lighting conditions and variable data quality remains an open problem.

Previous studies indicate the need for a comprehensive evaluation of the robustness and stability of RGB-to-NIR reconstruction models under realistic and heterogeneous illumination conditions, including the quantification of performance dispersion, failure cases, and cross-domain generalization. In response to this gap, the present work provides a systematic analysis of the influence of illumination variability on RGB-to-NIR spectral reconstruction. Specifically, we quantify model performance and stability across diverse lighting scenarios and introduce a Variational Autoencoder (VAE)–based generative model for NIR reconstruction, benchmarking it against a pix2pix cGAN baseline under identical conditions. This comparative evaluation enables a rigorous assessment of the robustness and generalization capabilities of both architectures across varying illumination conditions. The proposed framework and findings offer new insights into performance dispersion under heterogeneous lighting across different solar incidence angles and acquisition environments, demonstrating that the VAE-based approach provides improved stability for NIR reconstruction in outdoor scenarios. These results support the development of illumination-aware and uncertainty-aware spectral reconstruction models for remote sensing applications.

The remainder of this work is organized as follows. Section 1 provides an overview of the generative image reconstruction algorithms, algorithms features and hardware specifications. Section 2 presents the results of the experiments performed to convert from RGB to NIR images. Section 3 includes a detailed discussion of the results obtained. Finally, Section 4 presents the conclusions of this work.

1 Material and Methods

This section describes the experimental design adopted in this work, including the data acquisition strategy, illumination conditions, and testing protocol. As illustrated in Fig. 1, the proposed pipeline is structured into four main stages: (i) multispectral data acquisition through aerial surveys under varying illumination conditions; (ii) VAE-CNN and cGAN architectures; (iii) construction and training of generative reconstruction models based on VAE-CNN and cGAN architectures; and (iv) quantitative performance evaluation of the generative models under different illumination scenarios.

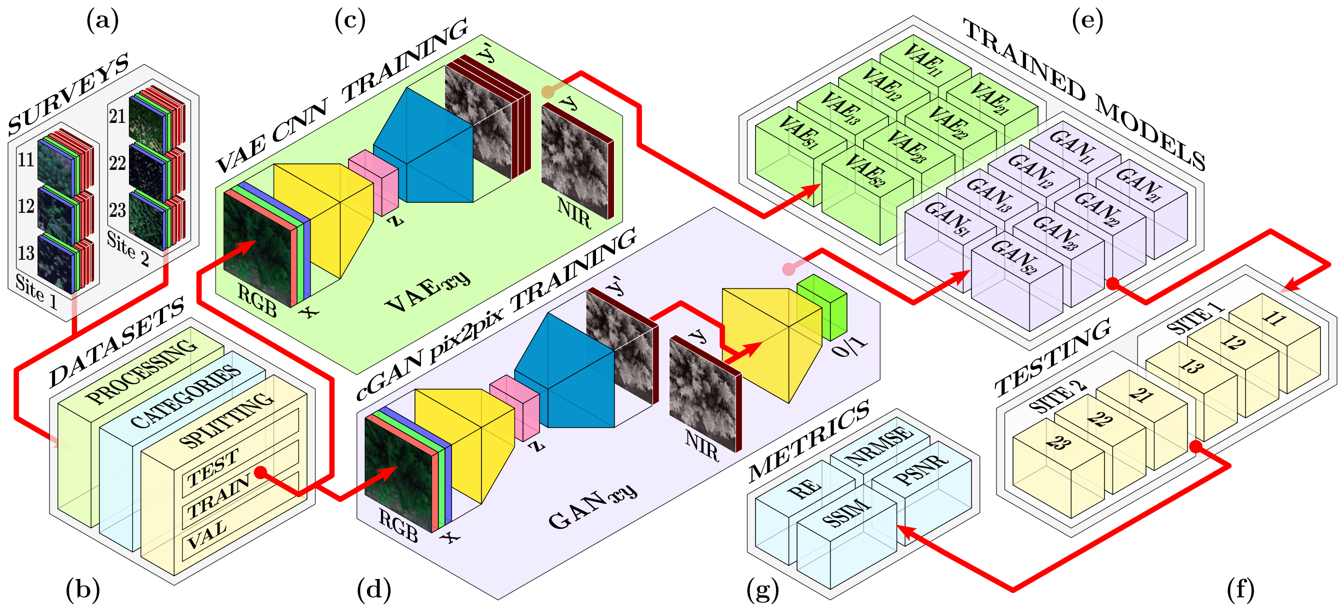

Figure 1. Representative illustation of the proposed methodology for NIR band reconstruction from RGB imagery. (a) Experimental surveys comprising six UAV acquisitions (three per study site) conducted under distinct illumination conditions. (b) Dataset preprocessing, including radiometric alignment and splitting into training and testing subsets. (c) Generative reconstruction frameworks: the VAE-CNN model and (d) the conditional GAN based on the pix2pix architecture. (e) Sets of trained models corresponding to each survey and study site. (f) Prediction of NIR images using the trained models on the test subsets. (g) Quantitative evaluation using SSIM, PSNR, and NRMSE metrics, together with relative error visualizations to assess reconstruction performance. Note: This pipeline integrates data preparation, model training, and performance analysis within a unified experimental framework.

1.1 Data acquisition

Multispectral imagery was acquired using a MicaSense RedEdge camera mounted on a Tarot T18 octocopter. Although the sensor captures five spectral bands, this study focuses on four bands, blue (475 nm), green (560 nm), red (668 nm), and near-infrared (840 nm), at a spatial resolution of 1280×960 pixels. All acquisitions were performed in a nadir configuration following the flight parameters reported in Table 1. Two forest study sites were selected, corresponding to a twenty-year-old (site 1) and a six-year-old (site 2) *Pinus insigne* plantation with native understory shrubs. Each study site covered an area of approximately 1 ha and was chosen to provide vegetation scenes with comparable structural characteristics while exhibiting different canopy maturity levels. The study sites are located at 35°13′19.3″S, 72°09′29.3″W and 35°13′14.2″S, 72°09′27.4″W, respectively, in the Maule region, Chile. All images are aligned with the tools and procedures provided by the camera manufacturer[33,34].

The quality of spectral reconstruction is highly sensitive to solar geometry during octocopter data acquisition. To provide a reproducible and quantitative framework, lighting scenarios are categorized based on the physical relationship between solar elevation (θ_s) and the projected shadow length (L) of an object with height (H), defined by the trigonometric identity $L = H\cot(\theta_s)$. Briefly,

- No shadows: Surveys conducted under overcast conditions where diffuse radiation predominates, resulting in negligible shadow projection.
- Minimal shadows: Solar elevation between 60° and 90°; in this range, the shadow is significantly smaller than the object's height ($L < 0.58H$), minimizing canopy occlusion.
- Short shadows: Solar elevation between 30° and 60°, representing a transitional state where the shadow length is comparable to the height ($0.58H \leq L < 1.73H$).
- Moderate shadows: Solar elevation between 10° and 30°, where the shadow length grows significantly ($1.73H \leq L \leq 5.67H$), increasing the radiometric complexity due to extensive partial shading.

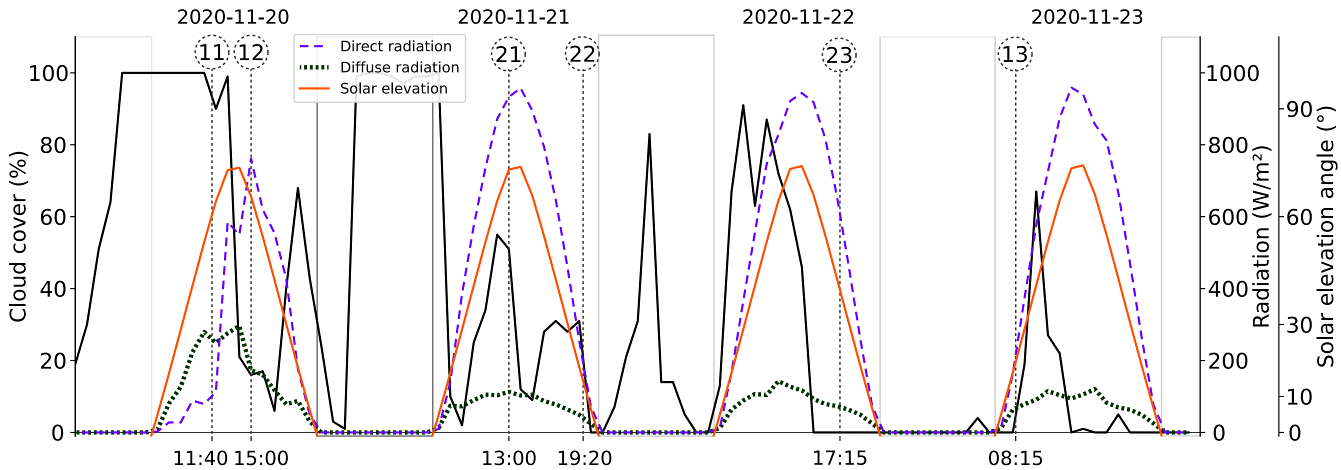

Figure 2. Timeline of the surveys, showing their capture timestamps (dotted black lines). The red line represents the solar elevation angle (degrees above the horizon), while the solid black line shows the cloud cover percentage. The short-dashed and dashed blue lines correspond to the diffuse and direct solar radiation, respectively. The black shaded areas indicate the shadow categories for each survey. Cloud cover, solar radiation, and solar elevation data were obtained from Open-Meteo[37].

Table 1. Collected data from aerial surveys conducted at the two study sites located in the Maule region, Chile. Note: Forest stands differ in age, affecting flight altitude and acquisition speed.

Study site	Survey	Altitude (m)		Speed (m s^{-1})	Captures	Day	Hour (GMT-3)	Shadow
		min	max					
S1	11	50	60	3–4	310	20	11:40	No
	12				324	20	15:00	Minimal
	13				403	23	08:15	Moderate
S2	21	30	40	4–5	440	21	13:00	Minimal
	22				440	21	19:20	Moderate
	23				539	22	17:15	Short

- Long shadows: Solar elevation between 0° and 10°; at these grazing angles, the shadow length exceeds 5.67 times the object's height. It should be noted that no samples were acquired under these conditions in the present dataset due to extreme spatial occlusion and low signal-to-noise ratios typical of near-horizon illumination.

All flights were conducted following a grid pattern at constant altitude, with the camera mounted in a nadir configuration. Data acquisition was performed under stable wind conditions (<10 m/s) to minimize motion-induced artifacts. Surveys were scheduled at different times of the day with the objective of capturing a range of illumination conditions, classified into three categories: no shadows, minimal shadows, and moderate shadows. These categories were defined based on solar elevation angle thresholds, as detailed in Fig. 2. Survey 11 was acquired under predominantly cloudy conditions; due to the dominance of diffuse solar radiation, it presents negligible shadowing and was therefore assigned to the no-shadow category. Surveys 12 and 21 were collected under minimal-shadow conditions, corresponding to high solar elevation angles (60° to 90°). Survey 23 was acquired under short-shadow conditions (30° to 60°), while Surveys 13 and 22 were obtained under moderate-shadow conditions associated with lower solar elevations (10° to 30°). No acquisitions were performed under long-shadow conditions (solar elevation 0° to 10°). This classification was adopted as it offers a systematic framework to organize the dataset based on shadow intensity and illumination geometry, consistent with established approaches in the remote sensing literature[35,36].

1.2 Generative Models

The reconstruction of the NIR band from RGB imagery is performed using two generative models: VAE-CNN and cGAN based on the pix2pix framework. To ensure a controlled comparison, all models are trained and evaluated under identical data splits, preprocessing steps, and illumination-specific training conditions.

1.2.1 VAE-CNN models

The VAE-based models implemented follow an encoder–decoder convolutional architecture designed to reconstruct the NIR band from RGB inputs through a probabilistic latent representation. Unlike deterministic reconstruction models, VAEs learn a distribution in latent space, allowing the model to capture uncertainty and variability induced by heterogeneous illumination conditions[38,39].

The encoder maps the input RGB image into a low-dimensional latent space characterized by a mean and variance, while the decoder reconstructs the target NIR image from samples drawn from this latent distribution. Convolutional layers are employed in both encoder and decoder to preserve spatial coherence and exploit local image structure. The training objective combines a reconstruction loss between the predicted and target NIR images with a Kullback–Leibler divergence term that regularizes the latent space toward a unit Gaussian distribution. This formulation promotes smooth latent representations and improves robustness to domain shifts caused by illumination variability.

The predicted outputs consist of three channels due to the use of a pre-trained feature loss network operating in a three-channel configuration. To enable direct comparison with the single-channel NIR reference images and the application of evaluation metrics, the three output channels were averaged to produce a single reconstructed NIR image.

1.2.2 GAN-pix2pix models

To preserve structural and semantic consistency between RGB inputs and reconstructed NIR images, encoder–decoder architectures with skip connections are commonly adopted in image-to-image translation tasks[40]. In particular, the U-Net architecture enables the retention of spatial information across network depths, making it suitable for spectral reconstruction problems[41]. In this context, the GAN-based baseline implemented follows the standard pix2pix framework[19].

The generator adopts a U-Net architecture composed of eight downsampling and eight upsampling blocks, with skip connections linking corresponding encoder and decoder layers. Downsampling blocks consist of two-dimensional convolutions, leaky rectified linear units, and batch normalization, while upsampling blocks are implemented using transposed convolutions, batch normalization, and rectified linear unit activations. This design allows the generator to combine high-level semantic features with fine-grained spatial information.

The discriminator corresponds to the default pix2pix discriminator. By operating at the patch level, this discriminator emphasizes local texture consistency and high-frequency detail rather than global image structure. The pix2pix model is used exclusively as the GAN-based baseline in this study, ensuring a controlled and consistent comparison with the VAE-CNN models under varying illumination conditions.

The selection of CNN-VAE and cGAN architectures for this work is based on their complementary strengths and proven performance in multispectral reconstruction tasks. First, CNN-VAEs were chosen for their ability to learn structured latent representations that ensure global radiometric coherence and stability across heterogeneous scenes. In contrast, cGANs were included to leverage adversarial training for enhancing sharp localized textural details, providing a comprehensive comparison between global consistency and local fidelity. Even though more recent generative frameworks such as Diffusion Models or Normalization Flows offer high-quality synthesis, they were excluded due to their prohibitive computational costs—a critical barrier for practical monitoring applications—and the need to isolate the effects of solar elevation using well-characterized baselines. This choice ensures that performance deviations are primarily attributable to illumination geometry rather than the training instabilities often associated with more complex generative paradigms.

1.3 Models training and testing

The NIR band reconstruction process is addressed using sixteen generative models, comprising eight VAE-based architectures and eight GAN-based architectures based on the pix2pix framework. Each model is trained independently using data from a single survey, enabling illumination-specific learning and subsequent cross-survey evaluation under mismatched lighting conditions. Briefly,

- A custom dataloader is implemented to pair RGB images as inputs with their corresponding single-band NIR images as reconstruction targets. To ensure architectural compatibility and reduced computational complexity across both generative frameworks, all images are resized to 256×256 pixels before to training. Although this reduction reduces fine-scale spatial detail, it ensure fair and consistent comparison between models.

- Data augmentation process is applied to the training images to enhance spatial generalization. The augmentation strategy consisted of three square crops (centre, left, and right), each subjected to four rotations (0°, 90°, 180°, and 270°) and their corresponding horizontal mirror transformations. No illumination-specific augmentation is applied, enabling the models to learn illumination effects directly from the data.

- The dataset distribution for training (80%), validation (10%), and testing (10%) is summarized in Table 2. For the VAE-CNN models, the reconstruction loss was adapted to minimize the discrepancy between the predicted output and the

Table 2. Dataset split for training, testing and validation to train every prediction model. Every survey has its own trained model by multispectral band, allowing cross validation between surveys later on.

Study Site	Survey	Multispectral Captures		
		RGB-to-NIR		
		80% Train	10% Test	10% Validation
S1	11	5,952 (248)	31	31
	12	6,192 (258)	33	33
	13	7,704 (321)	41	41
S2	21	8,448 (352)	44	44
	22	8,448 (352)	44	44
	23	10,344 (431)	54	54

target NIR band rather than reconstructing the RGB input, as in standard autoencoder formulations. For the GAN-based pix2pix models, paired RGB–NIR datasets were generated prior to training using the preprocessing tools provided by the original authors[42].

- As illustrated in Fig. 1, each survey at each study site is associated with a dedicated generative model trained exclusively on data from that survey. This design enables a controlled analysis of illumination-domain specificity and facilitates cross-survey evaluation, where models trained under one illumination condition are tested on datasets acquired under different shadow regimes.
- All models are trained and tested on a high-performance workstation equipped with an Intel Core i7-12650H CPU (10 cores, 3.5/4.7 GHz), an NVIDIA RTX 3070M GPU with 8 GB GDDR6 memory, and running Ubuntu 22.04 LTS. After training, each model was used to predict the test split of every survey dataset, yielding six test datasets for the NIR band.

Beyond training illumination-specific models for each individual survey, the experimental design implicitly incorporates a leave-one-survey-out evaluation within each shadow category. Because multiple surveys were acquired under comparable illumination conditions (i.e., no-shadow, minimal-shadow, and moderate-shadow), models trained on one survey are systematically evaluated on other surveys belonging to the same shadow class. This configuration enables the assessment of robustness to intra-class variability arising from temporal, atmospheric, and acquisition-related factors, while preserving comparable solar geometry.

In addition, site-specific models were trained by combining three surveys acquired at the same study site under distinct illumination conditions. This configuration allows the evaluation of illumination-invariant reconstruction performance within a fixed structural and ecological context, effectively disentangling illumination effects from canopy maturity and vegetation structure. As a result, the proposed framework supports a multi-level generalization analysis, including survey-specific, shadow-class-specific, and site-specific illumination variability, without relying on conventional k-fold cross-validation schemes that would mix illumination domains and obscure the targeted analysis.

1.4 Evaluation metrics

Following the guidelines described in[19], the performance of each generative model is evaluated using three metrics: **(i)** Peak signal to noise ratio (PSNR); **(ii)** Structural similarity index (SSIM); and **(iii)** Normalised root mean square error (NRMSE).

The PSNR measures the ratio between the maximum possible signal intensity (MAX) and the mean squared error (MSE) between the original and reconstructed images, as defined in (1). Note that, higher PSNR values show lower reconstruction error and, therefore, higher reconstruction fidelity.

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{1}$$

The NRMSE measures the square mean of the MSE normalised by the difference between the maximum and minimum (MIN) value. In (2), a lower NRMSE value indicates a smaller discrepancy between the original and reconstructed image.

$$\mathrm{NRMSE} = \frac{\sqrt{\mathrm{MSE}}}{\mathrm{MAX} - \mathrm{MIN}} \tag{2}$$

The SSIM evaluates the perceptual similarity between original and reconstructed images by measuring luminance, contrast, and structural consistency. Unlike PSNR and NRMSE that quantify pixel-wise radiometric errors, SSIM evaluates the preservation of structural information under varying illumination conditions in this work. The SSIM is computed as follows,

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{3}$$

where μ_x and μ_y are local mean intensities of the refence and reconstructed images, respectively. ${\sigma_x}^2$ and ${\sigma_y}^2$ represent the local variances. C_1 and C_2 are constants used to stabilize the division under low-variance conditions.

While this work focuses on pixel-wise metrics (SSIM, PSNR, and NRMSE) to enable direct comparison with previous works, we acknowledge that such metrics may not fully capture perceptual or spectral consistency in multispectral reconstruction tasks.

Spectral metrics such as the Spectral Angle Mapper (SAM) are widely used in remote sensing to evaluate spectral fidelity by measuring the angular difference between spectral vectors, independently of absolute magnitude. For a reconstructed spectral vector $\mathbf{g}$ and its reference counterpart $\mathbf{t}$, SAM is defined as:

$$\text{SAM}(\mathbf{g},\mathbf{t}) = \arccos\left(\frac{\mathbf{g}\cdot\mathbf{t}}{\|\mathbf{g}\|_2\,\|\mathbf{t}\|_2}\right). \tag{4}$$

However, in the present work, the reconstruction task focuses on a single spectral band, for which SAM becomes degenerate and offers limited discriminative power. Consequently, SAM is not included in the quantitative or qualitative evaluation, but is identified as a relevant complementary metric for future extensions involving multi-band or hyperspectral reconstruction.

1.5 Spatial error quantification

In addition to global image-level metrics, a spatial analysis of reconstruction errors was conducted using relative error maps. For each reconstructed sample, the relative error was computed as

$$RE = \frac{(I_{\text{gen}} - I_{\text{tgt}})}{(|I_{\text{tgt}}| + \varepsilon)} \tag{5}$$

where I_{gen} and I_{tgt} denote the generated and target NIR images, respectively, and ε prevents division by zero.

From each relative error map, summary statistics were extracted to characterize spatial error patterns, including the mean and median relative error (to assess systematic bias), the mean absolute relative error (MAE) to quantify error magnitude, and the standard deviation to capture spatial dispersion. For visualization purposes, error maps were displayed using a diverging color scale centered at zero, with extreme values clipped at the 99th percentile.

It is worth noting that relative error metrics are highly sensitive to low-intensity target values; therefore, spatial relative error statistics are interpreted in conjunction with absolute image-level metrics and visual inspection.

2 Experimental Results

Figure 3 presents the reconstruction performance of VAE-CNN and GAN-pix2pix models under different shadow conditions using SSIM, PSNR, and NRMSE metrics. VAE-based models consistently exhibit more compact interquartile ranges, indicating lower variability and improved robustness to illumination changes. Although GAN-based models occasionally achieve higher maximum SSIM and PSNR values (see Fig. 3a and Fig. 3b), they also present substantially higher dispersion and lower minimum values, particularly under moderate shadow conditions. This behaviour suggests that adversarial training enhances local perceptual sharpness at the expense of increased sensitivity to illumination-induced domain shifts. Furthermore, this effect is supported by the NRMSE distributions shown in Fig. 3c, where GAN-based models exhibit a larger number of high-error outliers, while VAE-based models maintain more stable error bounds across all shadow categories. Overall, these results indicate that VAE-based architectures provide more reliable and illumination-robust NIR reconstructions.

It is possible to see that reconstruction metrics improves as shadow intensity increases across all trained models; however, this performance improvement is accompanied by increased variability, particularly under moderate shadow conditions. In terms of SSIM (see Fig. 3a), both models exhibit their greatest dispersion in the moderate shadow category, where VAE-based models achieve the highest SSIM value of 0.959, while GAN-based models reach the lowest value of 0.132. Regarding PSNR (see Fig. 3b), moderate shadows yield the highest PSNR for the VAE models (35.06 dB), whereas GAN-based models exhibit their poorest performance under minimal shadow conditions (6.64 dB). Finally, analysing the NRMSE (see Fig. 3c), both

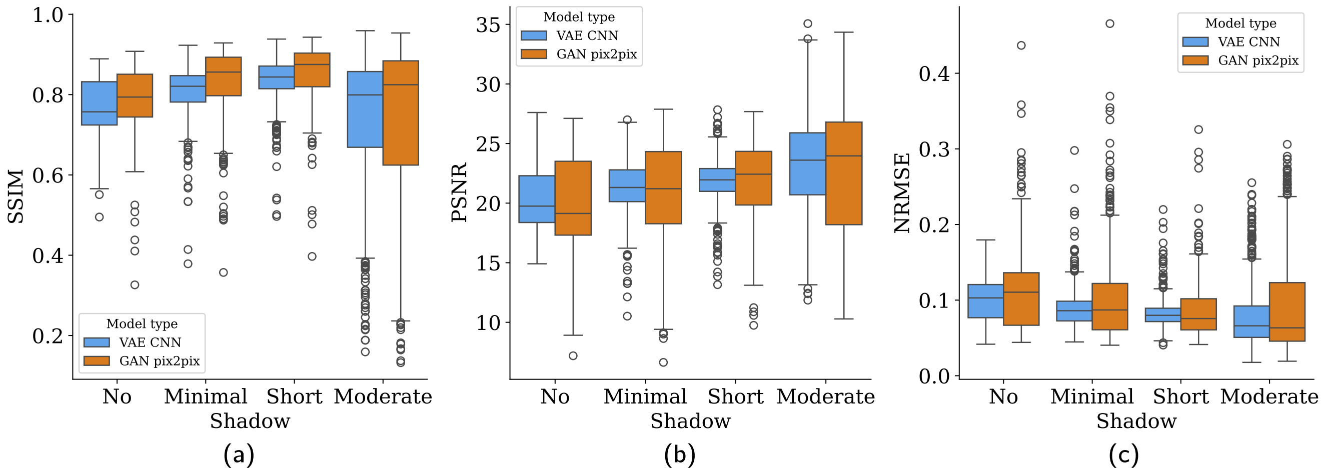

Figure 3. Main findings for NIR multispectral band reconstructions obtained with VAE–CNN and GAN–pix2pix models under different shadow conditions. **(a)** SSIM, **(b)** PSNR, and **(c)** NRMSE. The boxplots summarize the variability and comparative performance of both model types across the defined shadow categories.

Table 3. Summary of SSIM results for the VAE-CNN and GAN-pix2pix models in NIR multispectral band reconstruction. Columns report the maximum (*Max*), minimum (*Min*), upper and lower outliers (O^+ and O^-), median with interquartile range (*Median* ($Q1$–$Q3$)), and standard deviation (σ). Note: The highest value in each column is highlighted in bold, and the lowest is underlined.

Shadow (samples)	Best model: SSIM↑ for NIR band							Worst model: SSIM↑ for NIR band						
	Model	*Max*↑	O^+	*Median*($Q1$-$Q3$)	O^-	σ↓	*Min*↑	Model	*Max*↑	O^+	*Median*($Q1$-$Q3$)	O^-	σ↓	*Min*↑
No (31)	GAN_{11}	0.91	0	0.88 (**0.88-0.89**)	5	0.02	0.82	GAN_{22}	0.75	0	0.66 (0.62-0.70)	5	0.10	0.33
	VAE_{11}	0.89	0	0.86 (0.85-0.87)	1	**0.02**	0.80	VAE_{22}	0.78	0	0.72 (**0.69-0.73**)	7	**0.07**	0.50
Minimal (77)	GAN_{21}	0.93	0	0.88 (0.79-0.92)	0	0.07	0.70	GAN_{22}	0.88	0	0.76 (0.68-0.84)	1	0.12	0.36
	VAE_{S2}	0.92	0	0.89 (0.86-0.90)	2	0.03	0.79	VAE_{22}	0.84	0	0.77 (0.74-0.81)	9	0.09	0.38
Short (54)	GAN_{23}	0.94	0	**0.92** (0.90-0.93)	3	0.02	**0.85**	GAN_{22}	**0.89**	0	**0.83** (0.74-0.86)	4	0.11	0.40
	VAE_{S2}	0.94	0	0.90 (0.89-0.92)	2	0.02	0.80	VAE_{22}	0.86	0	0.82 (0.74-0.85)	4	0.09	**0.50**
Moderate (85)	GAN_{22}	0.95	0	0.91 (0.87-0.94)	1	0.04	0.74	GAN_{11}	0.72	0	0.43 (0.36-0.59)	0	0.15	0.13
	VAE_{S2}	**0.96**	0	0.90 (0.85-0.95)	0	0.06	0.75	VAE_{11}	0.82	0	0.51 (0.42-0.68)	0	0.18	0.16

approaches attain comparable minimum values across shadow categories. Nevertheless, the VAE models achieve the lowest error of 0.018 under moderate shadows, while the GAN models record the highest error of 0.466 under minimal shadows. Finally, GAN-based models show a larger number of high-error outliers in NRMSE, indicating reduced stability and increased sensitivity to illumination variability.

Tables 3 to 5 present a detailed statistical analysis of the best and worst-performing models per class. It is worth noting the three metrics presented similar patterns for both models. Thus, the pooled dataset model (VAE_{S2}) achieve the best overall performance with a 0.96 SSIM, 35.06 db PSNR, and 0.018 NRMSE under moderate shadows, as shown in Figs. 4 to 6. Although, GAN-based models achieve similarly high medians with 28.13 dB PSNR and 0.039 NRMSE, they present higher dispersion, which confirms that adversarial training enhances local detail at the expense of global consistency (reduced consistency across samples). In contrast, it is possible to note that models trained on individual surveys present more irregular generalization. In particular, VAE_{22} and GAN_{22} (PSNR of 6.64 dB and NRMSE of 0.466) are generally the worst models. In addition, models models trained on Survey 11 (VAE_{11} and GAN_{11}) present pronounced degradation when evaluated under moderate shadows despite stronger performance in no-shadow conditions. These results underscore the importance of illumination-domain alignment between training and evaluation.

Under no-shadow conditions, model S1 trained on Surveys 11, 12, and 13 achieved the highest medians and most compact interquartile ranges, indicating internal coherence and strong generalization within their illumination domain. However, their performance progressively declines with increased shadow intensity, especially under moderate shadows, where all single-survey models present higher variability and reduced reconstruction fidelity. Across illumination scenarios, VAE-

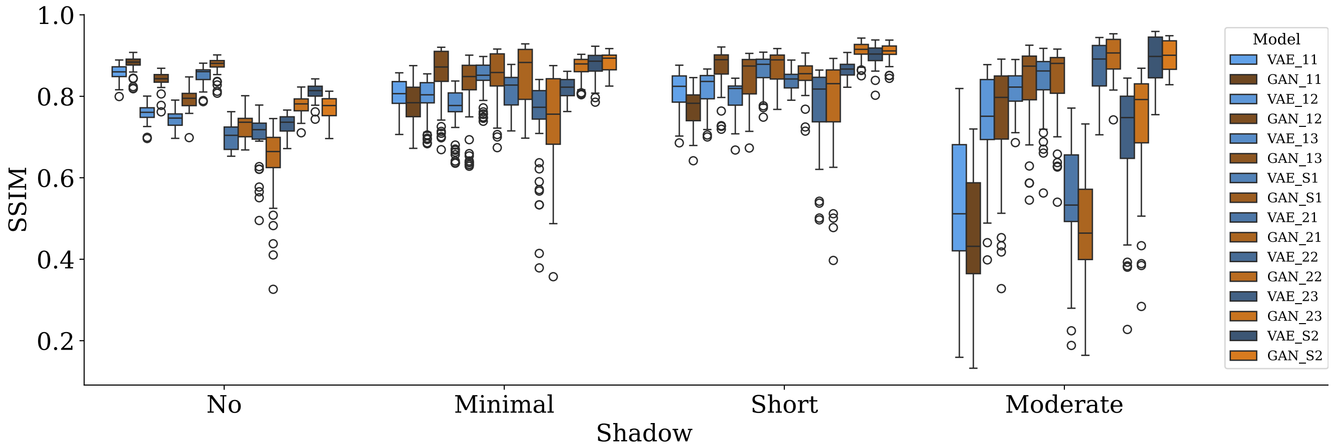

Figure 4. Boxplots of SSIM values for the VAE-CNN and GAN-pix2pix models in NIR band reconstruction from RGB imagery. Note: Models are grouped by shadow category (no, minimal, short, and moderate), and each VAE model is paired with its corresponding GAN model trained on the same dataset to enable direct performance comparison.

Table 4. Summary of PSNR results for the VAE-CNN and GAN-pix2pix models in NIR multispectral band reconstruction. Columns report the maximum (*Max*), minimum (*Min*), upper and lower outliers (O^+ and O^-), median with interquartile range (*Median* ($Q1$–$Q3$)), and standard deviation (σ). Note: The highest value in each column is highlighted in bold, and the lowest is underlined.

Shadow	**Best model: PSNR↑ for NIR band**							**Worst model: PSNR↑ for NIR band**						
(samples)	**Model**	*Max*↑	O^+	*Median*($Q1$-$Q3$)	O^-	$\sigma\downarrow$	*Min*↑	**Model**	*Max*↑	O^+	*Median*($Q1$-$Q3$)	O^-	$\sigma\downarrow$	*Min*↑
No	GAN$_{11}$	27.11	1	25.12 (24.70-25.49)	1	**0.77**	**23.50**	GAN$_{22}$	<u>15.98</u>	0	<u>12.62</u> (11.37-13.55)	1	2.02	7.19
(31)	VAE$_{11}$	27.60	1	24.35 **(23.91-24.65)**	<u>6</u>	1.33	21.34	VAE$_{22}$	20.61	0	18.24 **(17.54-18.77)**	3	**1.26**	**14.91**
Minimal	GAN$_{12}$	27.88	0	<u>22.11</u> <u>(18.02-26.14)</u>	0	<u>4.44</u>	<u>12.45</u>	GAN$_{22}$	23.15	0	15.61 <u>(13.02-18.94)</u>	0	<u>3.97</u>	<u>6.64</u>
(77)	VAE$_{S2}$	<u>26.99</u>	0	23.48 (21.26-24.83)	1	2.45	15.53	VAE$_{22}$	23.47	0	20.52 (19.81-21.48)	<u>9</u>	2.65	10.52
Short	GAN$_{23}$	27.68	0	25.43 (24.86-26.24)	4	1.41	21.38	GAN$_{22}$	**24.05**	0	18.39 (15.92-20.53)	0	3.54	9.75
(54)	VAE$_{S2}$	27.84	0	23.88 (22.04-25.16)	0	2.20	17.93	VAE$_{22}$	22.99	0	**20.99** (18.85-22.09)	2	2.44	13.16
Moderate	GAN$_{S2}$	34.34	0	**28.13** (25.15-30.40)	0	3.27	21.65	GAN$_{11}$	18.61	0	13.24 (12.07-15.20)	0	2.00	10.28
(85)	VAE$_{S2}$	**35.06**	0	27.29 (24.41-30.04)	0	3.91	20.03	VAE$_{11}$	23.33	0	16.47 (14.70-18.94)	0	2.94	11.86

based models consistently outperform GAN-based models in preserving global radiometric balance and structural coherence under heterogeneous lighting conditions. In contrast, GAN-based models tend to exhibit greater sensitivity to illumination inconsistencies, leading to broader metric distributions. In particular, VAE$_{11}$, GAN$_{11}$, VAE$_{21}$, and GAN$_{21}$ models are the lowest performers in this condition. A comparative analysis across metrics confirms that VAE-based models maintain narrower performance distributions, higher minimum PSNR values, and lower NRMSE dispersion, showing greater overall stability. Although, GAN models are capable of locally generating sharper reconstruction (e.g., The GAN$_{23}$ model achieves an SSIM of 0.92 under short shadows category), they present increased variability and reduced robustness when illumination conditions deviate from those observed during training.

Using the spatial error statistics defined in Section 1.5, Figures 7 and 8 provide a qualitative and quantitative comparison of the spatial distribution of reconstruction errors.

Visual comparisons in Figs. 7 and 8 complement the quantitative results. In best-performing cases, relative error maps present slight overestimation (see red pixels) over vegetated regions and underestimation in shaded areas (see blue pixels), indicating that both architectures capture broad radiometric patterns but remain challenged by fine-scale shadow compensation. On the other hand, GAN-based models tend to systematically underestimate NIR intensity in the no-, minimal-, and short-shadow categories in worst-performing cases. In addition, VAE-based models present a bias toward underestimating non-vegetated areas and overestimating vegetation, especially under no and short shadows. Under moderate shadows, both architectures mainly overestimate shaded vegetation. Also, illuminated vegetation is reconstructed more accurately, as indicated by near-zero relative error regions (see white pixels).

In the best-performing cases (Fig. 7), GAN-based models under no-shadow conditions exhibit near-zero mean relative error

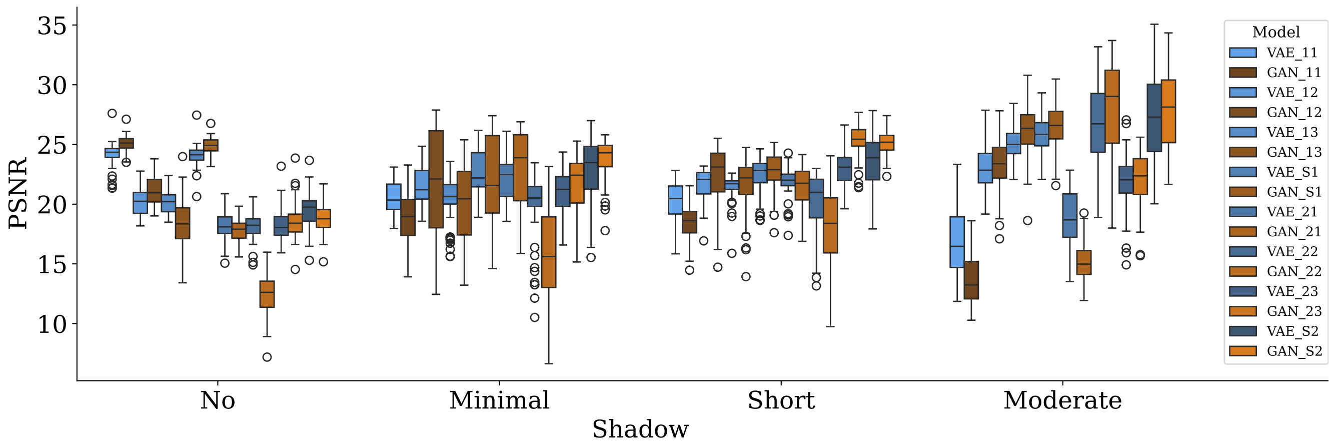

Figure 5. Boxplots of PSNR values for the VAE-CNN and GAN-pix2pix models in NIR band reconstruction from RGB imagery. Note: Models are grouped by shadow category (no, minimal, short, and moderate), and each VAE model is paired with its corresponding GAN model trained on the same dataset to enable direct performance comparison.

Table 5. Summary of NRMSE results for the VAE-CNN and GAN-pix2pix models in NIR multispectral band reconstruction. Columns report the maximum (*Max*), minimum (*Min*), upper and lower outliers (O^+ and O^-), median with interquartile range (*Median* ($Q1$–$Q3$)), and standard deviation (σ). Note: The highest value in each column is highlighted in bold, and the lowest is underlined.

Shadow (samples)	**Best model: NRMSE↓ for NIR band**							**Worst model: NRMSE↓ for NIR band**						
	Model	$Min\downarrow$	O^-	$Median(Q1\text{-}Q3)$	O^+	$\sigma\downarrow$	$Max\downarrow$	**Model**	$Min\downarrow$	O^-	$Median(Q1\text{-}Q3)$	O^+	$\sigma\downarrow$	$Max\downarrow$
No	GAN_{11}	0.044	1	0.055 **(0.053-0.058)**	2	0.005	0.067	GAN_{22}	<u>0.159</u>	0	<u>0.234</u> (0.210-0.270)	1	0.060	0.437
(31)	VAE_{11}	0.042	1	0.061 (0.059-0.064)	<u>6</u>	**0.010**	0.086	VAE_{22}	0.093	0	0.123 (0.115-0.133)	3	**0.019**	**0.180**
Minimal	GAN_{12}	0.040	0	<u>0.078</u> <u>(0.049-0.126)</u>	0	<u>0.050</u>	<u>0.238</u>	GAN_{22}	0.070	0	0.166 <u>(0.113-0.223)</u>	1	<u>0.082</u>	<u>0.466</u>
(77)	VAE_{S2}	<u>0.045</u>	0	0.067 (0.057-0.086)	3	0.024	0.167	VAE_{22}	0.067	0	0.094 **(0.084-0.102)**	<u>9</u>	0.043	0.298
Short	GAN_{23}	0.041	0	0.053 (0.049-0.057)	5	0.010	0.085	GAN_{22}	**0.063**	0	**0.120** (0.094-0.160)	4	0.061	0.326
(54)	VAE_{S2}	0.041	0	0.064 (0.055-0.079)	1	0.018	0.127	VAE_{22}	0.071	0	0.089 (0.079-0.114)	4	0.034	0.220
Moderate	GAN_{S2}	0.019	0	**0.039** (0.030-0.055)	0	0.016	**0.083**	GAN_{11}	0.117	0	0.218 (0.174-0.249)	0	0.046	0.306
(85)	VAE_{S2}	**0.018**	0	0.043 (0.031-0.060)	0	0.022	0.100	VAE_{11}	0.068	0	0.150 (0.113-0.184)	0	0.047	0.255

and minimal spatial dispersion. Specifically, GAN_{11} achieves a mean relative error close to zero ($\mu_{RE} = -0.008$) with very low variability ($\sigma_{RE} < 0.2$), indicating negligible systematic bias and stable spatial behaviour.

In contrast, the corresponding VAE-based model (VAE_{11}) presents substantially larger relative error magnitudes, despite a median relative error close to zero. The elevated mean and dispersion values observed for VAE-based models in these cases are primarily attributed to relative error amplification in low-intensity NIR regions rather than to large absolute reconstruction errors. Under minimal and short shadow conditions, best-performing VAE-based models trained on pooled datasets (VAE_{S2}) exhibit increased relative error magnitudes, with mean values exceeding 70 and 200, respectively. Nevertheless, median relative error values remain close to zero, indicating the absence of strong global bias. GAN-based counterparts (GAN_{21} and GAN_{23}) show lower mean relative error values and reduced dispersion, although with increased sensitivity to illumination variability compared to no-shadow scenarios. Under moderate shadow conditions, the best-performing VAE-based configuration (VAE_{S2}) exhibits a marked reduction in relative error magnitude ($\mu_{RE} = 16.501$) and spatial dispersion, outperforming the corresponding GAN-based model. This result indicates improved robustness of VAE-based architectures when training and evaluation illumination conditions are better aligned.

The worst-performing cases shown in Fig. 8 reveal a pronounced degradation in spatial error behaviour for both model families. VAE-based models trained on single surveys exhibit extreme amplification of relative error, particularly under moderate shadows. For instance, VAE_{11} reaches a mean relative error exceeding 500, with spatial dispersion on the order of 10^5, reflecting severe instability of the relative error metric in deeply shadowed regions. GAN-based worst-performing models generally present lower relative error magnitudes than their VAE counterparts but exhibit consistent negative bias across illumination conditions. Under no, minimal, and short shadow categories, GAN_{22} presents median relative errors below -0.5,

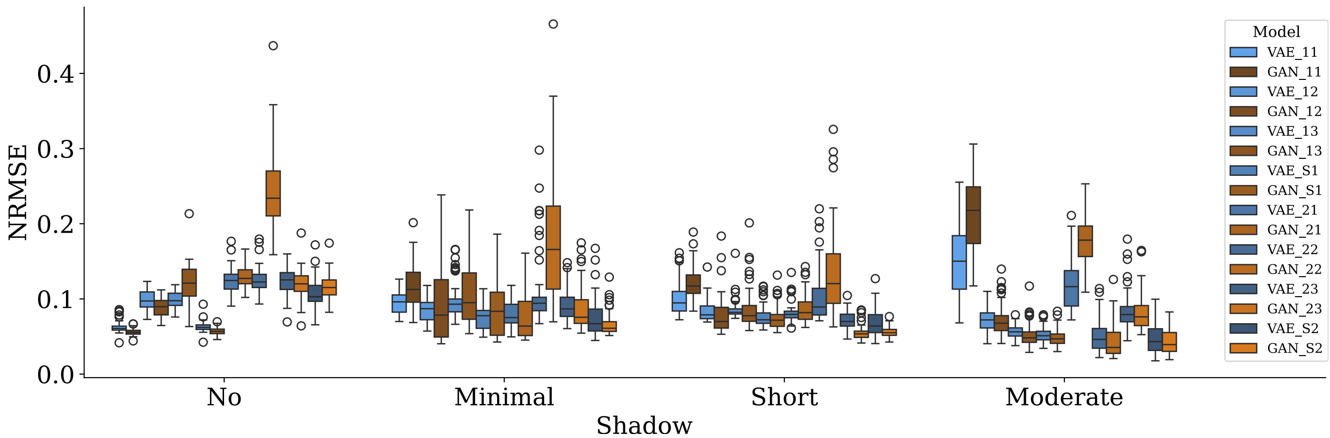

Figure 6. Boxplots of NRMSE values for the VAE-CNN and GAN-pix2pix models in NIR band reconstruction from RGB imagery. Note: Models are grouped by shadow category (no, minimal, short, and moderate), and each VAE model is paired with its corresponding GAN model trained on the same dataset to enable direct performance comparison.

indicating systematic underestimation of NIR reflectance. Under moderate shadows, both architectures exhibit elevated relative error values; however, VAE-based models display significantly higher dispersion, confirming their increased sensitivity of relative error metrics to near-zero target values.

Overall, the NIR spatial error analysis confirms that relative error metrics are highly sensitive to low-reflectance regions and should be interpreted jointly with absolute reconstruction metrics and visual inspection. GAN-based models tend to exhibit lower relative error dispersion under favourable illumination conditions but are more prone to systematic bias, whereas VAE-based models preserve global radiometric structure at the expense of amplified relative error under strong shadowing. These trends are consistent with the behaviour observed in the RedEdge reconstruction task and reinforce the importance of illumination-aware training data for robust multispectral reconstruction.

Overall, the combined results from SSIM, PSNR, NRMSE, and relative error visualizations confirm that VAE-based architectures provide more stable and reliable reconstructions across varying illumination scenarios, maintaining consistent structural and luminance fidelity. GAN-based models, while occasionally achieving locally higher detail, remain more susceptible to dataset bias and illumination variability. These findings underscore the importance of dataset diversity and domain balance to improve generalization and robustness in multispectral generative modeling.

3 Discussion

Within the evaluated datasets, VAE-based models consistently outperformed GAN-based models in terms of reconstruction stability, exhibiting lower performance dispersion and fewer extreme failure cases across all illumination scenarios. In contrast, GAN-based models showed higher variability under illumination-induced domain mismatch; however, under favourable conditions, they were capable of generating sharper local textures. These observations indicate that both architectures capture complementary aspects of the NIR band reconstruction problem, with VAEs prioritizing global structural and radiometric coherence, and GANs enhancing localized textural details.

A key observation emerging from the analysis is the influence of shadow intensity on reconstruction performance. A progressive improvement in SSIM and PSNR, accompanied by a reduction in NRMSE as shadow coverage increases, suggests that both architectures adapt more effectively to moderate-shadow scenarios. A plausible explanation for this behaviour lies in the geometric and radiometric properties of shadowed regions. As shadows become short to moderate, they tend to be more spatially coherent and radiometrically homogeneous. Given that multispectral bands such as NIR primarily capture surface reflectance, increased shadow coverage reduces high-frequency radiometric variation and saturation effects, effectively lowering the amount of discriminative spectral information that must be inferred from the RGB input. Consequently, regions under shadow exhibit smaller relative differences between RGB and NIR responses, simplifying the reconstruction task for the generative models.

This hypothesis is further supported by the training–evaluation consistency observed across datasets. The best results under moderate-shadow conditions were predominantly achieved by models trained on datasets acquired under minimal to moderate shadows, indicating that exposure to similar illumination geometries during training improves generalization. In contrast, the poorest performance under moderate shadows corresponded to models trained exclusively under no-shadow conditions,

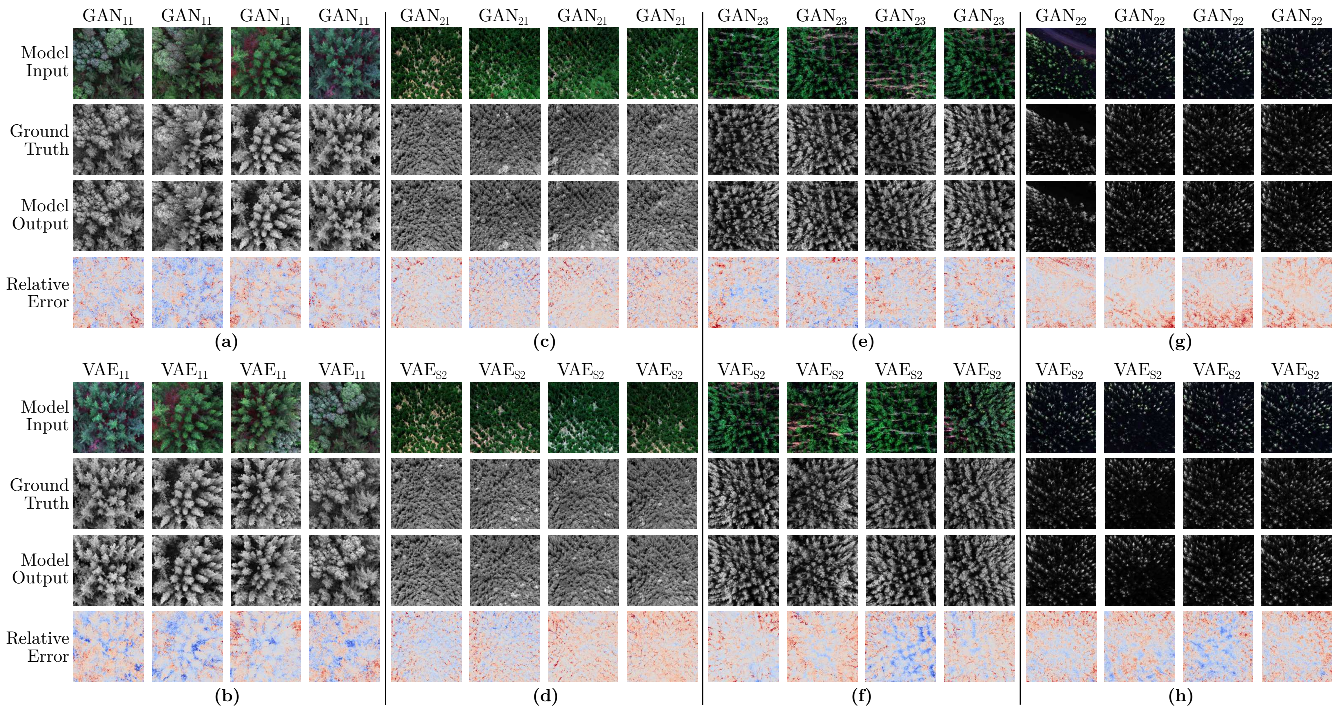

Figure 7. Visual comparison of the best-performing cases for each shadow category in the reconstruction of the NIR multispectral band from RGB input. Each column pair corresponds to the GAN-pix2pix (a,c,e,g) and VAE-CNN models (b,d,f,h) trained on the dataset associated with the respective shadow condition: (a,b) no shadows, (c,d) minimal shadows, (e,f) short shadows, and (g,h) moderate shadows. Within each subfigure, the four rows represent: RGB input, target NIR image, generated NIR reconstruction, and relative error map. The relative error maps are displayed using a diverging color scale centered at zero, where white denotes accurate reconstruction, blue indicates underestimation, and red represents overestimation with respect to the target NIR intensity. Overestimations are predominantly observed over vegetated regions, whereas underestimations tend to occur in shaded or low-intensity areas.

providing empirical evidence that illumination-domain mismatch negatively impacts reconstruction accuracy. These findings suggest that shadow-aware data diversity plays a critical role in learning robust spectral mappings.

The SSIM analysis reinforces this interpretation by revealing a systematic advantage for VAE-based models in terms of structural robustness. Higher minimum and maximum SSIM values across all shadow categories indicate that VAEs better preserve spatial organization under varying illumination. Although several GAN variants achieved competitive median SSIM values in specific scenarios, their broader interquartile ranges reveal greater sensitivity to illumination changes and dataset bias. In particular, models trained on Survey 22 consistently exhibited degraded performance, highlighting the limitations imposed by restricted illumination variability and reduced generalization capacity. Conversely, models S1, trained on Surveys 11, 12, and 13—acquired at the same site—performed best under no-shadow conditions, underscoring the importance of illumination-domain consistency between training and evaluation data.

Although SSIM, PSNR, and NRMSE are widely used for assessing image reconstruction quality, they are primarily pixel-based metrics and may not fully capture perceptual consistency or spectral fidelity, particularly in multispectral remote sensing applications. Metrics such as the Spectral Angle Mapper (SAM) or spectral information divergence could provide complementary insights by evaluating spectral similarity independently of absolute intensity differences. Future work will explore the integration of such metrics to better assess the ecological relevance of reconstructed NIR information.

The absence of spectral angle-based metrics in the reported results is a direct consequence of the single-band reconstruction setting, in which angular spectral measures do not provide additional discriminative information beyond pixel-wise errors.

Relative error visualizations provide additional spatial insight into reconstruction behaviour. Best-performing models preserve fine spatial details with limited deviations, although systematic biases remain. VAE-based models tend to slightly overestimate vegetation reflectance while underestimating shaded or darker regions, yet maintain coherent global structure. In worst-performing cases, GAN-based reconstructions systematically underestimate NIR reflectance under no, minimal, and short shadow conditions, whereas VAE models exhibit an inverse tendency, overestimating vegetation and underestimating non-vegetation areas. Under moderate shadows, both architectures predominantly overestimate shaded vegetation while maintaining

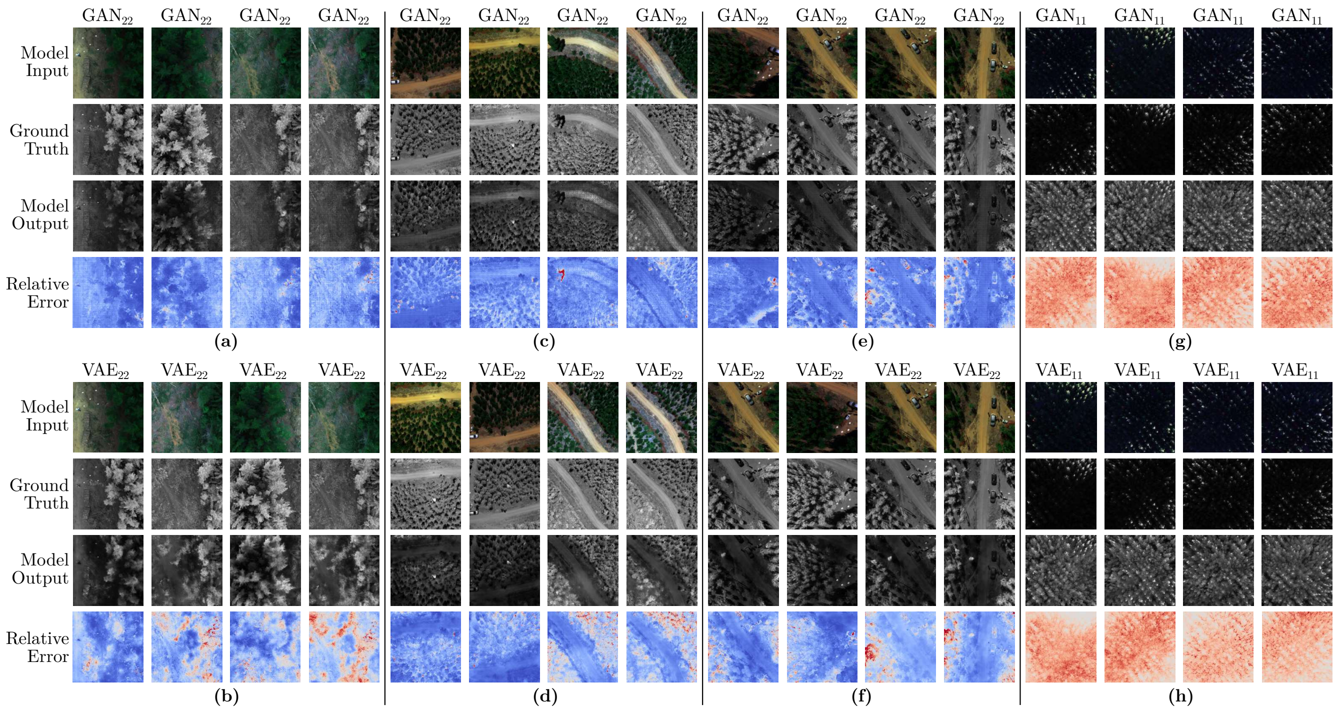

Figure 8. Visual comparison of the worst-performing cases for each shadow category in the reconstruction of the NIR multispectral band from RGB input. Each column pair corresponds to the GAN-pix2pix (a,c,e,g) and VAE-CNN models (b,d,f,h) trained on the dataset associated with the respective shadow condition: (a,b) no shadows, (c,d) minimal shadows, (e,f) short shadows, and (g,h) moderate shadows. The relative error maps reveal pronounced spatial error patterns in the worst-performing cases. For the no, minimal, and short shadow categories, both model families exhibit predominantly negative relative errors over non-vegetated regions and positive errors over vegetated areas. Under moderate shadow conditions, relative error patterns are dominated by positive errors over shaded vegetation, while illuminated vegetation areas exhibit near-zero or reduced relative error.

accurate reconstruction in illuminated regions, indicating persistent challenges in compensating for shadow-induced radiometric distortions.

Overall, the combined analysis of global metrics and spatial error patterns confirms that VAE-based architectures constitute a more reliable and robust solution for RGB-to-NIR reconstruction under heterogeneous illumination conditions. Their ability to maintain stable performance across shadow regimes, together with reduced sensitivity to illumination-domain mismatch, directly supports their suitability for operational forestry and precision agriculture monitoring, where acquisition conditions are inherently variable. While GAN-based models can enhance localized textural details under favourable illumination, their higher variability limits their practical reliability. Consequently, these findings demonstrate that effective RGB-to-NIR reconstruction is not determined solely by model expressiveness, but by the interaction between illumination-aware training data, shadow geometry, and the radiometric characteristics of vegetation surfaces. This work therefore advances the current state of RGB-to-NIR reconstruction by explicitly characterizing illumination robustness as a key design criterion for real-world deployment.

4 Conclusions

This work presented a comprehensive evaluation of VAE-CNN and GAN-pix2pix generative models for reconstructing the NIR multispectral band from RGB imagery under varying illumination conditions. The results demonstrate that VAE-based architectures provide a more stable and robust framework for RGB-to-NIR reconstruction in outdoor forestry scenarios, exhibiting higher minimum performance levels, narrower metric distributions, and reduced sensitivity to illumination-induced domain shifts. The best-performing configuration, VAE$_{S2}$, highlights the importance of training with diverse illumination conditions, particularly under moderate shadow regimes.

Beyond technical performance, these findings have direct implications for precision agriculture and forestry monitoring. Reliable NIR reconstruction from RGB imagery can reduce dependence on dedicated multispectral sensors, lowering acquisition costs while enabling high-resolution estimation of vegetation condition and canopy structure. This capability supports scalable monitoring in operational scenarios where multispectral data are sparse, inconsistent, or unavailable.

Future work will extend this methodology to additional multispectral bands, such as Red Edge, and to datasets comprising different or mixed vegetation species, including agricultural crops, to assess generalization under more complex ecological conditions. Moreover, the exploration of alternative generative paradigms, including diffusion models and normalizing flows, may help to identify architectural properties that better capture the radiometric characteristics of multispectral data. Together, these directions aim to establish generalizable principles for applying refined RGB-to-multispectral reconstruction methods across diverse forestry and agricultural datasets.

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

The author has no acknowledgments to declare.

Funding

This work has been supported by ANID under grant Fondecyt Iniciación 11241171.

Author contributions statement

Conceptualization, A.V.-D and O.M.-G; Methodology, A.V.-D and O.M.-G; Software, A.V.-D; Validation, A.V.-D, O.M.-G; Formal analysis, A.V.-D, R.G.-R, T.A.-R.,E.L.-E, O.M.-G; Investigation, A.V.-D and O.M.-G; Data curation, A.V.-D and T.A.-R; Writing—original draft, A.V., R.G.-R, T.A.-R, E.L.-E, R.D, and O.M.-G; Writing—review and editing, A.V.-D, R.G.-R, T.A.-R, E.L.-E, R.D, and O.M.-G; Visualization, A.V.-D; Supervision, O.M.-G; Project administration, O.M.-G; Funding acquisition, O.M.-G

References

1. Liu, Q., Shen, B. & Wen, X. Role of climate-smart agriculture in fighting against climate change in competitive supply chains. *Int. J. Prod. Econ.* **264**, 108978 (2023).
2. Finger, R. *et al.* Towards sustainable crop protection in agriculture: A framework for research and policy. *Agric. Syst.* **219**, 104037, DOI: https://doi.org/10.1016/j.agsy.2024.104037 (2024).
3. Nearing, M. *et al.* Modeling response of soil erosion and runoff to changes in precipitation and cover. *Catena* **61**, 131–154 (2005).
4. Delgado, J. A., Short Jr, N. M., Roberts, D. P. & Vandenberg, B. Big data analysis for sustainable agriculture on a geospatial cloud framework. *Front. Sustain. Food Syst.* **3**, 54 (2019).
5. Campoverde-Molina, M. & Luján-Mora, S. Cybersecurity in smart agriculture: A systematic literature review. *Comput. & Secur.* **150**, 104284, DOI: https://doi.org/10.1016/j.cose.2024.104284 (2025).
6. Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K. & Sun, X. A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Comput. Electron. Agric.* **222**, 109037, DOI: https://doi.org/10.1016/j.compag.2024.109037 (2024).
7. Myneni, R. B., Hall, F. G., Sellers, P. J. & Marshak, A. L. The interpretation of spectral vegetation indexes. *IEEE Transactions on Geosci. remote Sens.* **33**, 481–486 (1995).
8. Wang, J., Wang, Y., Li, G. & Qi, Z. Integration of remote sensing and machine learning for precision agriculture: A comprehensive perspective on applications. *Agronomy* **14**, DOI: 10.3390/agronomy14091975 (2024).
9. Lim, S. L., Sreevalsan-Nair, J. & Daya Sagar, B. Multispectral data mining: A focus on remote sensing satellite images. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **14**, e1522 (2024).
10. Zeng, L., Wardlow, B. D., Xiang, D., Hu, S. & Li, D. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote. Sens. Environ.* **237**, 111511 (2020).
11. Xie, C. & Yang, C. A review on plant high-throughput phenotyping traits using uav-based sensors. *Comput. Electron. Agric.* **178**, 105731 (2020).

12. Zhang, Y. *et al.* Two-step resup&down generative adversarial network to reconstruct multispectral image from aerial rgb image. *Comput. Electron. Agric.* **192**, 106617 (2022).

13. Signoroni, A., Savardi, M., Baronio, A. & Benini, S. Deep learning meets hyperspectral image analysis: A multidisciplinary review. *J. imaging* **5**, 52 (2019).

14. Zhang, J. *et al.* A survey on computational spectral reconstruction methods from rgb to hyperspectral imaging. *Sci. reports* **12**, 11905 (2022).

15. Shi, Z., Chen, C., Xiong, Z., Liu, D. & Wu, F. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1052–10528, DOI: 10.1109/CVPRW.2018.00139 (2018).

16. Zhao, Y., Po, L.-M., Yan, Q., Liu, W. & Lin, T. Hierarchical regression network for spectral reconstruction from rgb images. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1695–1704, DOI: 10.1109/CVPRW50498.2020.00219 (2020).

17. Zamir, S. W. *et al.* Restormer: Efficient transformer for high-resolution image restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5718–5729, DOI: 10.1109/CVPR52688.2022.00564 (2022).

18. Song, D., Sun, H., Ngumbi, E. & Kamruzzaman, M. Multispectral image reconstruction from rgb image for maize growth status monitoring based on window-adaptive spatial-spectral attention transformer. *Comput. Electron. Agric.* **239**, 111062, DOI: https://doi.org/10.1016/j.compag.2025.111062 (2025).

19. Arevalo-Ramirez, T. *et al.* Assessment of light environment conditions for reconstruction of multispectral images by conditional adversarial networks. *Biosyst. Eng.* **260**, 104305, DOI: https://doi.org/10.1016/j.biosystemseng.2025.104305 (2025).

20. Song, X. *et al.* Agricultural image processing: Challenges, advances, and future trends. *Appl. Sci.* **15**, DOI: 10.3390/app15169206 (2025).

21. Mak, H. W. L., Han, R. & Yin, H. H. F. Application of variational autoencoder (vae) model and image processing approaches in game design. *Sensors* **23**, DOI: 10.3390/s23073457 (2023).

22. Liu, J. Research on the application of variational autoencoder in image generation. In *ITM Web of Conferences*, vol. 70, 02001 (EDP Sciences, 2025).

23. Wei, R. & Mahmood, A. Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey. *IEEE Access* **9**, 4939–4956, DOI: 10.1109/ACCESS.2020.3048309 (2021).

24. Gkillas, A., Kosmopoulos, D. & Berberidis, K. Cost-efficient coupled learning methods for recovering near-infrared information from rgb signals: Application in precision agriculture. *Comput. Electron. Agric.* **209**, 107833, DOI: https://doi.org/10.1016/j.compag.2023.107833 (2023).

25. Liu, X., Gherbi, A., Wei, Z., Li, W. & Cheriet, M. Multispectral image reconstruction from color images using enhanced variational autoencoder and generative adversarial network. *IEEE Access* **9**, 1666–1679, DOI: 10.1109/ACCESS.2020.3047074 (2021).

26. Farooque, A. A. *et al.* Red-green-blue to normalized difference vegetation index translation: a robust and inexpensive approach for vegetation monitoring using machine vision and generative adversarial networks. *Precis. Agric.* **24**, 1097–1115 (2023).

27. Fan, K. *et al.* Rmsrgan: A real multispectral imagery super-resolution reconstruction for enhancing ginkgo biloba yield prediction. *Forests* **15**, DOI: 10.3390/f15050859 (2024).

28. Mingrone, M., Seracini, M. & Cevoli, C. Spectral reconstruction applied in precision agriculture: On-field solutions. *Appl. Sci.* **15**, DOI: 10.3390/app152010985 (2025).

29. Carnegie, A. J., Eslick, H., Barber, P., Nagel, M. & Stone, C. Airborne multispectral imagery and deep learning for biosecurity surveillance of invasive forest pests in urban landscapes. *Urban For. & Urban Green.* **81**, 127859, DOI: https://doi.org/10.1016/j.ufug.2023.127859 (2023).

30. Zhao, J. *et al.* Deep-learning-based multispectral image reconstruction from single natural color rgb image—enhancing uav-based phenotyping. *Remote. Sens.* **14**, DOI: 10.3390/rs14051272 (2022).

31. Zhang, Y. *et al.* Two-step resup&down generative adversarial network to reconstruct multispectral image from aerial rgb image. *Comput. Electron. Agric.* **192**, 106617, DOI: https://doi.org/10.1016/j.compag.2021.106617 (2022).

32. Zhang, X. *et al.* A robust multispectral reconstruction network from rgb images trained by diverse satellite data and application in classification and detection tasks. *Remote. Sens.* **17**, DOI: 10.3390/rs17111901 (2025).

33. MicaSense Inc. Micasense image processing documentation (2021). Accessed: 2025-08-01.

34. MicaSense Inc. Micasense imageprocessing (2021). Accessed: 2025-08-01.

35. Honkavaara, E., Markelin, L., Rosnell, T. & Nurminen, K. Influence of solar elevation in radiometric and geometric performance of multispectral photogrammetry. *ISPRS J. Photogramm. Remote. Sens.* **67**, 13–26, DOI: https://doi.org/10.1016/j.isprsjprs.2011.10.001 (2012).

36. Li, H., Xu, L., Shen, H. & Zhang, L. A general variational framework considering cast shadows for the topographic correction of remote sensing imagery. *ISPRS J. Photogramm. Remote. Sens.* **117**, 161–171, DOI: https://doi.org/10.1016/j.isprsjprs.2016.03.021 (2016).

37. Zippenfenig, P. Open-meteo.com weather api, DOI: 10.5281/zenodo.7970649 (2023).

38. Castillo-Girones, S. *et al.* Artificial neural networks in agriculture, the core of artificial intelligence: What, when, and why. *Comput. Electron. Agric.* **230**, 109938, DOI: https://doi.org/10.1016/j.compag.2025.109938 (2025).

39. Ditria, L. & Drummond, T. Long-term prediction of natural video sequences with robust video predictors. *arXiv preprint arXiv:2308.11079* (2023).

40. Yuan, X., Tian, J. & Reinartz, P. Generating artificial near infrared spectral band from rgb image using conditional generative adversarial network. *ISPRS Annals Photogramm. Remote. Sens. Spatial Inf. Sci.* **V-3-2020**, 279–285, DOI: 10.5194/isprs-annals-V-3-2020-279-2020 (2020).

41. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, 234–241, DOI: 10.1007/978-3-319-24574-4_28 (Springer International Publishing, Cham, 2015).

42. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks (2018). 1611.07004.