remote sensing

Article

Lightweight Near-Infrared Spectral Reconstruction from Red UAV Imagery Using Artificial Intelligence for Low-Cost Remote Sensing

Viorel Bostan , Nicu Drumea , Viorel Carbune , Valeriu Seinic , Igor Calmicov , Adriana Ursu and Maria Gutu *

Technical University of Moldova, 9/7 Student Street, MD-2045 Chisinau, Moldova;
viorel.bostan@adm.utm.md (V.B.); nicu.drumea@iis.utm.md (N.D.); viorel.carbune@calc.utm.md (V.C.);
valeriu.seinic@icg.utm.md (V.S.); igor.calmicov@calc.utm.md (I.C.); ursu.adriana@iis.utm.md (A.U.)
* Correspondence: maria.gutu@iis.utm.md

Highlights

What are the main findings?

- A compact pixel-wise multilayer perceptron with 609 trainable parameters estimates normalized NIR-band digital intensity using only the corresponding UAV-acquired red-band value, without additional spectral bands or spatial-context information.
- The proposed MLP achieved error-based performance comparable to the Random Forest baseline while obtaining the highest Pearson correlation coefficient among the evaluated models.

What are the implications of the main findings?

- The substantially smaller saved-model size relative to Random Forest demonstrates a favorable trade-off between predictive performance and architectural compactness.
- The proposed methodology provides an empirical reference for investigating single-band NIR estimation in exploratory and resource-constrained remote sensing workflows, with broader applicability subject to sensor-specific validation.

Abstract

Near-infrared imagery is essential for vegetation monitoring, precision agriculture, and environmental remote sensing, but multispectral UAV systems remain significantly more expensive and less accessible than conventional RGB imaging platforms. This study presents a lightweight artificial intelligence framework for reconstructing the NIR spectral band exclusively from the red spectral band acquired by a UAV. The proposed methodology formulates the reconstruction task as a pixel-wise nonlinear regression problem and employs a compact multilayer perceptron (MLP) containing only 609 trainable parameters, without exploiting spatial neighborhood information. The framework was developed and evaluated using 280 synchronized multispectral UAV image sets acquired with a DJI Phantom 4 Multispectral platform over a heterogeneous agricultural landscape in the Republic of Moldova. Of these, 252 image sets were used for model development, and 28 were reserved as a held-out within-mission test subset. Quantitative evaluation on a held-out test dataset from the same acquisition mission yielded a mean squared error of 0.010329, a root mean squared error of 0.101632, a mean absolute error of 0.079883, a coefficient of determination of 0.253383, and a Pearson correlation coefficient of 0.683637 between measured and reconstructed normalized NIR digital intensities. The results indicate that the model captures part of the red–NIR relationship under the evaluated acquisition conditions; however, the moderate coefficient of determination suggests that the reconstructed

Check for updates

Academic Editors: Hao Zhang, Qiwen Jin, Chengli Peng and Xingyu Jiang

Received: 4 July 2026
Revised: 22 August 2026
Accepted: 24 August 2026
Published: 4 September 2026

 https://doi.org/10.3390/rs18173015

values are an approximation rather than a replacement for measured NIR observations. An illustrative NDVI-based assessment showed that broad spatial vegetation patterns remained identifiable. Rather than introducing a new neural network architecture, this work establishes a compact empirical baseline to investigate the practical performance and limitations of pixel-wise NIR reconstruction from a single red-band value with minimal model complexity.

Keywords: lightweight deep learning; near-infrared reconstruction; spectral reconstruction; UAV remote sensing; multispectral imaging; artificial intelligence

1. Introduction

Unmanned aerial vehicles (UAVs) have become an indispensable component of modern geomatics, enabling the acquisition of high-spatial-resolution imagery through flexible flight planning, rapid deployment, and relatively low operational costs. Compared with satellite remote sensing, UAV platforms offer centimeter-level spatial resolution and enable image acquisition under user-defined environmental conditions, making them particularly suitable for precision agriculture, environmental monitoring, forestry, infrastructure inspection, and natural resource management [1]. Continuous advances in UAV technology, multispectral sensors, and artificial intelligence (AI)-based image analysis have considerably expanded the role of aerial imagery in producing geospatial information that supports data-driven decision-making across a wide range of environmental and engineering applications [2–6].

Among the spectral bands acquired by multispectral sensors, the near-infrared band plays a central role in remote sensing because of its strong sensitivity to vegetation physiology and surface characteristics. Healthy vegetation strongly absorbs radiation in the red spectral region due to chlorophyll pigments, while the internal cellular structure of leaves produces high digital intensities in the near-infrared region. This distinctive spectral behavior underlies numerous vegetation indices, particularly the Normalized Difference Vegetation Index (NDVI), which remains one of the most widely used indicators for assessing vegetation vigor, biomass, crop stress, canopy development, and photosynthetic activity [7–12]. Consequently, accurate NIR measurements are essential not only for precision agriculture but also for environmental monitoring, ecosystem assessment, land-cover analysis, and rapid evaluation of vegetation conditions following natural or anthropogenic disturbances.

Despite these advantages, multispectral imaging systems remain substantially more expensive than conventional imaging solutions. Dedicated multispectral cameras require radiometric calibration, increase payload complexity, and are generally available only on specialized UAV platforms. Their effective use may also depend on specialized processing software, radiometric workflows, computational infrastructure, and personnel with expertise in multispectral image analysis. These requirements can create a substantial entry barrier for small farms and individual producers, even when the potential benefits of precision-agriculture monitoring are recognized. Conversely, conventional RGB cameras are integrated into most commercial drones and have generated extensive aerial image archives over the past decade. Although these datasets contain valuable visual information, the absence of the near-infrared spectral band prevents their direct use in calculating NIR-dependent vegetation indices [13]. This limitation is particularly relevant in operational scenarios where rapid deployment and low equipment costs are critical, such as large-area environmental surveys, emergency response, post-disaster reconnaissance, and routine UAV monitoring missions. Under such circumstances, computational methods capable of approximating missing spectral information from limited inputs could extend the analytical capabilities of lower-cost UAV platforms, reduce dependence on specialized multispectral-processing resources, and broaden access to selected precision-agriculture practices [13–17].

Recent advances in deep learning have significantly improved spectral reconstruction and vegetation-index estimation. Convolutional neural networks (CNNs), conditional generative adversarial networks (GANs) [18], encoder–decoder architectures, and transformer-based models have demonstrated promising performance in reconstructing NIR imagery or estimating vegetation indices from visible-spectrum observations [14–16,19–25]. Zhao et al. [20] proposed a deep learning framework for multispectral image reconstruction to support UAV-based crop phenotyping. In contrast, Yuan et al. [16] introduced a conditional generative adversarial network for generating artificial NIR spectra. Gkillas et al. [14] investigated computationally efficient learning strategies for recovering NIR information in precision agriculture, while Moscovini et al. [15] developed an open-source machine learning framework for pixel-wise NDVI prediction from calibrated RGB imagery. Other studies have explored GAN-based RGB-to-NDVI translation [21], CNN-based vegetation segmentation using virtual NIR channels [22], lightweight multilayer perceptron architectures for spectral super-resolution [24], and deep learning approaches for NDVI estimation throughout crop growth cycles [17,19,25]. Collectively, these studies, together with recent advances in hyperspectral vegetation-index retrieval and artificial intelligence for agricultural monitoring [11,12], demonstrate the growing potential of AI-based methods for extracting and reconstructing meaningful spectral information in remote sensing [14–16,19–25].

Nevertheless, most existing methods rely on computationally intensive architectures that jointly exploit spectral and spatial information through convolutional operations, attention mechanisms, or image-to-image translation networks. Despite their generally high reconstruction accuracy, such approaches often depend on large paired datasets and substantial computational resources. These requirements may limit deployment on embedded UAV platforms and edge computing systems, particularly in operational remote sensing scenarios that require computationally efficient inference [14,20,23,24].

Beyond computational complexity, another important research gap concerns the interpretation of the reconstructed spectral information itself. Most published studies primarily aim to maximize reconstruction accuracy using increasingly sophisticated deep-learning architectures. The present study addresses a complementary question: how much near-infrared information can be inferred solely from the spectral responses of individual pixels, and what is the trade-off between the available spectral information, prediction accuracy, and model complexity? Because convolutional and transformer-based models simultaneously exploit both spectral and spatial context, it is difficult to distinguish the contribution of the intrinsic spectral relationship from that of neighboring image structures. From a geomatics perspective, this distinction is particularly important because numerous remote sensing workflows, including radiometric correction, pixel-wise classification, vegetation-index computation, thematic mapping, and environmental assessment, process pixels independently.

It is acknowledged that NIR reflectance cannot generally be uniquely determined from red reflectance alone. Different combinations of vegetation structure, chlorophyll absorption, canopy density, soil background, illumination, and shadow conditions may produce similar red-band responses while corresponding to different NIR values. For this reason, the present study does not assume a deterministic physical equivalence between the two bands. Instead, the red-only configuration is intentionally used as a constrained experimental setting to quantify how much predictive information about NIR can be statistically recovered from a single visible-band observation under controlled acquisition conditions.

To address this research gap, the present study investigates whether the NIR spectral response can be reconstructed solely from red-band digital intensity measurements at the pixel level. Unlike image-to-image translation approaches, the proposed methodology employs a lightweight multilayer perceptron (MLP) to model the nonlinear relationship between red-band digital intensities and their corresponding NIR response without incorporating spatial neighborhood information. The reconstructed NIR band is subsequently used to generate NDVI maps, enabling the practical usefulness of the predicted spectral information to be evaluated alongside conventional regression metrics.

The proposed methodology was validated using multispectral UAV imagery acquired over agricultural areas in the Republic of Moldova with a DJI Phantom 4 Multispectral platform. The dataset comprises heterogeneous agricultural scenes containing cultivated fields, harvested parcels, unpaved agricultural roads, vegetation strips, isolated trees, and transition zones, thereby providing representative spectral variability for evaluating within-mission predictive performance on held-out images. Although the experimental validation focuses on agricultural environments, the underlying pixel-wise formulation may also be investigated in other remote sensing contexts in future studies. The present work, however, primarily addresses accessible vegetation assessment and preliminary crop-monitoring applications.

The contribution of this study lies in the systematic empirical investigation of the extent to which normalized NIR digital intensity can be estimated from a single red-band value in a compact pixel-wise configuration and of the resulting trade-off between predictive accuracy, limited input information, and model complexity, without relying on additional spectral bands or spatial-context processing. A compact MLP is evaluated against a mean predictor, linear regression, second-degree polynomial regression, and Random Forest using the same held-out test subset and common evaluation metrics. Predictive performance is examined alongside architectural complexity, training and inference times, and saved model size. This combined evaluation characterizes the trade-off between predictive capability and computational requirements and determines the extent to which useful NIR-related information can be obtained from a minimal spectral input. By reducing input and model-complexity requirements, the framework provides a foundation for more accessible vegetation-monitoring workflows, with particular relevance to small farms and individual producers for whom specialized multispectral-processing infrastructure and expertise may be economically impractical. In addition, vegetation-index maps derived from measured and reconstructed NIR data are compared to assess the preservation of broad spatial vegetation patterns.

2. Materials and Methods

2.1. Methodological Framework

The proposed methodology was designed to investigate the feasibility of reconstructing the near-infrared spectral response from the red spectral band using a supervised machine learning approach. In contrast to conventional multispectral image processing, where dedicated sensors directly acquire all spectral channels, the present framework estimates normalized NIR digital intensities as a nonlinear function of the corresponding red digital intensities at each image pixel.

The reconstruction problem was formulated as a pixel-wise regression task, where each pixel from the multispectral imagery constitutes an independent training sample. Consequently, the learning process focuses exclusively on modeling the functional relationship between normalized red-band and NIR-band digital intensity values, without considering image texture, neighborhood information, or spatial feature extraction. Similar pixel-wise learning strategies have previously been investigated for vegetation index estimation and spectral reconstruction [14,15,24]. In contrast, most recent studies employ image-based deep learning architectures that combine both spectral and spatial information [16,19–25].

The complete methodological workflow is illustrated in Figure 1. The proposed framework consists of six consecutive stages: (1) acquisition of multispectral UAV imagery, (2) extraction of the red (R) and near-infrared spectral bands, (3) numerical normalization and construction of a pixel-wise dataset, (4) supervised training of a lightweight multilayer perceptron neural network, (5) reconstruction of the NIR band for previously unseen imagery, and (6) quantitative evaluation using regression metrics and an illustrative NDVI-based assessment.

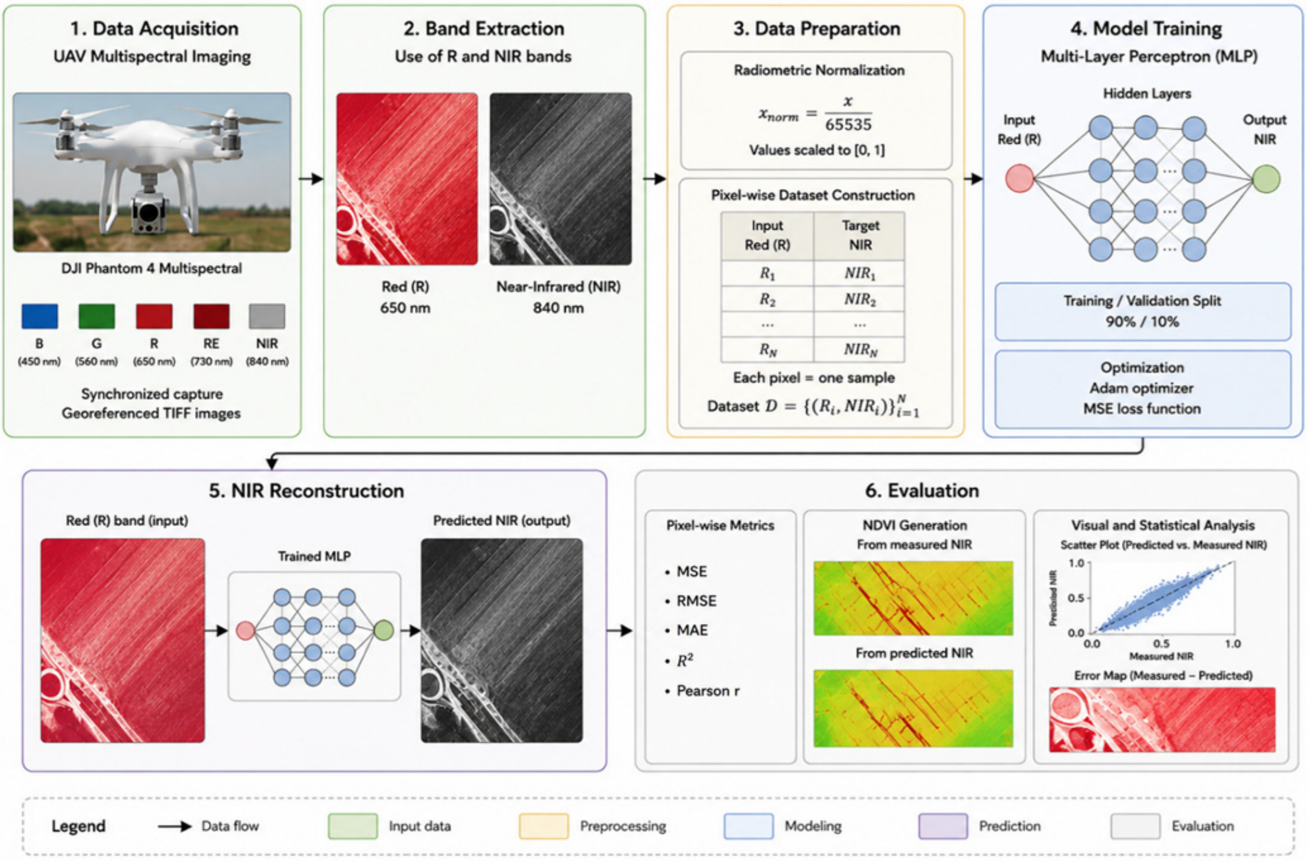


Figure 1. Methodological workflow of the proposed red-to-NIR reconstruction framework.

The first stage involves acquiring synchronized multispectral imagery that simultaneously records the red and NIR bands. The measured NIR-band digital intensities serve as the reference target for supervised model training and quantitative evaluation.

Following image acquisition, the red and NIR bands are extracted and numerically normalized, and each image is then converted to a pixel-wise dataset. Each observation, therefore, consists of one normalized red-band digital-intensity value and the corresponding measured normalized NIR-band digital intensity value. This representation transforms the reconstruction problem in image processing into a nonlinear regression problem, enabling the application of standard supervised learning techniques.

The normalized dataset is subsequently used to train a lightweight multilayer perceptron that approximates the nonlinear mapping from input red digital intensities to target normalized NIR digital intensities. After training, the network is applied to previously unseen UAV imagery to predict the NIR response for every image pixel independently.

The reconstructed NIR band is finally evaluated using two complementary approaches. The first consists of statistical regression metrics that quantify the agreement between the measured and reconstructed digital intensity values. The second evaluates the practical usefulness of the reconstructed spectral information by generating NDVI and visually comparing it with NDVI calculated from the original multispectral imagery [26]. Similar validation strategies that combine numerical accuracy with application-oriented assessment have become increasingly common in remote sensing and vegetation-monitoring studies [14–16,19–25].

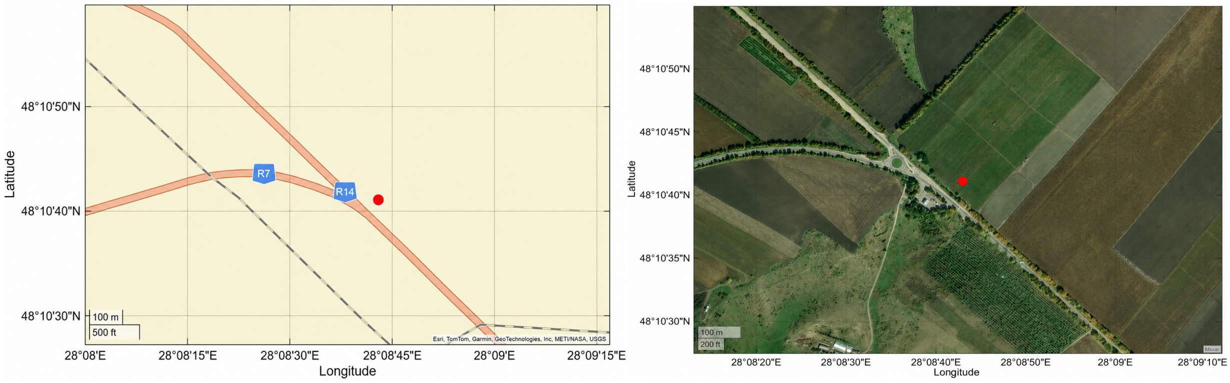
Therefore, the proposed methodological framework combines conventional UAV image preprocessing with a lightweight neural regression model specifically developed for pixel-wise red-to-NIR reconstruction. Unlike most existing image-to-image reconstruction approaches that rely on computationally demanding convolutional architectures [16,19–25], the proposed method was intentionally designed to minimize computational complexity. Since each prediction depends only on a single input value, inference requires only elementary scalar operations, making the model suitable for execution on conventional general-purpose processors. Because each pixel-wise prediction depends on a single input value and the network contains only 609 trainable parameters, the proposed architecture has limited storage and arithmetic requirements. Its execution characteristics were evaluated on a conventional CPU-based system without GPU acceleration. Although implementation on embedded or resource-constrained platforms was not examined in the present study, the compact architecture provides a suitable basis for future platform-specific optimization and deployment studies.

2.2. Study Area and UAV Data Acquisition

The experimental dataset was collected on agricultural land located in a low-relief plain in southern Moldova. The study area comprises a heterogeneous agricultural landscape including cultivated fields, recently harvested arable land, row-crop parcels, unpaved agricultural roads, small patches of spontaneous vegetation, agricultural infrastructure, vegetation strips, isolated trees, and transition zones between vegetated and non-vegetated surfaces. At the time of acquisition, the harvested fields were in a post-harvest stage, whereas the surrounding grassland and spontaneous vegetation remained photosynthetically active. This diversity of land-cover types provides a broad range of spectral responses. This surface heterogeneity provides a useful within-mission setting for examining the spatial distribution of red-to-NIR reconstruction errors.

Figure 2a shows the geographical location of the study area within the Republic of Moldova. In contrast, Figure 2b presents a satellite image highlighting the exact location of the UAV multispectral image acquisition site. The red dot indicates the georeferenced position associated with a representative image from the experimental dataset. The corresponding spatial coordinates enabled the identification of the acquisition site on both the cartographic reference map and the satellite imagery. The corresponding map views were subsequently extracted and prepared for presentation using MATLAB (version R2025a). These images provide the geographical and spatial context of the experimental area used in this study.

Image acquisition was performed using a DJI Phantom 4 Multispectral unmanned aerial vehicle (DJI, Shenzhen, China). The UAV survey was conducted on 7 September 2020, starting at approximately 11:30 local time. The flight lasted 27 min 32 s and covered approximately 20 ha at a relative flight altitude of approximately 99.7 m. The camera was maintained in a near-nadir orientation, with a representative gimbal pitch of $-89.9°$. A total of 280 synchronized multispectral image sets were acquired during the mission.



(**a**) (**b**)

Figure 2. Study area and UAV data-acquisition site: (**a**) cartographic reference map showing the geographical location of the study area in the Republic of Moldova; (**b**) satellite image showing the UAV multispectral data-acquisition area. The red dot marks the acquisition location of a representative georeferenced image from the experimental dataset.

According to observations recorded at the Ceadîr-Lunga meteorological station (https://rp5.ru/Arhiva_meteo_%C3%AEn_Cead%C3%AEr-Lunga, accessed on 21 July 2026), the air temperature was 28.6 °C, the relative humidity was 41%, and the mean wind speed was 3 m/s from the southeast. No precipitation or cloud cover was reported, and horizontal visibility was approximately 20 km. These conditions indicate that the UAV survey was conducted under clear weather and relatively stable illumination conditions.

The UAV was equipped with an integrated six-camera imaging system comprising one RGB sensor and five monochrome sensors. The five monochrome sensors simultaneously acquire images in blue (450 ± 16 nm), green (560 ± 16 nm), red (650 ± 16 nm), red-edge (730 ± 16 nm), and near-infrared (840 ± 26 nm), together with a conventional RGB image acquired by the separate RGB sensor. All six sensors are mounted on the same stabilized gimbal, and image acquisition is synchronized. The red and NIR images used in this study were acquired simultaneously by two separate monochrome sensors within the same multispectral camera array. Each multispectral image was stored as a 16-bit TIFF. This synchronized acquisition facilitates the alignment of the spectral bands while minimizing temporal discrepancies caused by platform motion or illumination changes, which is essential for supervised pixel-wise learning tasks [1]. The multispectral platform was used to obtain synchronized red and NIR measurements, with the measured NIR-band digital intensities serving as reference values for supervised training and evaluation. Because no additional empirical radiometric calibration was performed, these reference data represent sensor digital intensities rather than calibrated surface reflectance.

Of the five spectral bands recorded by the multispectral camera, this study used only the red and near-infrared bands. The red band was selected as the sole input variable because of its strong physical relationship with vegetation spectral response and its direct role in the computation of the Normalized Difference Vegetation Index, one of the most widely used vegetation indicators in remote sensing [7–10]. The NIR band served exclusively as the reference target during supervised learning. The remaining spectral bands (blue, green, and red-edge) were intentionally excluded to investigate the capability of reconstructing NIR information using only the spectral response of the red band. The resulting experimental dataset, whose main characteristics are summarized in Table 1, was subsequently used for supervised training and independent evaluation of the proposed neural network.

Table 1. Dataset characteristics and image-level partitioning.

Parameter	Value
UAV	DJI Phantom 4 Multispectral
Number of synchronized multispectral image sets	280
Spectral bands used	Red (input), NIR (target)
Image format	16-bit GeoTIFF
Model-development subset	252 (90%)
Within-mission held-out test subset	28 (10%)

As shown in Table 1, the experimental dataset comprises 280 multispectral images acquired with the DJI Phantom 4 Multispectral platform. The red and near-infrared spectral bands were used to generate paired input–target samples for the proposed Red-to-NIR reconstruction task. The dataset was subsequently partitioned at the image-set level into a model-development subset and a held-out test subset at a 9:1 ratio, with the test subset reserved exclusively for final model evaluation. This experimental design evaluates model performance on previously unseen image files from the same acquisition mission. Because the dataset was acquired with substantial image overlap and no spatial blocking was applied, the evaluation should be interpreted as within-mission held-out performance rather than fully spatially independent generalization. This partitioning follows standard machine learning validation practices for remote sensing [1]. The remaining spectral bands (blue, green, and red-edge) were intentionally excluded to investigate the capability of reconstructing NIR information using only the spectral response of the red band.

2.3. Dataset Preparation and Pre-Processing

The acquired multispectral images were stored as 16-bit TIFF files containing unsigned digital numbers (DNs). No additional empirical radiometric calibration using a reflectance calibration panel or field spectrometer was performed. The images were processed using the manufacturer's default sensor settings and the Pix4D workflow. Before constructing the supervised learning dataset, the pixel values of both the red and near-infrared bands were normalized to the interval [0, 1] according to the following:

$$\mathrm{x_{norm}} = \frac{\mathrm{x}}{65,535}, \tag{1}$$

where x denotes the original 16-bit pixel intensity, and $\mathrm{x_{norm}}$ represents the normalized digital intensity. The denominator 65,535 represents the maximum value theoretically encodable by an unsigned 16-bit integer ($2^{16} - 1$) and was used as a fixed numerical scaling factor. Its use does not imply that the sensor measurements occupied the entire 0–65,535 range. Because no empirical radiometric calibration was performed, the normalized values represent relative digital intensities rather than calibrated surface reflectance.

This numerical scaling improves numerical stability during neural network optimization, accelerates convergence, and ensures that all input values are represented on a common numerical scale, thereby facilitating efficient gradient-based learning [27,28].

Unlike conventional image-based deep learning approaches that process image patches or entire images, the proposed methodology adopts a pixel-wise learning strategy, treating each pixel as a within-mission held-out training subset. For every pixel, the normalized digital intensities from the red spectral band were used as the input feature, while the corresponding normalized near-infrared digital intensities served as the regression target:

$$\mathrm{x_i} = [\mathrm{R_i}], \mathrm{y_i} = \mathrm{NIR_i}, \tag{2}$$

where R_i means the normalized digital intensities of the red spectral band, and NIR_i represents the corresponding normalized near-infrared digital intensities measured at the same pixel location.

Accordingly, the supervised learning dataset can be represented as

$$D = \{(R_i, NIR_i)\}_{i=1}^{N}, \tag{3}$$

where N represents the total number of pixel samples extracted from all training images.

This pixel-wise formulation differs from convolutional neural networks because neighboring pixels do not contribute to feature extraction. Each prediction relies exclusively on the normalized red digital intensity of one pixel, allowing the red–NIR relationship to be investigated independently of spatial context. The formulation was deliberately selected as a compact single input. It provides a controlled baseline for determining how effectively NIR-related information can be approximated from a single red-band value while minimizing spectral input and architectural requirements.

A random hold-out strategy partitioned the 280 synchronized multispectral image sets into a model-development subset of 252 (90%) and a held-out test subset of 28 (10%). This separation prevents complete held-out image files from being used during model fitting and enables evaluation on previously unseen image files from the same acquisition mission. It is a standard evaluation protocol in machine learning applications for remote sensing [29]. The implementation relied on the default system-initialized pseudo-random state rather than specifying a fixed random seed. Pixel samples extracted from the 252 model-development image sets were subsequently divided into training and validation subsets. In contrast, all pixel samples from the 28 held-out image sets were excluded from the model and reserved for final testing.

The UAV survey was conducted with a longitudinal image overlap of approximately 75–85% and a lateral overlap of approximately 65–75%. Consequently, two consecutive images could share a substantial portion of the same ground area. Because the random partitioning was performed without spatial blocking, some held-out test images may spatially overlap with images assigned to the model-development subset, even though they were excluded as complete image files from model training and validation. Therefore, the reported results characterize held-out image performance within the same UAV mission and under the same acquisition conditions. Although the model may also be applicable to spatially independent areas, other sites, or different acquisition missions, its generalization under such conditions requires further validation using a within-mission held-out subset.

2.4. Neural Network Architecture

The proposed reconstruction model was implemented as a fully connected feed-forward artificial neural network based on a multilayer perceptron (MLP) architecture. Unlike convolutional neural networks (CNNs), which simultaneously exploit both spectral and spatial neighborhood information, the proposed model performs pixel-wise nonlinear regression, treating each pixel independently. Consequently, the network learns a nonlinear mapping between normalized red digital intensities and corresponding near-infrared digital intensities without incorporating spatial context.

The overall architecture of the proposed neural network is illustrated in Figure 3. The model receives a single normalized red-band digital-intensity value as input and predicts the corresponding normalized NIR-band digital intensity value. As shown in Figure 3, the network consists of one input neuron, two fully connected hidden layers containing 32 and 16 neurons, respectively, and a single output neuron that estimates the normalized NIR-band digital intensity value. Thus, the network follows a 1–32–16–1 architecture, where the first value denotes the single input feature, the intermediate values indicate the number of neurons in the two hidden layers, and the final value represents the predicted NIR output.

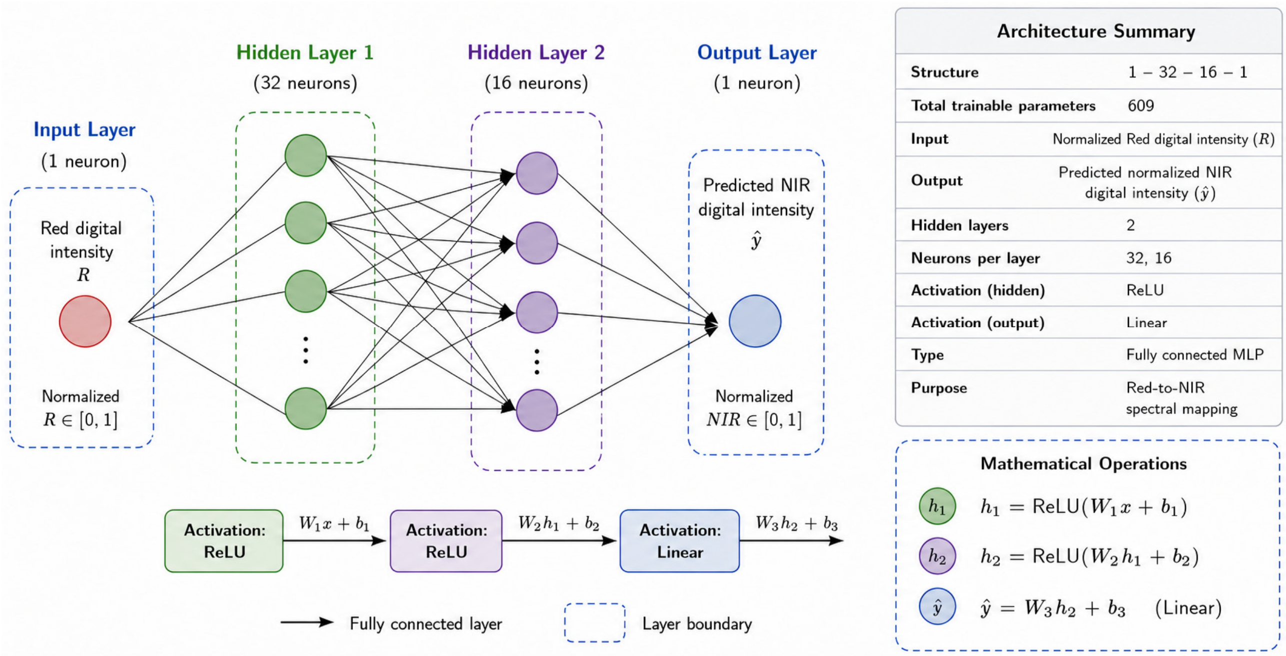


Figure 3. Architecture of the lightweight MLP used for pixel-wise reconstruction of normalized NIR-band digital intensity from normalized red-band digital intensity.

The hidden layers containing 32 and 16 neurons were selected to balance nonlinear modeling capacity with architectural compactness. The architecture was not selected through an exhaustive hyperparameter search, as the purpose of the study was not to optimize a neural architecture for maximum predictive accuracy. Instead, the 32–16 configuration was selected as a compact nonlinear baseline that provides substantially greater modeling capacity than linear and polynomial regression while maintaining only 609 trainable parameters. The progressively decreasing layer width enables nonlinear feature representation followed by compression before the scalar NIR output. The study focused on evaluating a deliberately lightweight configuration intended for prospective resource-constrained applications.

Each hidden layer employs the Rectified Linear Unit (ReLU) activation function due to its computational efficiency, its ability to alleviate the vanishing gradient problem, and its widespread use in deep learning for remote sensing and computer vision [28]. The output layer uses a linear activation function, which is appropriate for continuous regression problems involving normalized digital intensity values.

The forward propagation through the network can be defined as follows:

$$\mathrm{h_1} = \mathrm{ReLU}(\mathrm{W_1 x_i} + \mathrm{b_1}), \mathrm{h_2} = \mathrm{ReLU}(\mathrm{W_2 h_1} + \mathrm{b_2}), \quad (4)$$

while the predicted NIR value is obtained as

$$\hat{\mathrm{y}}_\mathrm{i} = \mathrm{W_3 h_2} + \mathrm{b_3}, \quad (5)$$

where $\mathrm{W_1}$, $\mathrm{W_2}$, and $\mathrm{W_3}$ denote the weight matrices associated with the first hidden layer, second hidden layer, and output layer, respectively, while $\mathrm{b_1}$, $\mathrm{b_2}$, and $\mathrm{b_3}$ denote their corresponding bias vectors or bias terms.

As illustrated in Figure 3, the proposed architecture contains 609 trainable parameters, providing sufficient nonlinear modeling capacity while maintaining a lightweight computational structure suitable for rapid pixel-wise inference over large UAV datasets. This balance between predictive capability and architectural compactness supports further investigation of the model in resource-constrained implementations.

The MLP was employed as a conventional nonlinear regression model to represent relationships that linear regression cannot adequately capture. Although substantially simpler than convolutional or transformer-based architectures, the proposed network models the nonlinear red–NIR relationship using only one input value and 609 trainable parameters. Its relevance is therefore assessed empirically by comparing it with constant, linear, polynomial, and tree-based regression models, along with analyses of predictive performance, storage requirements, and execution characteristics. This compact formulation makes the model a relevant candidate for lightweight spectral reconstruction workflows and for future implementation on resource-constrained platforms.

2.5. Model Training

The proposed neural network was trained using a supervised learning strategy, where each normalized red digital intensity value was an input sample, and the corresponding normalized near-infrared digital intensities served as the target. During training, the model learned its parameters by minimizing the discrepancy between the predicted and measured normalized NIR digital intensity values across the training dataset.

The optimization problem can be formulated as

$$\Theta^{*} = \underset{\Theta}{\operatorname{argmin}}\, \mathrm{L}(\Theta), \tag{6}$$

where Θ denotes the complete set of trainable weights and biases of the neural network, Θ^{*} represents the optimal parameter set obtained after training, and $\mathrm{L}(\Theta)$ represents the objective function to be minimized during optimization.

The loss function was defined as the Mean Squared Error (MSE):

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2, \tag{7}$$

where $\mathrm{y_i}$ and $\hat{\mathrm{y}}_\mathrm{i}$ denote the measured and predicted normalized NIR digital intensity values, respectively, and N is the total number of training samples.

The neural network was implemented in Python 3.11.9 using TensorFlow 2.21.0 and Keras and was trained on a system equipped with an 11th Gen Intel Core i5-1135G7 processor and 8 GB of RAM. The Adam optimizer was configured with a learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-7}$. Adam was selected because it combines adaptive learning-rate estimation with momentum-based optimization, providing stable convergence and efficient optimization for nonlinear regression problems involving large datasets [27]. The MSE loss function was chosen because it is the standard objective function for continuous regression tasks and strongly penalizes large prediction errors.

The extracted pixel samples were randomly split into training and validation sets at a 9:1 ratio. Such a partition optimizes the learning process while simultaneously monitoring the model's performance on unseen validation data, thereby reducing the risk of overfitting and supporting optimization monitoring and model selection within the development subset [29].

Training was performed using mini-batch gradient descent with a batch size of 4096 samples. Since the dataset consists of millions of independent pixel observations extracted from multispectral imagery, a relatively large batch size improves computational efficiency while maintaining stable parameter updates during optimization.

The network weights were initialized using the Glorot-uniform initializer, while all bias terms were initialized to zero. Although the maximum number of training epochs was set to 100, the validation loss reached its minimum at epoch 20. Therefore, the model weights corresponding to epoch 20 were retained for final evaluation. The principal hyperparameters and implementation settings used to train the proposed neural network are listed in Table 2. These parameters were selected to balance computational efficiency and stable optimization while maintaining the lightweight architecture proposed in this study.

Table 2. Training configuration of the proposed neural network.

Parameter	Value
Model	Multilayer Perceptron
Input feature	Normalized red-band digital intensity
Output	Normalized NIR-band digital intensity
Hidden layers	32–16 neurons
Activation function	ReLU
Output activation	Linear
Loss function	Mean Squared Error
Optimizer	Adam
Batch size	4096
Training framework	TensorFlow 2.21.0/Keras 3.15.1
Training/Validation split	90%/10%
Python version	3.11.9
Learning rate	0.001
Adam β_1	0.9
Adam β_2	0.999
Adam ε	10^{-7}
Weight initialization	Glorot uniform
Bias initialization	Zeros
Maximum epochs	100
Completed epochs	20
Hardware	11th Gen Intel Core i5-1135G7 CPU, 8 GB RAM

As shown in Table 2, the proposed MLP employs two hidden layers with ReLU activation and is optimized using the Adam algorithm with the Mean Squared Error (MSE) loss function. The selected configuration provides a compact architecture while ensuring efficient convergence during supervised training.

During training, the validation subset was used to monitor the evolution of the loss function and evaluate the generalization capability of the proposed model. The progressive reduction in both training and validation errors indicated that the network successfully learned the nonlinear mapping between the red and NIR spectral responses without exhibiting unstable optimization behavior.

After the training stage, the optimized model was applied to previously unseen UAV imagery to reconstruct the NIR band from the normalized red digital intensity values. The pixel-wise formulation and the compact 609-parameter architecture result in a small saved-model size. Its practical execution characteristics were evaluated separately through the training and inference measurements reported in Section 3.3.

2.6. Baseline Models and Comparative Evaluation

To contextualize the performance of the proposed MLP, four baseline approaches were evaluated: a constant mean predictor, linear regression, second-degree polynomial regression, and Random Forest regression. These models were selected to represent progressively more complex red-to-NIR mappings, ranging from a constant prediction to linear, simple nonlinear parametric, and nonlinear ensemble-based regression.

https://doi.org/10.3390/rs18173015

The mean predictor assigned every test pixel the mean normalized NIR digital intensities calculated exclusively from the training dataset and did not use the red-band input. Linear regression modelled the relationship between the red and NIR digital intensities using a first-degree function:

$$\widehat{\mathrm{NIR}} = a\mathrm{R} + b, \tag{8}$$

where R denotes the normalized red-band digital intensities, while a and b are the regression coefficient and intercept, respectively.

Second-degree polynomial regression was included to determine whether a simple nonlinear function could represent the red-to-NIR relationship more adequately:

$$\widehat{\mathrm{NIR}} = a\mathrm{R}^2 + b\mathrm{R} + c, \tag{9}$$

where a, b, and c are coefficients estimated from the training data.

Random Forest regression was included as a nonlinear machine learning baseline. The model used an ensemble of 80 regression trees and a fixed random seed of 42 to ensure reproducibility.

All baseline models used the same normalized red–NIR training pairs and were evaluated on the same held-out test images employed for the final evaluation of the proposed MLP. No pixel from the held-out test images was used for model fitting or hyperparameter selection. The models were compared using MSE, RMSE, MAE, R^2, and Pearson's correlation coefficient. For the constant mean predictor, Pearson's correlation coefficient was undefined because all predicted values were identical.

In addition to predictive accuracy, the saved model size was recorded to compare the storage requirements of the proposed MLP with those of the baseline models. Model size was treated separately from training and inference speed because a smaller saved model does not necessarily imply faster execution.

2.7. Performance Evaluation

The proposed methodology was examined from two complementary perspectives: quantitative NIR reconstruction performance and an illustrative assessment of the influence of reconstruction errors on NDVI. The numerical evaluation quantified agreement between reconstructed and measured normalized NIR digital intensity values, whereas the application-oriented evaluation assessed whether the reconstructed NIR band preserved sufficient spectral information for vegetation index generation [26]. Similar evaluation strategies combining statistical accuracy with practical remote sensing applications have recently been adopted in studies on spectral reconstruction and NDVI estimation [14,15,20–22,25].

To quantitatively assess the regression performance of the proposed MLP and the baseline models, five complementary statistical indicators were employed: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), the coefficient of determination (R^2), and Pearson's correlation coefficient (r). The MSE was computed using Equation (7), while the RMSE, MAE, and R^2 were calculated using Equations (10)–(12) as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2}, \tag{10}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum\nolimits_{\mathrm{i}=1}^{\mathrm{N}}|\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}|, \tag{11}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{y_i} - \overline{\mathrm{y}})^2}. \tag{12}$$

MAE represents the average absolute magnitude of the prediction errors, whereas R^2 indicates the proportion of variability in the measured NIR values explained by the model relative to a constant reference based on the mean of the test observations. The mean predictor yielded an R^2 of zero, indicating that the constant prediction did not explain any variance in the NIR values of the held-out test subset.

The Pearson correlation coefficient was additionally used to quantify the linear relationship between measured and reconstructed NIR values:

$$r = \frac{\sum_{i=1}^{N} (y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum_{i=1}^{N} (y_i - \overline{y})^2} \sqrt{\sum_{i=1}^{N} (\hat{y}_i - \overline{\hat{y}})^2}}, \tag{13}$$

where $\overline{\hat{y}}$ denotes the mean predicted normalized NIR digital intensities.

These complementary metrics evaluate different aspects of model performance. MSE and RMSE quantify the average reconstruction error, whereas the Pearson correlation coefficient evaluates the strength of the linear relationship between the measured and predicted digital intensity values.

Because the ultimate objective of NIR reconstruction is not only to minimize prediction errors but also to preserve the spectral information required for subsequent remote sensing analyses, an additional application-oriented validation was performed using the Normalized Difference Vegetation Index:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}}, \tag{14}$$

where NIR denotes the reconstructed or measured near-infrared digital intensities, and R represents the digital intensities measured in the red spectral band.

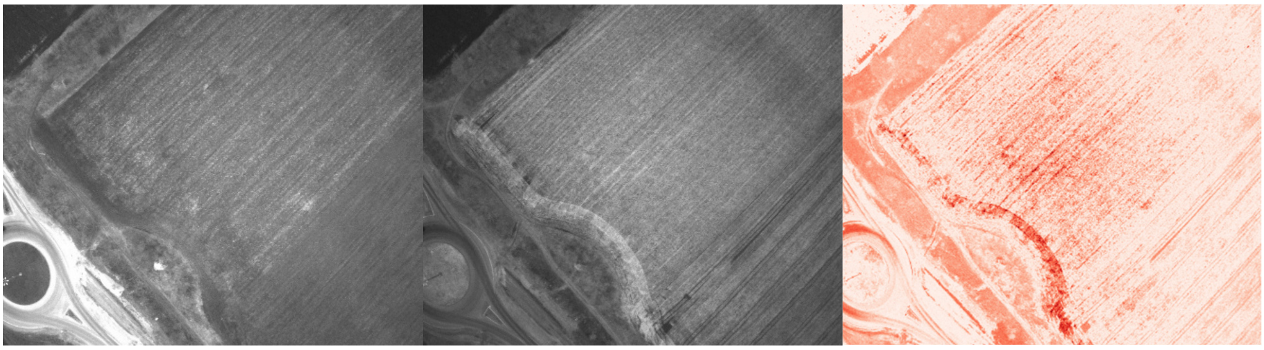
For each image included in the held-out test subset, two NDVI maps were generated: one using the original measured NIR band and another using the reconstructed NIR band predicted by the proposed neural network. The comparison of these products enabled the practical impact of spectral reconstruction on vegetation-index generation to be assessed, complementing the numerical evaluation based on regression metrics.

In addition to the statistical analysis, a qualitative visual assessment was performed by comparing the original NIR imagery, the reconstructed NIR imagery, and the corresponding absolute error maps. This complementary evaluation provides insight into the spatial distribution of reconstruction errors. It facilitates the interpretation of model performance over different land-cover classes, including cultivated fields, harvested parcels, roads, vegetation strips, isolated trees, and transition zones.

3. Results

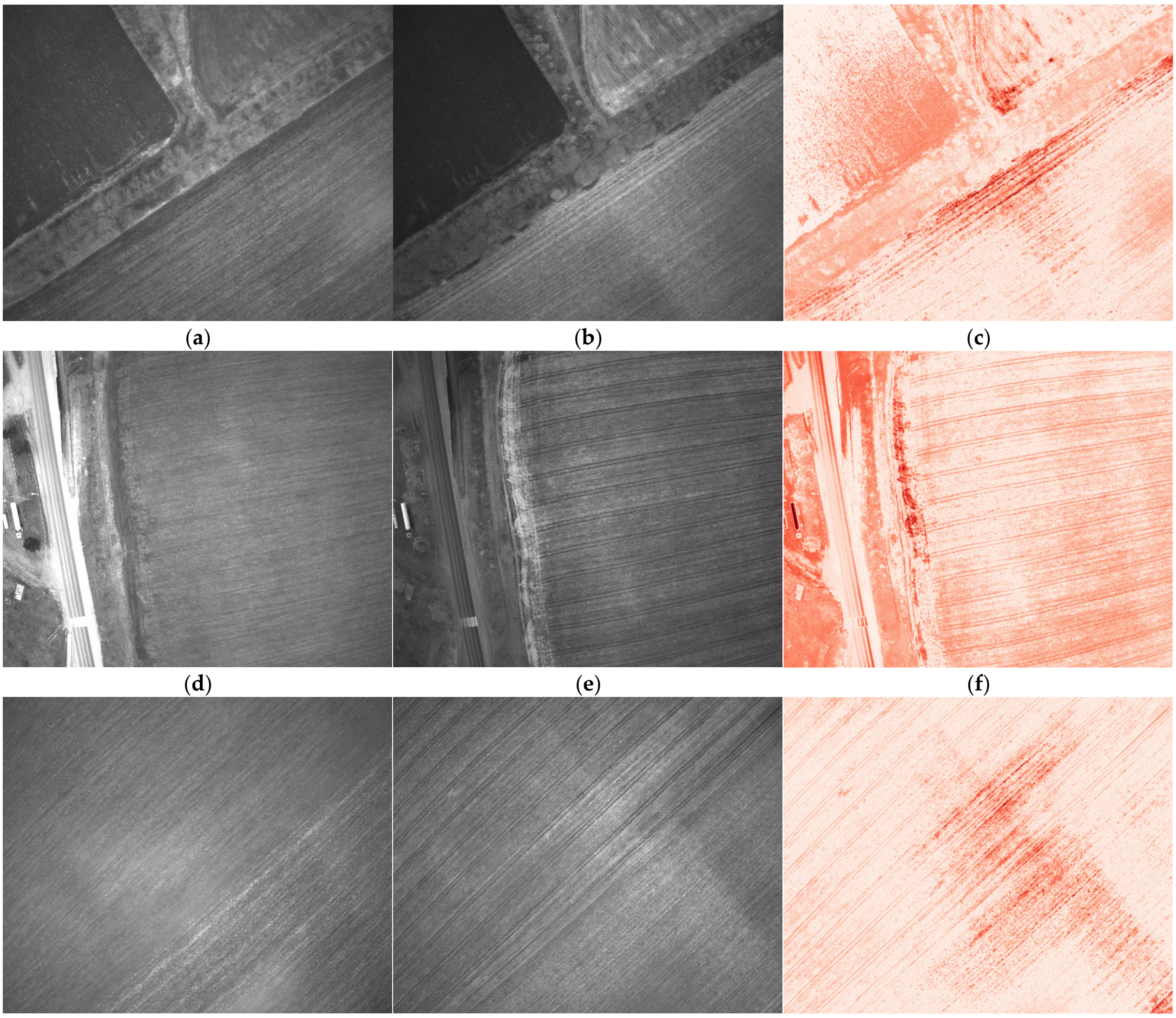
3.1. Spatial Assessment of NIR Reconstruction Errors

The spatial distribution of NIR reconstruction errors was examined using representative images from the training and within-mission held-out test subsets. Figure 4 presents the normalized red-band input image, the corresponding measured NIR-band image, and the absolute reconstruction error map for a representative image from the training subset. Figure 5 presents the same three components for a representative image from the within-mission held-out test subset. The absolute error maps were calculated as the pixel-wise absolute difference between the measured and reconstructed normalized NIR-band digital-intensity values, thereby illustrating the spatial distribution of reconstruction errors across the two representative images.



(**a**) (**b**) (**c**)

Figure 4. Representative reconstruction results for the training subset: (**a**) normalized red-band input image; (**b**) corresponding measured NIR-band image; (**c**) absolute error map between the measured and reconstructed normalized NIR-band digital intensities.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

(**g**) (**h**) (**i**)

Figure 5. Representative spatial reconstruction-error results for three images from the within-mission held-out test subset. The rows illustrate different agricultural surface configurations. The columns show (**a**,**d**,**g**) normalized red-band input images; (**b**,**e**,**h**) corresponding measured NIR-band images; and (**c**,**f**,**i**) absolute error maps between measured and reconstructed normalized NIR-band digital intensities.

https://doi.org/10.3390/rs18173015

The normalized red-band and measured NIR-band images show that agricultural fields, vegetation strips, roads, field boundaries, and other landscape features are clearly distinguishable in the selected representative scenes. The corresponding absolute error maps show that reconstruction differences are spatially nonuniform and vary according to the characteristics of the observed surfaces.

Lower reconstruction differences are observed across several relatively homogeneous agricultural surfaces, whereas larger differences are concentrated primarily along field boundaries, roads, vegetation strips, and transition zones characterized by abrupt digital-intensity changes. These regions represent more challenging reconstruction conditions because the proposed model performs independent pixel-wise predictions without explicitly incorporating information from neighboring pixels.

Therefore, the spatial error maps indicate that the model reconstructs normalized NIR-band digital intensity more consistently over relatively homogeneous surfaces, whereas larger discrepancies occur in heterogeneous and transitional regions. This spatial behavior illustrates the limitations of applying a single global pixel-wise red-to-NIR mapping to surfaces with different spectral characteristics.

3.2. Quantitative Reconstruction Performance

The quantitative performance of the proposed MLP was evaluated by comparing its predicted NIR digital intensities with the corresponding measured values from the within-mission held-out test subset. To contextualize its performance, the MLP was compared with a constant mean predictor, linear regression, second-degree polynomial regression, and Random Forest regression. All models were evaluated on the same held-out test data using MSE, RMSE, MAE, R^2, and Pearson's correlation coefficient. The comparative results are presented in Table 3, while the distribution of the pixel-wise squared errors produced by the MLP is illustrated in Figure 6.

Table 3. Predictive performance of the proposed MLP and baseline regression models on the within-mission held-out test subset.

Model	MSE	RMSE	MAE	R^2	Pearson r
Mean predictor	0.013834	0.117620	0.094462	0.000000	N/A
Linear regression	0.013586	0.116557	0.094653	0.017994	0.134172
Polynomial regression (degree 2)	0.011482	0.107152	0.086679	0.170073	0.412442
Random Forest regressor	0.010319	0.101585	0.079824	0.254077	0.504062
Proposed MLP	0.010329	0.101632	0.079883	0.253383	0.683637

As shown in Table 3, the mean predictor yielded an R^2 value of 0.000000, representing the reference performance obtained without using the red-band input. Its undefined Pearson correlation coefficient results from the constant predicted values, which have zero variance. Therefore, Pearson's r is reported as N/A (not applicable) for this baseline.

Linear regression provided only a limited improvement over the mean predictor, obtaining an RMSE of 0.116557, an R^2 of 0.017994, and a Pearson correlation coefficient of 0.134172. Second-degree polynomial regression achieved lower prediction errors, reducing RMSE to 0.107152 and MAE to 0.086679, while increasing R^2 to 0.170073 and Pearson's r to 0.412442. These results indicate that a simple linear function cannot adequately represent the relationship between the red and NIR digital intensities and that nonlinear modeling provides a meaningful improvement.

Among the evaluated models, the Random Forest regressor produced the lowest prediction errors, with an MSE of 0.010319, an RMSE of 0.101585, and an MAE of 0.079824. It also achieved the highest R^2, at 0.254077. The proposed MLP produced nearly identical error-based results: an MSE of 0.010329, an RMSE of 0.101632, an MAE of 0.079883, and an R^2 of 0.253383. However, the MLP achieved the highest Pearson correlation coefficient among the evaluated models (r = 0.683637), compared with 0.504062 for Random Forest, indicating a stronger linear association between the MLP predictions and the measured NIR digital intensities.

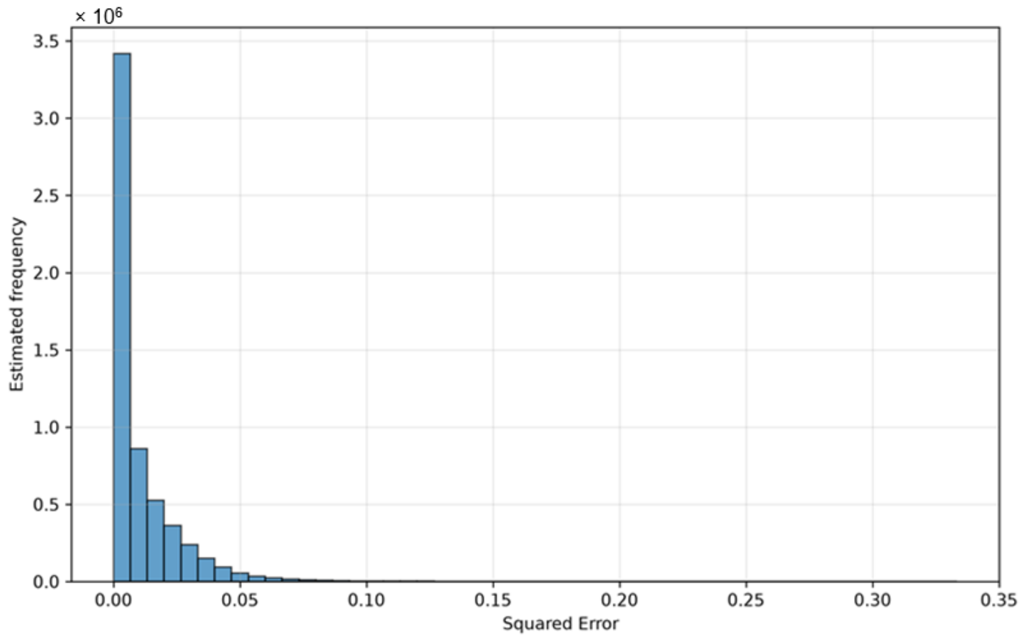


Figure 6. Distribution of the pixel-wise squared errors between the measured and reconstructed normalized NIR-band digital intensities in the within-mission held-out test subset.

Therefore, both nonlinear machine learning approaches outperformed the mean, linear, and second-degree polynomial references. Random Forest achieved marginally lower error values and a marginally higher R^2, whereas the proposed MLP achieved a substantially higher Pearson correlation coefficient. The very small differences in MSE, RMSE, MAE, and R^2 indicate that the two nonlinear models provided practically comparable error-based performance on the evaluated test subset. Nevertheless, the moderate R^2 values show that red-band intensity alone explains only part of the variability in the measured NIR data.

The statistical indicators are complemented by the error distribution presented in Figure 6, which provides a more detailed characterization of the reconstruction quality. The histogram reveals a strongly right-skewed distribution, with the overwhelming majority of pixel-wise squared errors concentrated near zero. Only a relatively small proportion of pixels exhibit larger reconstruction errors, forming a gradually decreasing tail toward higher error values. This distribution indicates that prediction errors are generally low across most pixels, with large deviations observed only in a limited number of observations.

The concentration of squared errors near zero indicates that many test pixels had relatively small reconstruction differences. Consequently, the histogram complements the global statistical metrics by showing that the reported MSE and RMSE values summarize a predominantly right-skewed error distribution in which relatively large errors occur less frequently.

The combined quantitative evidence demonstrates that the proposed lightweight MLP captures a meaningful nonlinear relationship between the Red and NIR digital intensities. Despite relying on only one red-band pixel value and excluding additional spectral and spatial information, the model reproduced relevant NIR variations and achieved error-based performance comparable to that of the substantially larger Random Forest model. These results support the feasibility of using reconstructed normalized NIR digital intensities as an accessible source of approximate spectral information for vegetation-oriented and other application-driven remote sensing analyses.

3.3. Computational Characteristics

The proposed MLP comprises only 609 trainable parameters. A single pixel-wise prediction requires approximately 560 multiplications, 560 additions, and 48 ReLU evaluations. In addition to its architectural compactness, the storage requirements of the trained MLP were compared with those of the evaluated baseline models. The resulting saved model sizes are presented in Table 4.

Table 4. Computational characteristics of the evaluated regression models.

Model	Training Time (s)	Inference Time (s)	Saved Model Size
Linear regression	24.250	0.003410	0.403 KB
Polynomial regression (degree 2)	58.500	0.003759	0.411 KB
Random Forest regressor	912.940	0.100530	1712.230 KB
Proposed MLP	4730.455	1.398822	31.401 KB

As shown in Table 4, linear and second-degree polynomial regression had the smallest storage requirements because they retained only a limited number of fitted coefficients. The proposed MLP occupied approximately 31.4 KB in its saved form, whereas the Random Forest regressor required approximately 1712.23 KB. Consequently, the saved MLP was approximately 56 times smaller than the Random Forest model. Although the MLP was larger than the two parametric regression baselines, it provided substantially better predictive performance than these simpler models, as reported in Table 3.

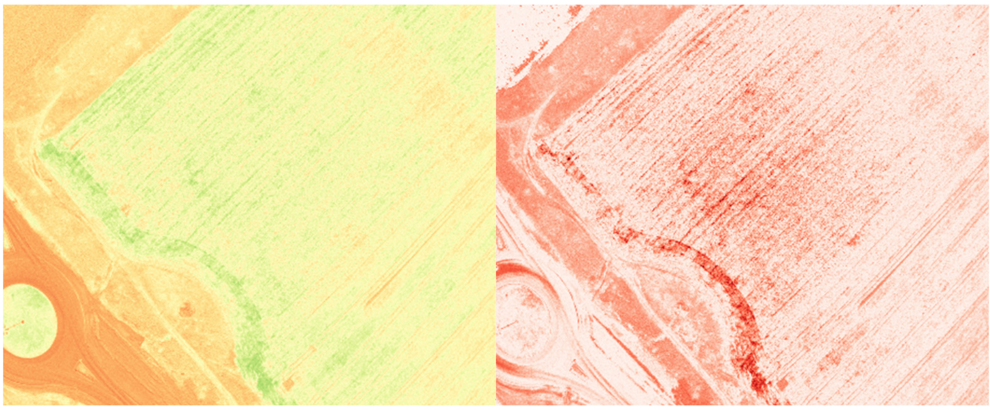
All execution times reported in Table 4 were measured without GPU acceleration on the same system equipped with an 11th Gen Intel Core i5-1135G7 processor and 8 GB of RAM. Training times were measured using the complete development dataset comprising 252 images, whereas inference times were measured for a single held-out test image with a spatial resolution of 1600×1300 pixels. Under these experimental conditions, the proposed MLP required 4730.455 s for training and 1.398822 s to process the test image. In comparison, the Random Forest regressor required 912.940 s for training and 0.100530 s for inference on the same image, while the linear and polynomial regression models exhibited substantially shorter execution times. Thus, the compact storage representation of the MLP did not translate into faster training or inference under the current software and hardware configuration.

Therefore, the results demonstrate that the principal computational advantage of the proposed MLP lies in its architectural compactness and reduced storage requirement relative to the Random Forest baseline. However, saved model size and execution time represent distinct computational characteristics. Although the MLP was approximately 56 times smaller than Random Forest in terms of saved size, it required longer training and inference times under the evaluated conditions. The reported execution times depend on the hardware configuration, software implementation, dataset size, and runtime conditions and should therefore be interpreted as system-specific measurements rather than universally reproducible benchmarks.

3.4. NDVI Reconstruction Assessment

The influence of the proposed Red-to-NIR reconstruction methodology on a subsequent remote sensing product was further examined using the Normalized Difference Vegetation Index. Since NDVI is one of the most widely used vegetation indices for vegetation monitoring, precision agriculture, and environmental assessment, this analysis provides an application-oriented illustration of how NIR reconstruction errors may affect the resulting vegetation index, complementing conventional regression metrics.

To evaluate the influence of the reconstructed NIR band on vegetation-index generation, two NDVI maps were produced for each image in the held-out test subset. The first NDVI map was computed from the original NIR band acquired by the multispectral sensor. In contrast, the second was generated from normalized NIR digital intensities predicted by the proposed MLP, while preserving the original red spectral band. In addition, an absolute NDVI difference map was calculated to visualize the spatial distribution of reconstruction errors. Figure 7 presents the measured NDVI map and the corresponding absolute difference map for a representative test image.



(**a**) (**b**)

Figure 7. Illustrative NDVI assessment for a representative image from the within-mission held-out test subset: (**a**) NDVI calculated using the measured NIR-band data; (**b**) absolute difference between NDVI calculated using measured and reconstructed NIR-band data.

As illustrated in Figure 7, the measured NDVI map shows the principal spatial distribution of vegetation in the representative test image. The absolute NDVI difference map indicates that the differences between measured and reconstructed NDVI are spatially variable and appear more pronounced along some field boundaries and heterogeneous transition regions, where abrupt spectral changes may make pixel-wise prediction more challenging. Lower visual differences can be observed in several relatively homogeneous agricultural areas. However, because this comparison is qualitative and no aggregated NDVI error metrics or land-cover-specific assessments were calculated, these observations should not be interpreted as quantitative evidence of reconstruction accuracy.

To complement the image-based assessment, a theoretical sensitivity analysis was conducted to illustrate how a fixed perturbation in the reconstructed NIR digital intensity may propagate into the calculated NDVI. Normalized red and NIR digital intensities within the interval [0, 1] were considered. For each valid combination for which the NDVI denominator was nonzero, the absolute difference between the original NDVI and the NDVI calculated after adding a fixed positive perturbation to the NIR value was determined. A perturbation of $\Delta\mathrm{NIR} = 0.1$ was selected because it approximates the test-set RMSE of the proposed MLP (0.101632). This value represents an illustrative perturbation on the normalized digital-intensity scale and should not be interpreted as an absolute surface-reflectance difference or as the actual prediction error of every test pixel.

The propagated NDVI difference was calculated as follows:

$$\Delta\mathrm{NDVI} = \left| \frac{\mathrm{NIR} + \Delta\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \Delta\mathrm{NIR} + \mathrm{R}} - \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}} \right|, \quad (15)$$

As shown in Figure 8, the effect of a fixed NIR perturbation on NDVI is nonlinear and depends on both the red and NIR digital intensities. The largest absolute NDVI differences occur when the sum of the red and NIR intensities is relatively small, particularly at low NIR values. As NIR intensity increases, the absolute NDVI difference generally decreases. Thus, NIR errors of similar magnitude may produce different NDVI deviations depending on the spectral values of the analyzed pixel. This theoretical analysis illustrates the sensitivity of NDVI to NIR reconstruction errors.

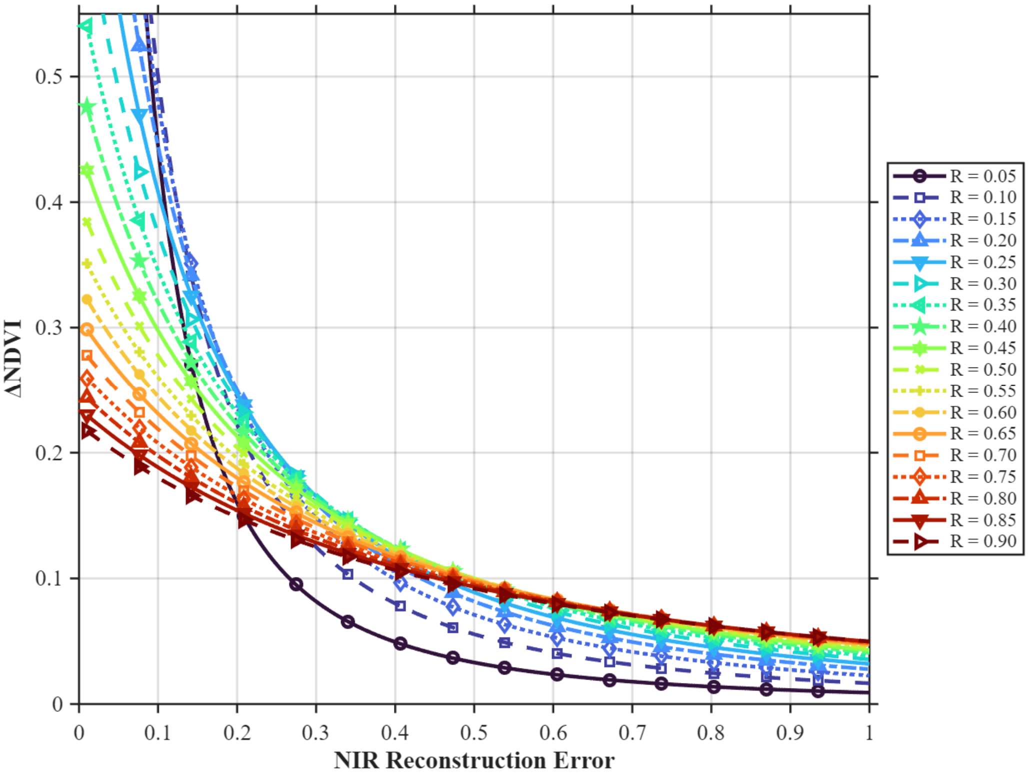


Figure 8. Theoretical propagation of a fixed positive NIR perturbation (ΔNIR = 0.1) into the resulting NDVI variation (ΔNDVI) as a function of normalized NIR digital intensity. Each colored curve corresponds to a different normalized Red-band digital-intensity value, denoted by R (R = 0.05–0.90).

Therefore, the NDVI-based assessment provides an application-oriented illustration of how the reconstructed NIR data affect the spatial representation of vegetation. The preservation of broad vegetation patterns indicates potential value for qualitative assessment, preliminary crop monitoring, and exploratory decision support in settings where dedicated multispectral resources are unavailable or economically impractical. The approach is not proposed as a universal substitute for calibrated NIR measurements, particularly in applications requiring high radiometric accuracy. Further quantitative NDVI validation, including error metrics, agreement analysis, and land-cover-specific assessment, will determine the operational contexts and accuracy requirements for which the reconstructed information is most suitable.

4. Discussion

4.1. Reconstruction Performance and Relation to Existing Studies

The present study investigated the feasibility of reconstructing the near-infrared spectral band from the red spectral band using a lightweight multilayer perceptron operating exclusively at the pixel level. Unlike most recent approaches, which employ computationally intensive convolutional neural networks, encoder–decoder architectures, generative adversarial networks, or transformer-based models, the proposed methodology intentionally eliminates spatial context and focuses solely on the nonlinear relationship between the normalized red-band and NIR-band digital intensities of individual pixels [11,14–16,19–25]. This simplified formulation enables investigation of the pixel-level statistical relationship between the two bands while maintaining low architectural complexity.

The experimental results indicate that the MLP captured part of the pixel-level relationship between the normalized red-band and NIR-band digital intensities without using neighboring-pixel information. The reconstructed NIR imagery and the derived vegetation-index maps visually preserved the dominant spatial organization of agricultural parcels, vegetation strips, and transition zones, although local discrepancies remained visible. These observations suggest that the red-band input contains predictive information about the corresponding NIR-band values during the evaluated mission. However, the moderate coefficient of determination and the absence of cross-site or cross-date validation indicate that this relationship should not be interpreted as universally stable.

Recent advances in artificial intelligence have considerably improved spectral reconstruction from RGB and multispectral imagery. Zhao et al. [20] demonstrated that deep convolutional architectures can accurately reconstruct multispectral information for UAV-based crop phenotyping. In contrast, Yuan et al. [16] successfully employed conditional generative adversarial networks to generate artificial NIR information and vegetation indices from visible imagery. Moscovini et al. [15] proposed a machine learning framework for pixel-wise NDVI estimation from calibrated RGB images, while Gong et al. [24] investigated lightweight neural architectures for spectral super-resolution. Additional studies have shown that encoder–decoder and transformer-based architectures can further improve reconstruction accuracy by jointly exploiting spectral and spatial information [18]. Although these approaches generally achieve higher predictive performance, they also require substantially larger training datasets, considerably more trainable parameters, and significantly higher computational resources.

4.2. Baseline Comparison and Computational Characteristics

The proposed methodology was intentionally designed to prioritize computational compactness and implementation simplicity rather than maximizing predictive accuracy. The MLP contains 609 trainable parameters and occupies approximately 31.4 KB in saved form, compared with approximately 1.712 MB for the evaluated Random Forest regressor. Thus, the saved MLP was approximately 56 times smaller than the Random Forest model. Nevertheless, the MLP required 4730.455 s for training and 1.398822 s for inference, exceeding the corresponding execution times of all evaluated baseline models under the reported experimental conditions. These results show that architectural complexity, saved model size, runtime memory use, and execution time represent distinct computational characteristics. Therefore, the main computational advantage demonstrated by the MLP is its compact architecture and reduced saved-model size relative to Random Forest, rather than faster training or inference.

Comparison with the baseline regression models provides a more nuanced interpretation of the proposed method. The MLP substantially improved upon the mean predictor, linear regression, and second-degree polynomial regression. Compared with Random Forest, the MLP produced nearly identical MSE, RMSE, MAE, and R^2 values, while achieving a considerably higher Pearson correlation coefficient. Random Forest remained marginally better according to the error-based metrics, but the differences were very small. Therefore, the proposed MLP provides predictive performance comparable to that of the evaluated Random Forest while retaining a substantially smaller architecture and saved model size. Its contribution lies in providing a compact neural baseline for pixel-wise Red-to-NIR reconstruction rather than establishing state-of-the-art predictive performance.

https://doi.org/10.3390/rs18173015

Unlike image-to-image reconstruction approaches that use several visible bands and spatial context, the proposed MLP evaluates a deliberately constrained single-input configuration. Its contribution is therefore characterized by the experimentally observed trade-off between available spectral information, predictive performance, and architectural compactness.

4.3. Error Sources and NDVI-Based Interpretation

The observed predictive performance reflects the amount of information available when NIR digital intensity is estimated from a single Red-band value. Similar red intensities may correspond to different NIR intensities due to variations in vegetation structure, soil background, shadows, illumination, and land cover type. Although land-cover-specific models could potentially represent these heterogeneous relationships more effectively, their development would require reliable land-cover labels or an additional classification stage. Consequently, the conditional variability of NIR for a given Red intensity defines the attainable accuracy of deterministic single-band reconstruction. Within this boundary, the proposed MLP offers a compact approximation that can provide access to NIR-related information in applications where direct multispectral acquisition and specialized processing are unavailable or economically impractical.

Based on the present results, the proposed method provides its most consistent estimates under acquisition conditions similar to those represented in the training data, particularly over relatively homogeneous agricultural surfaces and under comparable sensor and illumination conditions. The spatial error analysis further indicates that heterogeneous transition regions, mixed pixels, and abrupt land-cover boundaries represent more challenging prediction conditions because similar red-band responses may correspond to different NIR values. Within this experimentally defined scope, the reconstructed NIR values provide data-driven estimates of NIR-related information that may support qualitative vegetation assessment and exploratory remote sensing applications. These estimates are intended to complement preliminary, resource-constrained assessment workflows rather than to replace calibrated NIR measurements in applications requiring high radiometric accuracy.

An important contribution of the present study is the application-oriented assessment based on NDVI reconstruction. Many previous investigations assess spectral reconstruction primarily through numerical regression metrics such as MSE, RMSE, or the coefficient of determination [14–17,19–25]. While these indicators provide valuable information regarding prediction accuracy, they do not directly illustrate how reconstruction errors may affect subsequent remote-sensing products. In contrast, the present work additionally examines how the reconstructed NIR band influences the generation of a vegetation index widely used in precision agriculture and environmental monitoring [17]. In the representative test image, some broad spatial vegetation patterns remained identifiable in the NDVI product generated using the reconstructed NIR band. Because NDVI remains one of the most widely adopted indicators of crop vigor, biomass, canopy development, ecosystem condition, and plant health [7–10], this assessment illustrates the potential downstream relevance of NIR reconstruction errors. The importance of reliable vegetation-index retrieval has likewise been emphasized in recent studies on hyperspectral remote sensing. Plajer et al. [11] demonstrated the effectiveness of hyperspectral imagery for accurate NDVI computation, while Racoviteanu et al. [12] proposed spectral aggregation strategies for retrieving Sentinel-2 vegetation indices from PRISMA hyperspectral data. These studies further confirm the growing importance of artificial intelligence and advanced spectral processing for vegetation-index retrieval and for extracting vegetation-related information from remotely sensed imagery [11,12].

4.4. Accessibility, Practical Relevance, and Future Research

From an accessibility perspective, the significance of the proposed framework extends beyond reducing model size. Multispectral remote sensing may involve not only the acquisition cost of dedicated sensors but also radiometric processing software, computational infrastructure, and specialized expertise. These combined requirements can restrict the adoption of precision-agriculture technologies by small farms and individual producers. By investigating how NIR-related information can be approximated from a minimal visible-band input using a compact model, the present study provides a technical foundation for lowering these entry barriers. Subject to further validation with conventional RGB cameras, such an approach may enable qualitative vegetation assessment, preliminary crop monitoring, and early identification of areas requiring closer inspection without the full financial and technical infrastructure associated with conventional multispectral workflows. Accordingly, the compact formulation provides a technical basis for further evaluation within accessible, resource-constrained vegetation-monitoring workflows.

Several limitations should nevertheless be acknowledged. First, the experimental dataset was acquired using a single UAV multispectral platform over agricultural areas within a single geographical region of the Republic of Moldova. Additional experiments involving different crops, climatic conditions, acquisition seasons, geographical regions, and environmental conditions would provide a more comprehensive assessment of the robustness and transferability of the proposed methodology. Second, the present study intentionally used only the red spectral band as input to isolate its nonlinear relationship with the NIR response. Future investigations may examine whether incorporating additional visible bands, such as blue, green, or red-edge digital intensities, further improves reconstruction accuracy while maintaining a lightweight network architecture. Finally, although the adopted MLP provides a compact architecture and small saved-model size, future research may investigate implementation-level optimizations and hybrid lightweight architectures that combine pixel-wise spectral regression with limited spatial feature extraction. From a system-integration perspective, the proposed model could also be incorporated as a dedicated processing service within a modular UAV software architecture, enabling spectral estimation to operate alongside flight-control, data-acquisition, and geospatial-processing components [30]. Such developments could improve reconstruction performance and execution efficiency without sacrificing the accessibility-oriented design of the proposed framework [14–16,19–25].

Therefore, the results position the proposed MLP as a compact empirical baseline that balances attainable reconstruction performance with minimal spectral-input and architectural requirements. This balance is particularly relevant to accessible vegetation assessment and preliminary crop monitoring, while broader operational use requires validation across sensors, sites, and seasons, as well as with conventional RGB imaging platforms.

5. Conclusions

This study evaluated a compact pixel-wise MLP for approximating normalized NIR-band digital intensity from a single red-band value without using additional spectral bands or spatial-context information. The model contains 609 trainable parameters and was investigated as a lightweight empirical baseline rather than as a novel neural-network architecture. On the within-mission held-out test subset, it achieved an MSE of 0.010329, an RMSE of 0.101632, an MAE of 0.079883, an R^2 of 0.253383, and a Pearson correlation coefficient of 0.683637. Its error-based performance was comparable to that of the evaluated Random Forest regressor. However, the moderate R^2 indicates that a single red-band value cannot fully account for the variability in the NIR response. The main contribution of this work is therefore the empirical characterization of the predictive information that can be extracted from a single red-band observation using a minimal nonlinear model, together with the resulting trade-off between predictive accuracy, available input information, and architectural compactness.

The reconstructed data preserved broad spatial vegetation patterns in the illustrative NDVI comparison, indicating potential usefulness for qualitative vegetation assessment and preliminary remote sensing analysis when directly measured NIR data are unavailable. Nevertheless, the proposed method should not be regarded as a replacement for dedicated multispectral sensors or calibrated NIR measurements, particularly in applications requiring accurate quantitative spectral or vegetation-index information. Moreover, because all data originated from a single UAV mission and spatial blocking was not applied, the results demonstrate within-mission held-out performance rather than generalization across sites, seasons, crop types, or acquisition conditions.

Future research should validate the approach using spatially and temporally independent datasets acquired across diverse landscapes, crops, growth stages, seasons, and illumination conditions. It should also include quantitative validation of reconstructed vegetation indices, land-cover-specific error analysis, testing with conventional low-cost RGB cameras, and the investigation of compact models incorporating additional visible bands or limited spatial information.

Author Contributions: Conceptualization, V.B. and V.C.; Investigation, N.D., I.C., A.U. and M.G.; Methodology, V.B. and V.S.; Software, N.D., I.C. and A.U.; Validation, N.D., V.C., V.S. and M.G.; Formal Analysis, V.S. and M.G.; Writing—Original Draft Preparation, V.C., N.D. and M.G.; Writing—Review and Editing, M.G. and V.C.; Visualization, N.D. and A.U.; Project Administration, V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Research Excellence Stimulation Project "Research on Intelligent Methods and Efficient Technologies for Monitoring Forest Areas, Perennial Crops, and Water Surfaces Using Unmanned Aerial Vehicles", grant number 25.80012.5007.71SE.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors would like to thank Cart Engineering SRL for providing the multispectral UAV data used in this research. The availability of these data significantly contributed to the experimental validation of the proposed methodology.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rotaru, L.; Cojuhari, I.; Gutu, M.; Kazak, A. Mathematical Models and System-Level Analysis of Modern Geolocation Technologies. In Proceedings of the 2026 International Conference on Development and Application Systems (DAS), Suceava, Romania, 21 May 2026; pp. 134–141.
2. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A Review on UAV-Based Applications for Precision Agriculture. *Information* **2019**, *10*, 349. [CrossRef]
3. Aasen, H.; Honkavaara, E.; Lucieer, A.; Zarco-Tejada, P.J. Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows. *Remote Sens.* **2018**, *10*, 1091. [CrossRef]
4. Maes, W.H.; Steppe, K. Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [CrossRef] [PubMed]
5. Shahi, T.B.; Xu, C.-Y.; Neupane, A.; Guo, W. Machine Learning Methods for Precision Agriculture with UAV Imagery: A Review. *Era* **2022**, *30*, 4277–4317. [CrossRef]
6. Gutu, M.; Rotaru, L.; Alexei, V.; Kapusteanski, M. Advanced Drone-Based Monitoring of Agricultural, Forestry, and Aquatic Ecosystems: Technical Framework. *J. Eng. Sci.* **2025**, *32*, 108–121. [CrossRef]

7. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
8. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In *Proceedings of the Third Earth Resources Technology Satellite-1 (ERTS-1) Symposium*; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 1974; Volume 1.
9. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1353691. [CrossRef]
10. Zhang, H.K.; Roy, D.P.; Yan, L.; Li, Z.; Huang, H.; Vermote, E.; Skakun, S.; Roger, J.-C. Characterization of Sentinel-2A and Landsat-8 Top of Atmosphere, Surface, and Nadir BRDF-Adjusted Reflectance and NDVI Differences. *Remote Sens. Environ.* **2018**, *215*, 482–494. [CrossRef]
11. Plajer, I.C.; Băicoianu, A.; Majercsik, L.; Ivanovici, M. NDVI Computation from Hyperspectral Images. In Proceedings of the 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Athens, Greece, 31 October 2023; pp. 1–5.
12. Racoviţeanu, A.; Florea, C.; Niţu, A.; Ivanovici, M. PRISMA Hyperspectral Data Aggregation for Retrieval of Sentinel-2 Vegetation Indices. In Proceedings of the IGARSS 2025—2025 IEEE International Geoscience and Remote Sensing Symposium, Brisbane, Australia, 3 August 2025; pp. 8891–8895.
13. Collin, A.; James, D.; Mury, A.; Letard, M.; Guillot, B. Predicting the Infrared UAV Imagery Over the Coast. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 149–156. [CrossRef]
14. 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.* **2023**, *209*, 107833. [CrossRef]
15. Moscovini, L.; Ortenzi, L.; Pallottino, F.; Figorilli, S.; Violino, S.; Pane, C.; Capparella, V.; Vasta, S.; Costa, C. An Open-Source Machine-Learning Application for Predicting Pixel-to-Pixel NDVI Regression from RGB Calibrated Images. *Comput. Electron. Agric.* **2024**, *216*, 108536. [CrossRef]
16. Yuan, X.; Tian, J.; Reinartz, P. Generating Artificial Near Infrared Spectral Band from RGB Image Using Conditional Generative Adversarial Network. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *V-3-2020*, 279–285. [CrossRef]
17. Diykh, M.; Ali, M.; Jamei, M.; Abdulla, S.; Uddin, M.P.; Farooque, A.A.; Labban, A.H.; Alabdally, H. Empirical Curvelet Transform Based Deep DenseNet Model to Predict NDVI Using RGB Drone Imagery Data. *Comput. Electron. Agric.* **2024**, *221*, 108964. [CrossRef]
18. Arevalo-Ramirez, T.; Menéndez, O.; Villacrés, J.; Guevara, J.; Guamán-Rivera, R.; Demarco, R.; Auat Cheein, F. Assessment of Light Environment Conditions for Reconstruction of Multispectral Images by Conditional Adversarial Networks. *Biosyst. Eng.* **2025**, *260*, 104305. [CrossRef]
19. Doornbos, J.; Babur, Ö.; Valente, J. Evaluating Generalization of Methods for Artificially Generating NDVI from UAV RGB Imagery in Vineyards. *Remote Sens.* **2025**, *17*, 512. [CrossRef]
20. Zhao, J.; Kumar, A.; Banoth, B.N.; Marathi, B.; Rajalakshmi, P.; Rewald, B.; Ninomiya, S.; Guo, W. Deep-Learning-Based Multispectral Image Reconstruction from Single Natural Color RGB Image—Enhancing UAV-Based Phenotyping. *Remote Sens.* **2022**, *14*, 1272. [CrossRef]
21. Darmawan, H.; Yuliana, M.; Hadi, M.Z.S.; Sudibyo, R.W. Fine-Tuning GAN Models with Unpaired Aerial Images for RGB to NDVI Translation in Vegetation Index Estimation. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2024**, *14*, 1654–1662. [CrossRef]
22. Picon, A.; Bereciartua-Perez, A.; Eguskiza, I.; Romero-Rodriguez, J.; Jimenez-Ruiz, C.J.; Eggers, T.; Klukas, C.; Navarra-Mestre, R. Deep Convolutional Neural Network for Damaged Vegetation Segmentation from RGB Images Based on Virtual NIR-Channel Estimation. *Artif. Intell. Agric.* **2022**, *6*, 199–210. [CrossRef]
23. Lee, G.; Ghimire, D.; Kim, D.; Cho, S.; Kim, B.; Jeong, S. FUSE-Net: Multi-Scale CNN for NIR Band Prediction from RGB Using GNDVI-Guided Green Channel Enhancement. *Sensors* **2025**, *25*, 4076. [CrossRef]
24. Gong, L.; Yihong, L.; Zhiyuan, Z.; Gang, W.; Jiaojiao, L. Multifiltering MLP for Spectral Super-Resolution with Remote Sensing Image Verification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 16646–16658. [CrossRef]
25. Wang, J.; Chen, C.; Wang, J.; Yao, Z.; Wang, Y.; Zhao, Y.; Sun, Y.; Wu, F.; Han, D.; Yang, G.; et al. NDVI Estimation Throughout the Whole Growth Period of Multi-Crops Using RGB Images and Deep Learning. *Agronomy* **2024**, *15*, 63. [CrossRef]
26. Davidson, C.; Jaganathan, V.; Sivakumar, A.N.; Czarnecki, J.M.P.; Chowdhary, G. NDVI/NDRE Prediction from Standard RGB Aerial Imagery Using Deep Learning. *Comput. Electron. Agric.* **2022**, *203*, 107396. [CrossRef]
27. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization 2014. *arXiv* **2017**, arXiv:1412.6980.
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1026–1034.

29. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
30. Gorgos, I.; Cojocaru, V.; Carbune, V.; Gutu, M. A Microservice-Based Architecture for Unmanned Aerial Vehicle Control Systems. In Proceedings of the 2026 International Conference on Development and Application Systems (DAS), Suceava, Romania, 21 May 2026; pp. 317–324.