Hyperspectral Canopy Reflectance and Machine Learning for Threshold-Based Classification of Aphid-Infested Winter Wheat

Skendžić, Sandra; Novak, Hrvoje; Zovko, Monika; Pajač Živković, Ivana; Lešić, Vinko; Maričević, Marko; Lemić, Darija

Source / Izvornik: **Remote Sensing, 2025, 5, 1 - 24**

Journal article, Accepted version
Rad u časopisu, Završna verzija rukopisa prihvaćena za objavljivanje (postprint)

https://doi.org/10.3390/rs17050929

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:168:993023

remote sensing

MDPI

Article

Hyperspectral Canopy Reflectance and Machine Learning for Threshold-Based Classification of Aphid-Infested Winter Wheat

Sandra Skendžić [1,*], Hrvoje Novak [2], Monika Zovko [3], Ivana Pajač Živković [1], Vinko Lešić [2], Marko Maričević [4] and Darija Lemić [1]

[1] Department of Agricultural Zoology, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, 10000 Zagreb, Croatia; ipajac@agr.hr (I.P.Ž.); dlemic@agr.hr (D.L.)
[2] University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia; hrvoje.novak@fer.hr (H.N.); vinko.lesic@fer.hr (V.L.)
[3] Department of Soil Amelioration, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25, 10000 Zagreb, Croatia; mzovko@agr.hr
[4] Bc Institute for Breeding and Production of Field Crops, Rugvica, Dugoselska 7, 10370 Dugo Selo, Croatia; marko.maricevic@bc-institut.hr
* Correspondence: sskendzic@agr.hr

Abstract: Aphids are significant pests of winter wheat, causing damage by feeding on plant sap and reducing crop yield and quality. This study evaluates the potential of hyperspectral remote sensing (350–2500 nm) and machine learning (ML) models for classifying healthy and aphid-infested wheat canopies. Field-based hyperspectral measurements were conducted at three growth stages—T1 (stem elongation–heading), T2 (flowering), and T3 (milky grain development)—with infestation levels categorized according to established economic thresholds (ET) for each growth stage. Spectral data were analyzed using Uniform Manifold Approximation and Projection (UMAP); vegetation indices; and ML classification models, including Logistic Regression (LR), k-Nearest Neighbors (KNNs), Support vector machines (SVMs), Random Forest (RF), and Light Gradient Boosting Machine (LGBM). The classification models achieved high performance, with F1-scores ranging from 0.88 to 0.99, and SVM and RF consistently outperforming other models across all input datasets. The best classification results were obtained at T2 with an F1-score of 0.98, while models trained on the full spectrum dataset showed the highest overall accuracy. Among vegetation indices, the Modified Triangular Vegetation Index, MTVI (r_{pb} = −0.77 to −0.82), and Triangular Vegetation Index, TVI (r_{pb} = −0.66 to −0.75), demonstrated the strongest correlations with canopy condition. These findings underscore the utility of canopy spectra and vegetation indices for detecting aphid infestations above ET levels, allowing for a clear classification of wheat fields into "treatment required" and "no treatment required" categories. This approach provides a precise and timely decision making tool for insecticide application, contributing to sustainable pest management by enabling targeted interventions, reducing unnecessary pesticide use, and supporting effective crop protection practices.

Keywords: aphid infestation; winter wheat; canopy; spectral reflectance; vegetation indices; machine learning; classification; economic threshold

Academic Editors: Xinxin Liu, Bin Yang, Qian Shi and Lin Lei

Received: 31 January 2025
Revised: 27 February 2025
Accepted: 1 March 2025
Published: 5 March 2025

Citation: Skendžić, S.; Novak, H.; Zovko, M.; Pajač Živković, I.; Lešić, V.; Maričević, M.; Lemić, D. Hyperspectral Canopy Reflectance and Machine Learning for Threshold-Based Classification of Aphid-Infested Winter Wheat. *Remote Sens.* **2025**, *17*, 929. https://doi.org/10.3390/rs17050929

1. Introduction

Winter wheat (*Triticum aestivum* L.) is one of the most widely cultivated cereals globally, occupying nearly a third of European agricultural land [1]. However, its production is increasingly threatened by climate change and biotic stressors, particularly aphid infestations, which contribute to significant yield losses [2–4]. Aphids (Hemiptera: Aphididae) are major agricultural pests worldwide, causing crop damage by extracting photoassimilates and transmitting plant viruses [5], the most important ones in cereal crops being Barley Yellow Dwarf Virus (BYDV) and cereal yellow dwarf viruses (CYDVs) [6]. They are particularly problematic due to their high reproductive capacity, short life cycle, and polymorphic nature, leading to frequent outbreaks [7,8]. In Central and Southeastern Europe, cereals are attacked by more than seven aphid species, with four being predominant: English grain aphid (*Sitobion avenae*, F.), bird cherry-oat aphid (*Rhopalosiphum padi*, L.), rose-grain aphid (*Metopolophium dirhodum*, Walker), and greenbug aphid (*Schizaphis graminum*, Rondani), collectively known as the cereal aphid complex [3,9,10]. The latter species, *S. graminum*, has been increasing in importance due to climate change [11]. These aphids differ in feeding site preferences: *S. avenae* colonizes the ear and upper leaves [8], *R. padi* prefers the stem and lower leaf sheath [12], *M. dirhodum* feeds on the underside of leaves, and *S. graminum* primarily targets leaf tissues, often leading to chlorosis [13]. Aphid feeding disrupts the physiology of wheat by extracting sap, inducing chlorophyll loss and triggering oxidative stress, which leads to hypersensitive cell death and premature senescence [14]. This shortens the grain filling period, resulting in a lower grain weight and empty ears. Each day with a shortened vegetation period can lead to yield losses of up to 200 kg/ha [10]. Aphids feed exclusively on phloem sap with their stylet-like mouthparts, and their feeding diverts nutrients to the infested tissue, further increasing yield losses [15]. Severe infestations can cause direct yield reductions of 35–40% [16], while levels as low as 10 aphids per tiller have been linked to a 6% yield reduction [17,18].

Cereal aphid populations tend to increase rapidly in the field, often reaching levels where yield losses exceed treatment costs. Economic thresholds (ETs) for treatment vary according to the growth stage, with intervention recommended at 4 aphids per tiller from stem elongation to heading, 4–7 aphids per tiller during flowering, and 8–12 aphids per tiller during mid-grain development [19–22]. The effective control of aphids requires timely detection to prevent further spread and minimize yield losses. This is traditionally achieved through frequent scouting and, if necessary, the application of foliar insecticides [23]. While chemical control remains effective in reducing aphid populations [24], the indiscriminate use of pesticides increases production costs, contributes to environmental risks, and accelerates the development of resistance [25]. Given these limitations, there is an increasing need for advanced monitoring methods that improve efficiency and accuracy.

Traditional scouting methods are labor-intensive and spatially limited, hindering rapid aphid detection [26]. Remote sensing (RS) overcomes these limitations by using multi- or hyperspectral sensors on satellites, drones, or ground platforms to non-invasively collect spectral data, enabling the real-time monitoring of crop health and pest infestations. This enhances decision making efficiency in precision agriculture [27–30]. RS enables fast and scalable monitoring, enhancing decision making in precision agriculture [31–33]. Since pests are typically distributed in patches, precision agriculture technologies enable targeted, site-specific interventions, optimizing pest management and reducing unnecessary pesticide use [30,34].

The most important spectral ranges for plant studies include the visible range (VIS; 400–700 nm), the near-infrared (NIR; 750–1300 nm) and the shortwave infrared (SWIR; 1300–2500 nm), where stress responses are most evident [35]. However, due to the complexity of hyperspectral data, dimensionality reduction techniques such as Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) are applied to extract the most relevant spectral features, while noise reduction enhances

data quality [36–38]. PCA is a linear transformation method that projects data onto new orthogonal axes to maximize variance, compressing hyperspectral information while preserving key patterns. However, its linearity limits its ability to capture complex spectral variations [38]. In contrast, UMAP is a non-linear dimensionality reduction technique that preserves local and global data structures, making it highly effective for capturing complex, non-linear patterns in hyperspectral datasets [39–41]. RS sensors measure spectral reflectance, enabling vegetation index (VI) calculations, which are widely used for site-specific crop management and aphid detection [26,42–48]. The VI indicates mathematical combinations of reflectance values at different spectral bands, providing insights into plant health, biomass accumulation, and chlorophyll content. Mapping the VI enables a spatio-temporal analysis of crop conditions, supporting precision agriculture applications [49]. Ground-based hyperspectral data provide high-resolution measurements at the canopy level, allowing for precise characterization of plant stress responses. Additionally, they serve as a reference for calibrating large-scale RS applications, including drone- and satellite-based monitoring [50–53].

Machine learning (ML) combines computer science and statistics to recognize patterns and improve task performance through data learning [54]. In RS, ML algorithms provide a powerful approach for classifying high-dimensional spectral data, as they can model complex patterns, handle diverse input features, and operate without assumptions about the data distribution [55]. These capabilities make ML particularly effective for detecting and mapping vegetation stress, including pest-induced crop damage. ML models, such as Support vector machines (SVMs) and Random Forest (RF), are widely used in precision agriculture for their accuracy and robustness, while neural networks have also proven effective in crop classification tasks [56,57]. By integrating hyperspectral data with ML techniques, RS enables the early detection of aphid infestations, offering a scalable and efficient approach to pest management.

This study explores the application of hyperspectral data and ML models to classify aphid-infested winter wheat canopies across multiple growth stages, directly linking spectral data to the ET for treatment decisions. By leveraging advanced ML techniques and evaluating the VI's effectiveness at different growth stages, the findings provide new insights into optimizing RS-based aphid detection in winter wheat.

The specific objectives of this study include the following: (1) to analyze canopy-level spectral reflectance differences between aphid-infested wheat canopies above the ET and healthy canopies across different growth stages, (2) to identify VIs that exhibit the strongest correlations with aphid-induced stress in winter wheat, and (3) to develop ML classification models for distinguishing between healthy and aphid-infested wheat canopies, enabling treatment decision support.

2. Materials and Methods

2.1. Study Site and Experimental Design

The study was conducted on the experimental fields of the Bc Institute for Breeding and Production of Field Crops, d.d., in Zagreb, Croatia (45°44′50.0″N, 15°56′18.8″E) (Figure 1). The institute specializes in seed production and breeding programs for the most important arable crops. The experiment was conducted on a winter wheat field (variety: Bc Opsesija) consisting of eight plots, each 10 × 3 m in size. The wheat was sown in rows with a spacing of 12 cm and cultivated according to standard agronomic practices, including fertilization and fungicide application. The water supply for cultivation was solely dependent on natural rainfall. To evaluate the effects of aphid infestation on the reflectance of the wheat canopy, the plots were divided into two treatment groups: four plots were treated with insecticides to minimize aphid infestation, and the other four plots were left untreated to allow natural aphid infestation to develop. The study was conducted

from mid-April to mid-June 2022 and covered the main phenological stages for aphid infestation: stem elongation to heading (BBCH 30–55), flowering (BBCH 55–59), and milky grain development (BBCH 69–75). Visual assessments and spectral measurements were conducted on two consecutive days at each of these stages, focusing on the middle rows to minimize edge effects and ensure uniform data collection (Figure 2). The study site had been regularly infested by aphids in the past and was therefore suitable for this study. In addition, the environmental conditions in 2022 were favorable for aphid development, which contributed to a naturally high level of infestation in the untreated plots.

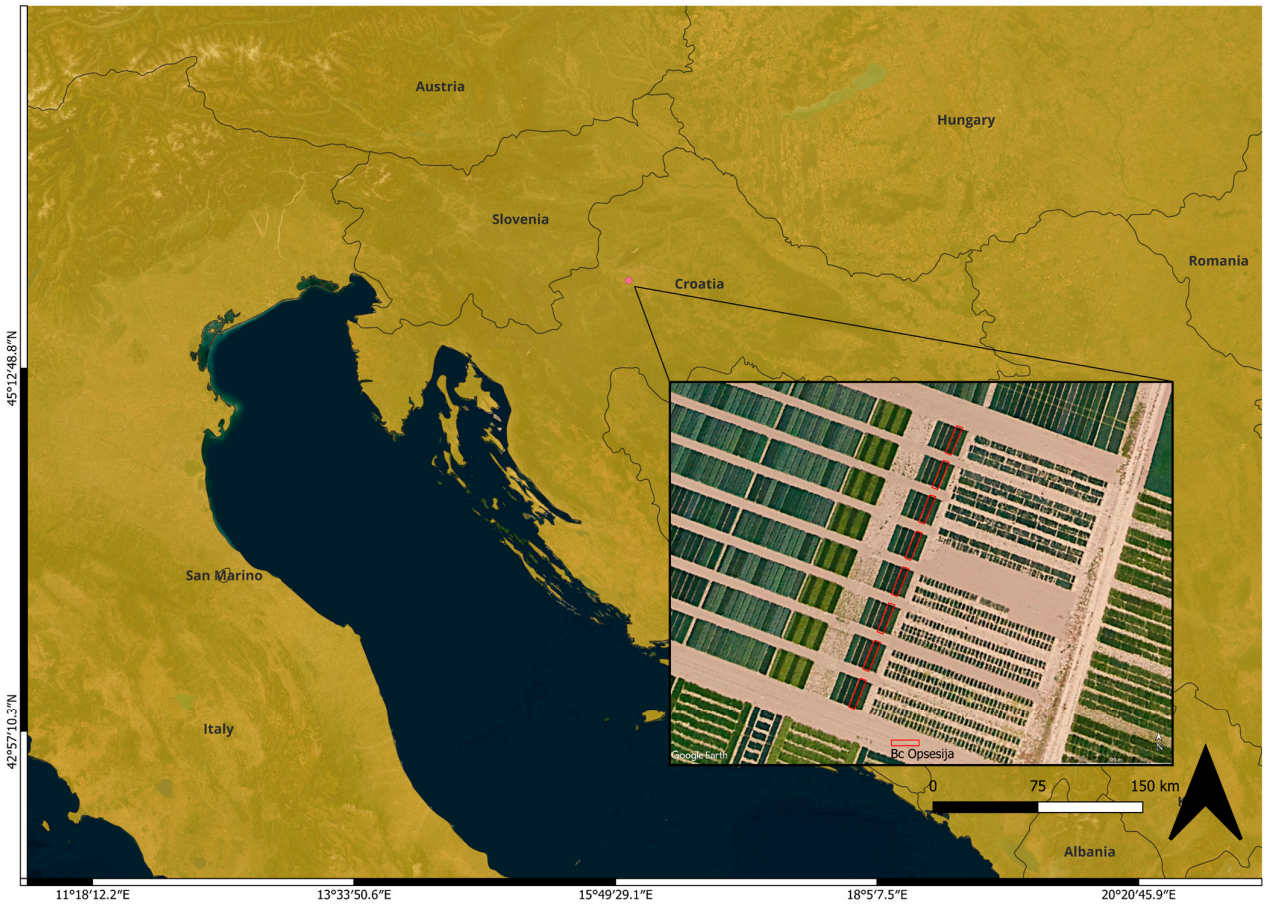

Figure 1. Study area and experimental design. Map created using QGIS and Google Earth Pro.

Figure 2. Fieldwork and canopy conditions during winter wheat growth stages. Spectral measurements were conducted using a portable spectroradiometer (left). Canopy images show healthy (A) and aphid-infested (B) wheat at three phenological stages: T1 (stem elongation–heading), T2 (flowering), and T3 (milky grain development).

2.2. Data Collection

2.2.1. Visual Assessment of Aphid Density

Scouting for cereal aphids began at the stem elongation stage and continued until the medium milk stage, in accordance with standard aphid population monitoring procedures. Field assessments were carried out on 8 plots described in Section 2.1. Close examination of the winter wheat stands was carried out using a magnifying glass to ensure accuracy. Whole tillers within the two inner rows of each plot were carefully inspected, and the number of aphids per tiller was recorded.

In addition to aphid counts, the species within the established aphid complex (*S. avenae*, *R. padi*, *M. dirhodum*, and *S. graminum*) were also identified during the assessments. The different aphid species preferred certain parts of the plant, with infestations often distributed over the ear, leaves, and stem. However, as the aphids were analyzed as a complex and not as individual species, their exact location on the plant was not a distinguishing feature. Aphid density was consistently assessed by the total number per tiller, which was classified as either below or above the economic threshold. The thresholds used in this study were derived from existing research [19–22] to ensure consistency with

established guidelines. The economic thresholds were determined based on the growth stage of the plant (see Table 1). Measurements were taken on two consecutive days during each of the three measurement periods: T1 (stem elongation to heading), T2 (flowering), and T3 (milky grain development). Tillers were then categorized based on aphid density relative to the ET. Tillers with aphid densities below the ET were classified as healthy (no visible symptoms of infestation or below ET), while those with densities above the ET were categorized as aphid-infested.

Table 1. Economic thresholds (ETs) for aphid infestation in winter wheat across growth stages T1, T2, and T3 [19–22].

Growth Stage	BBCH Stage	ET
T1 (stem elongation–heading)	30–55	4 aphids/tiller
T2 (flowering)	55–59	4–7 aphids/tiller
T3 (milky grain development)	69–75	8–12 aphids/tiller

2.2.2. Spectral Data Collection

In this study, spectral measurements were conducted at the canopy level of winter wheat stands using a pistol grip fiber holder attached to the Spectral Evolution® SR-2500 portable spectroradiometer (Haverhill, MA, USA). The device measures spectral reflectance in a range from 350 to 2500 nm, covering the visible (VIS), near-infrared (NIR), and short-wave infrared (SWIR) regions. It captures data in 2151 spectral bands and automatically resamples the spectra with a resolution of 1 nm for export. The spectral resolution varies between 5 nm in the 350–1000 nm range and 22 nm in the 1000–2500 nm range, allowing precise data acquisition across the entire spectrum. Measurements were performed by positioning the pistol grip fiber holder 50 cm above the surface of the canopy, with the device's 25° field of view covering an area of approximately 0.2 m^2 (based on the field of view geometry). All measurements were conducted under sunny, cloud-free conditions without wind, ensuring consistent lighting and minimizing environmental variability. Measurements were consistently taken between 10:00 and 14:00 to capture optimal sunlight angles and avoid shading effects [58–61]. This approach enabled accurate and consistent sampling of canopy reflectance. Calibration of the spectroradiometer was performed every 10 scans using a white $BaSO_4$ reference panel (99% Spectralon; SphereOptics GmbH, Herrsching, Germany) to ensure the accuracy of the spectral measurements. To ensure consistency, the fixture was held vertically above the canopy and perpendicular to the measured surface for all scans. Spectral measurements were taken at phenological stages BBCH 30–75, corresponding to growth stages T1, T2, and T3. The device was connected to a portable computer, and all data were stored using Darwin SP V1.5 software (Spectral Evolution, Haverhill, MA, USA), which allowed efficient data management. This setup minimized environmental disturbances, ensuring reliable and consistent data on canopy reflectance. In total, 431 spectral reflectance samples of wheat canopies were collected. The full raw spectral reflectance dataset used in this study is available as Supplementary Materials and can be accessed at [62].

2.3. Data Analysis and Processing

A Savitzky–Golay filter with a window size of 11 and a polynomial order of 2 was applied to smooth the canopy-level spectral reflectance curves, which are inherently noisy due to environmental factors. This method preserved key spectral features while reducing noise. Additionally, specific spectral regions affected by atmospheric noise or lacking informative content were removed. These included the ranges of 1350–1450 nm, 1800–1950 nm, and 2400–2500 nm, which correspond to atmospheric interference, and the range of

350–400 nm, which is considered non-informative. The resulting preprocessed data ensured that only meaningful spectral regions were retained for further analysis. For statistical analysis, Point Biserial Correlation (r_{pb}) coefficient was used as a special case of the Pearson Correlation, which is applied for measuring the relationship between a continuous variable, i.e., spectral reflectances and VIs, and a dichotomous variable, i.e., a variable with two possible values, healthy or aphid-infested. Point Biserial Correlation is calculated as

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_0}{s_x} \cdot \sqrt{\frac{n_1 \cdot n_0}{n^2}},$$

where $\bar{x}_1$ is the mean value of all x for which $y = 1$, $\bar{x}_0$ is the mean value of all x for which $y = 0$, s_x is the standard deviation, n_1 is the number of x_1, n_0 the number of x_0, and n the sample size.

2.3.1. Uniform Manifold Approximation and Projection (UMAP) Analysis

To analyze the clustering patterns of canopy spectral reflectance data, UMAP was applied as a dimensionality reduction technique [39,63]. UMAP was used to project high-dimensional spectral data (preprocessed using the Savitzky–Golay filter and cleaned of non-informative data and atmospheric noise) into a lower-dimensional space while preserving the local and global structure of the dataset. For the transformation, the UMAP parameter n neighbors was set to 20 so that the broader structure of the data was covered well, while the minimum distance parameter was set to 0.1, making the UMAP-transformed clusters denser and easier to visually analyze. For validation of the UMAP transformation, a silhouette analysis was performed [64], where the obtained silhouette score provides a numerical measure of how well are the transformed clusters of data separated, i.e., how well the UMAP procedure separated the high-dimensional spectral data into the lower-dimensional space.

2.3.2. Vegetation Indices (VI)

The selection of vegetation indices (VI) was based on a comprehensive review of the latest scientific literature [65–72]. These indices were selected primarily to monitor plant responses to aphid-induced stress in winter wheat, focusing on chlorophyll content, structural integrity and photosynthetic efficiency, earlier senescence, etc., which are known to change under biotic stress caused by direct feeding by sap-sucking insects such as aphids. Priority was given to indices that are highly sensitive to changes in canopy reflectance associated with stress. These include indices related to chlorophyll content and photosynthetic activity (NDVI, GNDVI, CCI, MTVI, and MTVI-2), which can indicate stress-induced chlorophyll degradation. Additionally, stress-specific indices such as the Aphid Index (AI), Aphid Stress Index (ASI-1), and Damage Sensitive Spectral Index-2 (DSSI-2) were included to assess plant stress responses more directly. Indices designed to detect leaf structural changes and pigment changes (SIPI, PSRI, and Chl SI-3) were also incorporated, as aphid feeding can alter the plant surface properties and leaf composition. Furthermore, indices related to water content and senescence detection (NDWI, SAVI, OSAVI, and TVI) were selected to capture broader plant responses to aphid infestations, including changes in water regulation and premature senescence caused by prolonged feeding. In total, 16 VI were selected (listed in Table 2) to analyze the hyperspectral data, each defined using reflectance (R) at specific wavelengths.

Table 2. List of vegetation indices (VI) considered for analysis, their formula, and source.

VI	Formula	Source
Normalized Difference Vegetation Index (NDVI)	$\mathrm{NDVI} = \frac{R_{800} - R_{670}}{R_{800} + R_{670}}$	Tucker (1979) [73]
Red Edge Normalized Difference Vegetation Index (NDVI750)	$\mathrm{NDVI}_{750} = \frac{R_{750} - R_{705}}{R_{750} + R_{705}}$	Gitelson and Merzlyak (1998) [74]
Green Normalized Difference Vegetation Index (GNDVI)	$\mathrm{GNDVI} = \frac{R_{800} - R_{550}}{R_{800} + R_{550}}$	de Souza et al. (2010) [75]
Normalized Difference Water Index (NDWI)	$\mathrm{NDWI} = \frac{R_{857} - R_{1241}}{R_{857} + R_{1241}}$	Gao (1996) [76]
Aphid Index (AI)	$\mathrm{AI} = \frac{R_{761} - R_{908}}{R_{712} - R_{719}}$	Mirik et al. (2006) [46]
Aphid Stress Index 1 (ASI-1)	$\mathrm{ASI\text{-}1} = \frac{R_{666} - R_{1462}}{R_{666} + R_{1462}}$	Chen et al., (2018) [68]
Damage Sensitive Spectral Index-2 (DSSI-2)	$\mathrm{DSSI\text{-}2} = \frac{R_{747} - R_{901} - R_{537} - R_{572}}{(R_{747} - R_{901}) + (R_{537} - R_{572})}$	Mirik et al. (2006) [47]
Chlorophyll Stress Index 3 (Chl SI-3)	$\mathrm{Chl\ SI\text{-}3} = \frac{R_{551}}{R_{915}}$	Zhao et al. (2005) [77]
Structure Insensitive Pigment Index (SIPI)	$\mathrm{SIPI} = \frac{R_{800} - R_{445}}{R_{800} - R_{680}}$	Kureel et al. (2022) [78]
Modified Triangular Vegetation Index (MTVI)	$\mathrm{MTVI} = 1.2 * [1.2 * (R_{800} - R_{550}) - 2.5 * (R_{670} - R_{550})]$	Haboudane et al. (2004) [79]
Modified Triangular Vegetation Index 2 (MTVI-2)	$\mathrm{MTVI\text{-}2} = \frac{1.5 * (1.2 * (R_{800} - R_{550}) - 2.5 * (R_{670} - R_{550})}{\sqrt{(2R800 + 1)^2 - (6\mathrm{R}800 - 5 * \sqrt{\mathrm{R}670}) - 0.5)}}$	Haboudane et al. (2004) [79]
Chlorophyll/Carotenoid Index (CCI)	$\mathrm{CCI} = \frac{R_{532} - R_{630}}{R_{532} + R_{630}}$	Springer et al. (2017) [80]
Pigment Senescence Reflectance Index (PSRI)	$\mathrm{PSRI} = \frac{R_{665} - R_{490}}{R_{740}}$	Punalekar et al. (2021) [81]
Soil-Adjusted Vegetation Index (SAVI)	$\mathrm{SAVI} = \frac{(R_{842} - R_{665}) * (1 + L)}{R_{842} + R_{665} + L}$ L=0,3	Huete (1988) [82]
Optimized Soil-Adjusted Vegetation Index (OSAVI)	$\mathrm{OSAVI} = \frac{(1 + 0.16) * (R_{800} - R_{670})}{R_{800} + R_{670} + 0.16}$	Haboudane et al. (2002) [83]
Triangular Vegetation Index (TVI)	$\mathrm{TVI} = 0.5 * [120 * (R_{750} - R_{550}) - 200 * (R_{670} - R_{550})]$	Mulla (2013) [42]

2.3.3. Machine Learning Analysis

The classification was performed by implementing a selection of machine learning models, with analyses conducted individually for each growth stage individually (T1, T2, and T3) and for all the growth stages unified in a single dataset (T123). Additionally, the classifications were carried out separately for VI, full spectral reflectance data, and finally on the UMAP transformed values. A supervised learning approach was applied, with predefined classes distinguishing between healthy plants and those under stress caused by aphid infestation. To ensure robust model evaluation and performance validation, a nested cross-validation (CV) procedure was conducted, with both the outer and inner loop performed in 5 separated stratified folds, with the data split into 80% for training and 20% for testing. The following selection of ML models was tuned and evaluated in this paper: Logistic Regression (LR), K-Nearest Neighbors (KNN), Support vector machine (SVM), Random Forest (RF), and Light Gradient-Boosting Machine (LGBM). The selected list of models was derived such that several aspects of ML based classification were covered well, including model and model tuning complexity (from simple linear and decision tree models to complex ensemble methods), variety of model structures (i.e linear, decision tree, and support vector machine), as well as model novelty (from baseline models such as SVM and RF to state-of-the-art LGBM). For the tuning of respective model hyperparameters (Appendix A), a Tree-structured Parzen Estimator implemented within the Optuna Python library was utilized [84]. Each tuning was performed over 200 trials, with an additional execution time threshold set at 2 h. A complete list of individual model hyperparameters is presented in Appendix A.1, Table A1, while other relevant implementation aspects are presented in Appendix A.2.

The models were evaulated using accuracy and F1 score classification metrics, calculated as follows:

$$\text{Overall accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

where TP denotes True Positive classifications, TN denotes True Negative classifications, FP denotes False Positive classifications, and FN denotes False Negative classifications.

3. Results

3.1. Response of Winter Wheat Canopy Spectra to Aphid Infestation

The spectral reflectance curves for winter wheat canopies—healthy (below ET or without visible symptoms) and aphid-infested (above ET)—are presented in Figure 3. These curves reveal distinct spectral patterns across the VIS, NIR, and SWIR regions, reflecting the physiological and structural changes induced by aphid infestations exceeding the ET. Across all growth stages, healthy plants consistently exhibited higher reflectance in the NIR region compared to infested plants, though the magnitude of spectral differences varied. At T1, reflectance differences were most pronounced in the red-edge (680–750 nm) and NIR regions, where healthy plants exhibited higher reflectance than infested plants. At T2, the red-edge differences became slightly less distinct, while NIR differences remained prominent. Additionally, infested plants displayed higher SWIR reflectance compared to healthy plants, diverging from the trends observed at other stages. At T3, the most significant differences were observed in the VIS and NIR regions. In the VIS region, especially from 550 (green peak) and 630 (red peak) to 680 nm, infested plants exhibited higher reflectance than healthy plants, leading up to the red-edge transition. In contrast, NIR reflectance was markedly lower in infested plants, representing the most pronounced spectral divergence at this stage (Figure 3).

The spectral reflectance curves for healthy winter wheat canopies across three growth stages exhibit progressive changes that reflect physiological and structural development over time. In the VIS region, reflectance increased progressively from T1 to T3, with the highest values observed at T3, particularly around 550 nm. The red-edge slope was most pronounced at T1 and T2, while at T3, the slope became less steep. In the NIR region, reflectance was highest at T2, followed by T3, with the lowest values recorded at T1. The largest difference was observed between T1 and T2, where reflectance increased sharply, while at T3, a slight decrease was noted compared to T2 (Figure 3).

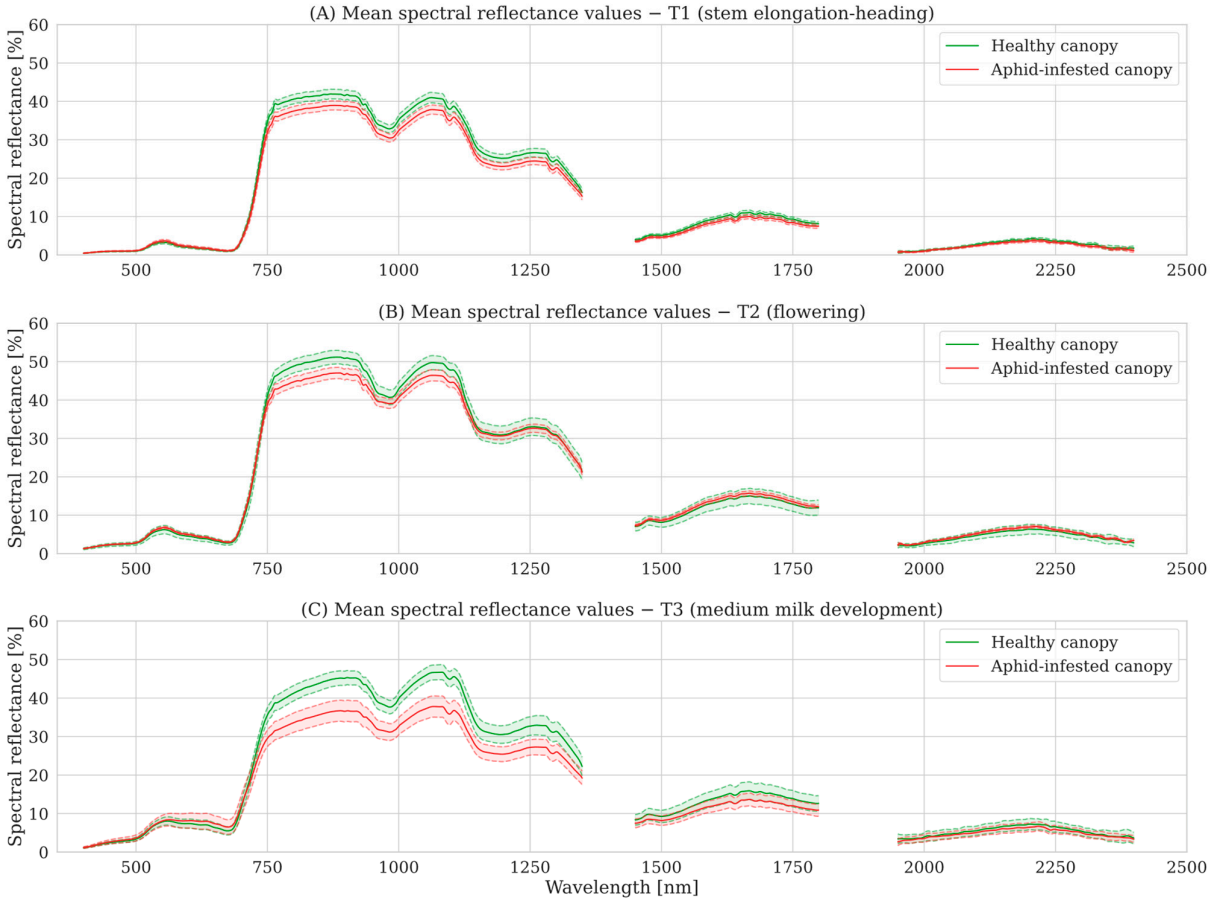

Figure 3. Mean and standard deviation of canopy spectral reflectance curves for healthy (green) and aphid-infested (red) winter wheat canopies across three growth stages: (**A**) T1—stem elongation–heading (BBCH 30–55), (**B**) T2—flowering (BBCH 55–59), and (**C**) T3—milky grain development (BBCH 69–75). Each mean spectral reflectance curve is based on 145 samples for T1, 141 samples for T2, and 145 samples for T3. Spectral regions affected by atmospheric noise (1350–1450 nm, 1800–1950 nm, and 2400–2500 nm) and non-informative content (350–400 nm) were excluded.

3.2. UMAP Visualization of Winter Wheat Canopy Spectra

The UMAP visualization (Figure 4) shows clear clustering of healthy (green) and aphid-infested (red) winter wheat canopies across all three measurement stages (T1, T2, and T3). In T1, the clusters for healthy and aphid-infested canopies are relatively distinct but show some overlap, primarily due to a few outliers. This resulted in a silhouette score of 0.76, which is still considered relevant despite being the lowest among all three growth stages. In T2, the separation between the two clusters remains clear, with only minor convergence visible, and a very confident silhouette score of 0.96. In T3, the distribution of samples is similar to that in T2, but with a few more incorrectly classified samples. Consequently, the silhouette score of 0.88, although marginally lower than that observed in T2, remains within a range indicative of high separation confidence. The presented UMAP transformation patterns indicate that differences in canopy spectral responses between healthy and aphid-infested wheat are most pronounced at the T2 growth stage. In T1, these differences remain only subtle, with less distinct reflectance profiles, resulting in less robust transformation of the collected data. In T3, aphid-induced stress signals merge more with natural senescence effects, leading to greater overlap (Figure 4).

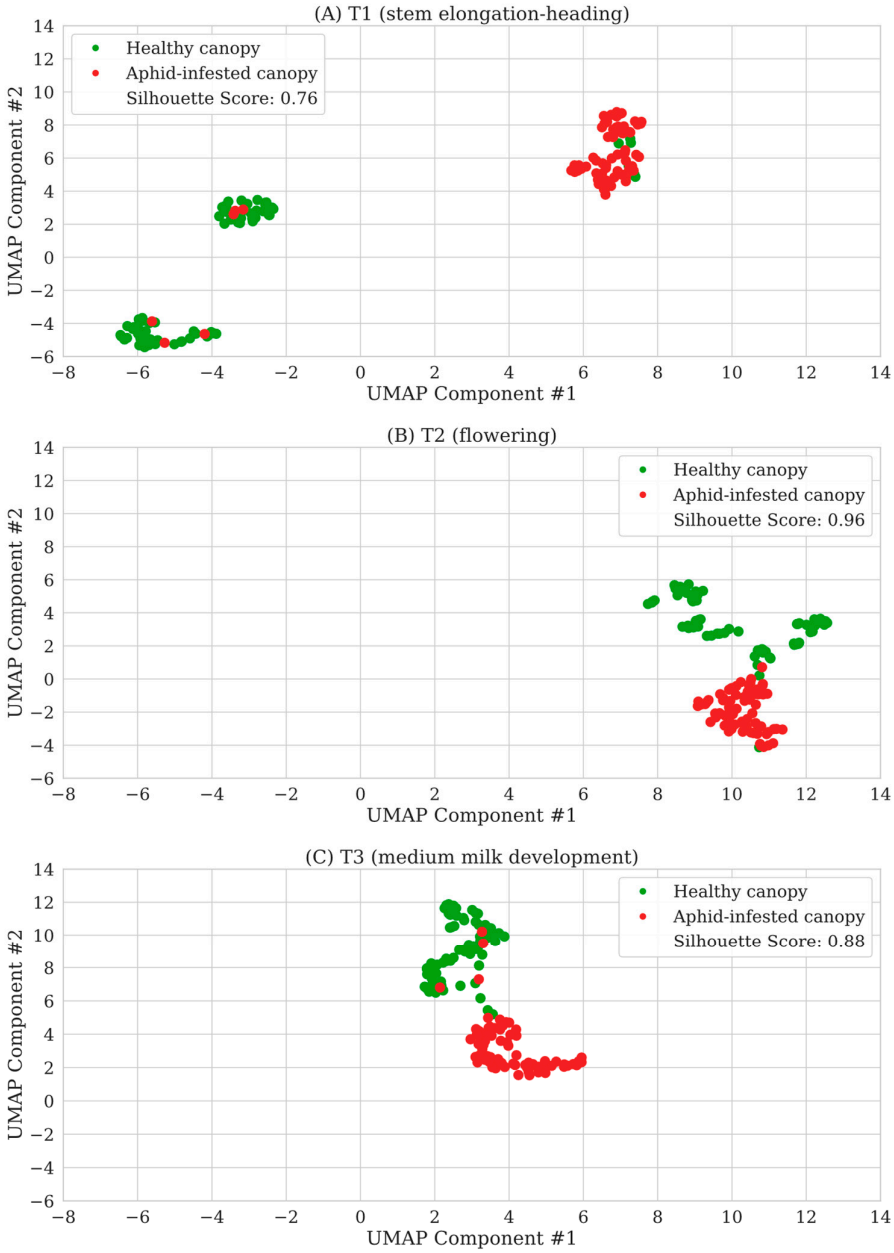

Figure 4. UMAP visualization of spectral data across three growth stages: (**A**) T1—stem elongation–heading (BBCH 30–55), (**B**) T2—flowering (BBCH 55–59), and (**C**) T3—milky grain development (BBCH 69–75), showing clustering of healthy (green) and aphid-infested (red) winter wheat canopies. Each UMAP projection is based on 145 samples for T1, 141 samples for T2, and 145 samples for T3.

3.3. Vegetation Indices Analysis

The boxplot analysis, as depicted in Figure 5, shows that indices such as NDVI, NDVI750, and OSAVI consistently decreased in aphid-infested plants, highlighting their sensitivity to reductions in vegetation greenness and canopy density. Conversely, stress-sensitive indices like Chl SI-3, ASI-1, PSRI, SIPI, and DSSI-2 exhibited higher values in infested plants, indicating their effectiveness in detecting pigment degradation and physiological stress. The most pronounced differences between healthy and aphid-infested plants were observed at T2, where the separation between groups was most distinct. This trend suggests that VI are particularly effective in differentiating aphid stress during the

flowering stage. For ease of interpretation and visualization, all VI values were standardized (Figure 5).

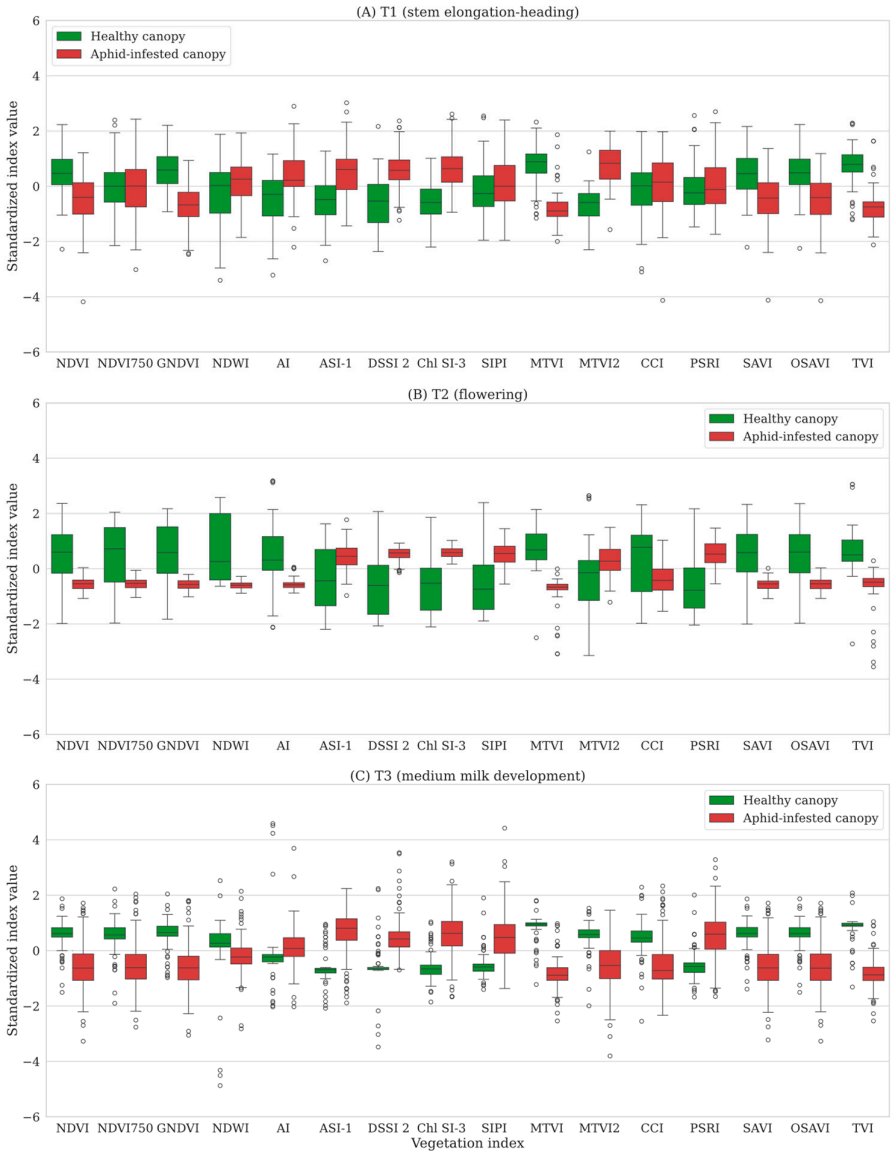

Figure 5. Boxplot distribution of vegetation indices across three growth stages: (**A**) T1—stem elongation–heading (BBCH 30–55), (**B**) T2—flowering (BBCH 55–59), and (**C**) T3—milky grain development (BBCH 69–75) for healthy (green) and aphid-infested (red) plants. Each boxplot is based on 145 samples for T1, 141 samples for T2, and 145 samples for T3.

In Figure 6, the Point Biserial Correlation coefficients (r_{pb}) for VI are shown across the growth stages T1, T2, and T3. Among all indices, MTVI (r_{pb} = −0.77, −0.79, −0.82) and TVI (r_{pb} = −0.75, −0.66, −0.66 consistently demonstrate the highest correlations with the canopy state across all stages, indicating their strong and stable performance in distinguishing between healthy and aphid-infested canopies.

Other indices that perform well and remain relatively stable across stages include GNDVI (r_{pb} = −0.66, −0.58, −0.57), Chl SI-3 (r_{pb} = 0.65, 0.59, 0.58), DSSI-2 (r_{pb} = 0.57, 0.53, 0.56), SAVI (r_{pb} = −0.47, −0.57, −0.56), OSAVI (r_{pb} = −0.48, −0.57, −0.56), and NDVI (r_{pb} = −0.47, −0.56, −0.55). These indices provide consistent differentiation between healthy and stressed

plants, with GNDVI and Chl SI-3 particularly standing out due to their ability to track chlorophyll content and canopy greenness effectively.

The correlation analysis (Figure 6) reveals that MTVI2 demonstrates a strong positive relationship with plant health during T1 (r_{pb} = 0.73), indicating its effectiveness in detecting early stage vegetation conditions. However, its correlation drops markedly during T2 (flowering, r_{pb} = 0.3) and becomes negatively correlated during T3 (milky grain development, r_{pb} = −0.52). This trend suggests that MTVI2's sensitivity to vegetation state is highly stage-dependent. A similar situation is observed with NDWI, which shows inconsistent behavior across growth stages, with a weak positive correlation during T1 (r_{pb} = 0.20) but negative correlations during T2 (r_{pb} = −0.60) and T3 (r_{pb} = −0.19). This variability likely reflects changes in canopy structure, chlorophyll saturation, and water dynamics as plants transition to reproductive stages, making both MTVI2 and NDWI less reliable for detecting aphid stress in mid-to-late growth stages (Figure 6).

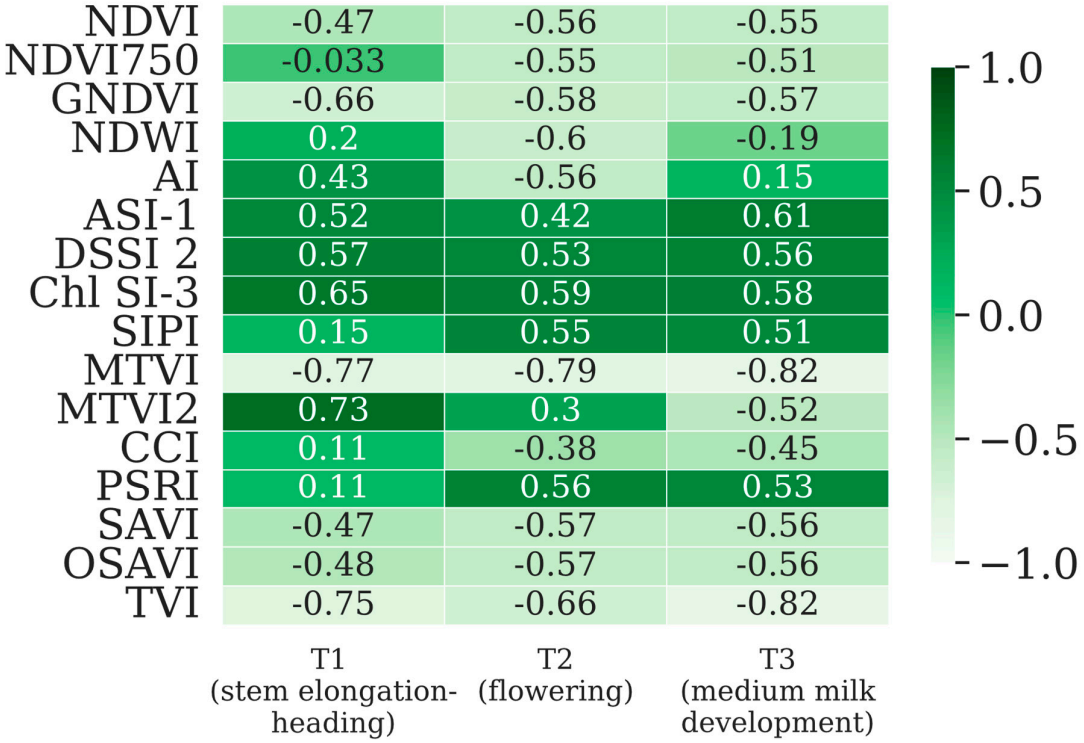

Figure 6. Point Biserial (r_{pb}) correlations for vegetation indices (VI) across three growth stages: T1—stem elongation–heading (BBCH 30–55), T2—flowering (BBCH 55–59), and T3—milky grain development (BBCH 69–75).

3.4. Machine Learning Classification

The results of machine learning-based classification of aphid-infested and healthy winter wheat canopies are presented in Tables 3–6. The classification performance is reported for three different input datasets: VI (Table 3), full spectral reflectance data (Table 4), and UMAP-transformed data (Table 5). Finally, in Table 6, the classification results are averaged across all ML models for the three corresponding input sets to provide an overall evaluation of model performance. Due to the use of a 5-fold nested cross-validation procedure, accuracy and F1-score metrics are presented as mean values with standard deviations, ensuring robust performance evaluation. Bolded values indicate the highest-performing model for each growth stage (T1, T2, T3, and T123).

The classification results demonstrate high overall performance, with accuracy ranging from 88% to 98% and F1-scores between 0.88 and 0.99. Among the tested ML models, SVM and RF consistently achieved the highest classification accuracy across different input sets and wheat growth stages. In Table 3 (VI-based classification), SVM exhibited the highest accuracy and F1-score in all growth stages, making it the most effective model for classification based on vegetation indices (93.10–98.57% accuracy and 0.93–0.99 F1 score). In Table 4 (full spectral reflectance data), RF provided the best classification results, with the highest accuracy and F1-score across most growth stages (89.66–98.57% accuracy and 0.89–0.99 F1 score). In Table 5 (UMAP-transformed data), RF and SVM showed the best performance, demonstrating strong classification accuracy across different growth stages (90.34–94.48% accuracy and 0.90–0.94 F1 score). Across the three individual growth stages, T2 exhibited the highest classification performance on average, indicating a clearer distinction between healthy and aphid-infested wheat canopies at this stage. In addition, models trained on the combined dataset (T123), which includes all growth stages, produced comparable results to models trained on individual growth stages.

Table 6 shows the overall classification performance averaged over all ML models for the three input datasets (VI, full-spectrum, and UMAP-transformed data). The results show that the highest overall accuracy and F1 score were obtained with the full spectrum dataset, especially at T2, where the highest classification performance was observed (97.86% accuracy and 0.98 F1 score). The VI dataset also performed well, showing consistent accuracy and F1-scores, with the best performance recorded at T2 (97.43% accuracy and 0.97 F1-score), making it a highly effective input type for classification. The UMAP-transformed data gave competitive results, especially at T3 (94.21% accuracy and 0.94 F1-score), showing that classification performance can be maintained despite dimensionality reduction. These results show that full spectral data provided the most robust classification performance, while the VI and UMAP-transformed data provided a reliable and computationally efficient alternative.

Table 3. Machine learning model' classification performance based on vegetation indices for aphid infestation detection in winter wheat across three phenological stages: T1 (stem elongation–heading), T2 (flowering), and T3 (milky grain development).

	LR		KNN		SVM		RF		LGBM	
	Acc.	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**
T1	**93.10 ± 5.45**	**0.93 ± 0.05**	92.41 ± 7.48	0.92 ± 0.08	**93.10 ± 5.45**	**0.93 ± 0.05**	90.34 ± 5.67	0.90 ± 0.06	91.72 ± 5.23	0.92 ± 0.05
T2	97.86 ± 1.96	0.98 ± 0.02	97.14 ± 2.99	0.97 ± 0.03	**98.57 ± 1.96**	**0.99 ± 0.02**	95.71 ± 2.99	0.96 ± 0.03	97.86 ± 3.19	0.98 ± 0.03
T3	92.41 ± 4.49	0.92 ± 0.04	91.72 ± 4.63	0.92 ± 0.05	**94.48 ± 3.08**	**0.94 ± 0.03**	**94.48 ± 3.08**	**0.94 ± 0.03**	88.28 ± 3.93	0.88 ± 0.04
T123	89.30 ± 2.52	0.89 ± 0.03	94.42 ± 0.94	0.94 ± 0.02	**95.35 ± 0.82**	**0.95 ± 0.01**	93.49 ± 1.95	0.94 ± 0.02	91.86 ± 3.94	0.92 ± 0.04

Bold values denote highest-performing models for each growth stage (T1–T123).

Table 4. Machine learning models' classification performance based on full spectral reflectance dataset across three phenological stages: T1 (stem elongation–heading), T2 (flowering), and T3 (milky grain development).

	LR		KNN		SVM		RF		LGBM	
	Acc.	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**
T1	90.34 ± 1.54	0.90 ± 0.02	90.34 ± 2.89	0.90 ± 0.03	89.66 ± 2.44	0.89 ± 0.02	89.66 ± 2.44	0.89 ± 0.02	**91.03 ± 3.08**	**0.91 ± 0.03**
T2	97.86 ± 1.96	0.98 ± 0.02	97.14 ± 2.99	0.97 ± 0.03	97.86 ± 1.96	0.98 ± 0.02	**98.57 ± 1.96**	**0.99 ± 0.02**	97.86 ± 1.96	0.98 ± 0.02
T3	94.48 ± 3.08	0.94 ± 0.03	93.79 ± 1.54	0.94 ± 0.02	90.34 ± 5.11	0.90 ± 0.05	**95.17 ± 3.08**	**0.95 ± 0.03**	93.79 ± 2.89	0.94 ± 0.03
T123	93.95 ± 2.78	0.94 ± 0.03	93.72 ± 1.94	0.94 ± 0.02	**95.35 ± 2.73**	**0.95 ± 0.03**	95.11 ± 0.97	0.95 ± 0.01	94.88 ± 1.33	0.95 ± 0.01

Bold values denote highest-performing models for each growth stage (T1–T123).

Table 5. Machine learning models' classification performance for aphid infestation detection using UMAP-transformed data across three phenological stages: T1 (stem elongation–heading), T2 (flowering), and T3 (milky grain development).

	LR		KNN		SVM		RF		LGBM	
	Acc.	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**
T1	**90.34 ± 3.78**	**0.90 ± 0.04**	89.66 ± 2.44	0.89 ± 0.02	**90.34 ± 3.78**	**0.90 ± 0.04**	**90.34 ± 3.78**	**0.90 ± 0.04**	**90.34 ± 3.78**	**0.90 ± 0.04**
T2	96.43 ± 3.57	0.96 ± 0.04	97.14 ± 3.91	0.97 ± 0.04	97.14 ± 3.91	0.97 ± 0.04	97.14 ± 3.91	0.97 ± 0.04	**97.86 ± 3.19**	**0.98 ± 0.03**
T3	93.79 ± 2.89	0.94 ± 0.03	**94.48 ± 1.89**	**0.94 ± 0.02**	**94.48 ± 1.89**	**0.94 ± 0.02**	**94.48 ± 1.89**	**0.94 ± 0.02**	93.79 ± 2.89	0.94 ± 0.03
T123	85.81 ± 2.89	0.86 ± 0.03	93.72 ± 3.25	0.94 ± 0.03	**94.19 ± 3.29**	**0.94 ± 0.03**	93.95 ± 3.53	0.94 ± 0.04	93.49 ± 2.92	0.93 ± 0.03

Bold values denote highest-performing models for each growth stage (T1–T123).

Table 6. Overall classification performance averaged across machine learning models using VI, full-spectrum, and UMAP-transformed data across three phenological stages: T1 (stem elongation–heading), T2 (flowering), and T3 (milky grain development).

	VI		Wwl		UMAP	
	Acc.	**F1**	**Acc.**	**F1**	**Acc.**	**F1**
T1	**92.14 ± 5.86**	**0.92 ± 0.06**	90.21 ± 2.48	0.90 ± 0.03	90.21 ± 3.51	0.90 ± 0.04
T2	97.43 ± 2.62	0.97 ± 0.03	**97.86 ± 2.16**	**0.98 ± 0.02**	97.14 ± 3.70	0.97 ± 0.04
T3	92.28 ± 3.84	0.92 ± 0.04	93.52 ± 3.14	0.93 ± 0.03	**94.21 ± 2.29**	**0.94 ± 0.02**
T123	92.88 ± 2.10	0.93 ± 0.02	**94.61 ± 1.95**	**0.95 ± 0.02**	92.23 ± 3.18	0.92 ± 0.03

Bold values denote highest-performing models for each growth stage (T1–T123).

4. Discussion

The spectral reflectance differences observed in this study reveal key physiological and structural changes in wheat canopies under aphid stress above ET (Figure 3). In the VIS region, aphid-infested plants exhibited slightly higher reflectance, particularly at T3, with the most pronounced separation from healthy plants around the green peak and red region (550–680 nm). This pattern aligns with previous findings that aphid feeding leads to chlorophyll degradation, increasing reflectance in the green and yellow spectral bands 7 [26]. As chlorophyll is crucial for photosynthesis [57,85], its decline at T3 coincides with peak aphid populations (8–12 aphids per tiller), particularly *S. avenae*, which predominantly feeds on the wheat ear, exacerbating stress and reducing yield [86,87]. The red-edge region (680–750 nm) showed consistent spectral differences between healthy and infested canopies across all stages, reinforcing its sensitivity to aphid-induced stress [88]. This finding aligns with previous research demonstrating the utility of red-edge indices for detecting soybean aphid stress on soybean canopies [89–91]. In the NIR region, aphid-infested canopies exhibited the greatest reduction in reflectance at T3, likely due to mesophyll damage, which disrupts the internal leaf structure, reducing scattering efficiency and increasing transmittance [92]. In contrast, healthy plants maintained a more intact mesophyll structure, efficiently scattering NIR radiation within green tissues such as leaves, flag leaves, glumes, and awns [93,94]. The SWIR region, strongly influenced by the leaf water content [95], exhibited variable responses across growth stages. At T2, infested canopies had higher SWIR reflectance, suggesting significant water stress, similar to findings in brown planthopper-infested rice [96] and aphid-infested wheat leaf [26]. However, at T1 and T3, healthy plants exhibited higher SWIR reflectance, possibly due to differences in canopy moisture regulation, cuticular changes, or leaf structural properties [97]. The pronounced differences in SWIR at T2 highlight the impact of aphid-induced water stress during flowering, a critical phase for grain formation (Figure 3). Increased SWIR reflectance in infested plants indicates water loss and structural disruption, which can impair pollination and grain development. As this stage is highly sensitive to water availability,

with drought known to reduce fertility [98], these findings underscore the importance of early detection and intervention to mitigate yield losses.

This finding on the SWIR region is further supported by VI analysis (Figure 5), where NDWI exhibited the strongest negative correlation (r_{pb} = −0.60) at T2, indicating reduced water content in aphid-infested wheat canopies [76]. In addition to NDWI, several other VIs demonstrated strong potential for detecting aphid-induced stress, with MTVI and TVI consistently exhibiting the highest correlations across all growth stages (Figure 5). Their strong negative correlations (r_{pb} = −0.77 to −0.82 for MTVI, r_{pb} = −0.66 to −0.82 for TVI) confirm that aphid feeding reduces vegetation greenness and canopy integrity, making these indices valuable for early stress detection. As an improved version of TVI, MTVI enhances sensitivity to structural canopy changes while minimizing pigment-related variations by incorporating reflectance at 800 nm instead of 750 nm [79]. Given its stability across growth stages, MTVI is particularly suitable for early stress detection and aphid hotspots mapping in site-specific pest management. TVI also performed well due to its sensitivity to chlorophyll content and canopy structure, as it is derived from green (550 nm), red (670 nm), and NIR (750 nm) reflectance [99]. However, its reliance on pigment-driven variations may limit its robustness compared to MTVI. MTVI2 demonstrated high sensitivity to early vegetation conditions (r_{pb} = 0.73 at T1) but showed declining performance in later stages (r_{pb} = 0.3 at T2, rpb = −0.52 at T3), indicating growth stage-dependent behavior. While its soil adjustment factor (L = 0.5) improves early stage greenness detection [99], its instability at T2 and T3 reduces its applicability for consistent aphid stress monitoring. Among stress-sensitive indices, DSSI-2 and Chl SI-3 showed strong correlations with aphid infestation, reinforcing their value for early stress detection. DSSI-2 remained positively correlated across all growth stages (r_{pb} = 0.53–0.57), effectively distinguishing between healthy and infested canopies. As an index designed to enhance contrast between vegetation and soil, DSSI-2 is particularly sensitive to pigment degradation and subtle canopy changes associated with aphid feeding [47]. Similarly, Chl SI-3, which tracks chlorophyll content, showed stable positive correlations (r_{pb} = 0.65 at T1, 0.58 at T2 and T3), confirming its reliability in detecting aphid-induced pigment breakdown [77]. GNDVI, a greenness-related index that incorporates the green spectral band instead of red, exhibited strong negative correlations across all growth stages (r_{pb} = −0.66 at T1, −0.58 at T2, −0.57 at T3), aligning with reductions in chlorophyll concentration and canopy greenness caused by aphid stress. Its effectiveness has also been demonstrated in cotton aphid detection [100].

The classification results demonstrated consistently high accuracy, with overall performance ranging from 88% to 98% and F1-scores between 0.88 and 0.99. Among the tested ML models, SVM and RF achieved the highest classification accuracy across different input datasets and wheat growth stages (Tables 3–6), reinforcing their robustness for aphid stress detection in field conditions. SVM's strong performance can be attributed to its ability to handle small and complex datasets effectively, leveraging its mathematical foundations to optimize classification boundaries [101]. This aligns with findings by Mountrakis et al. [37], who highlighted SVM's effectiveness in remote sensing applications. The highest classification accuracy was observed at T2 (flowering), confirming this stage as optimal for aphid stress detection. Since flowering is critical for grain formation, timely intervention is essential to prevent yield losses [102]. Among the input datasets, full spectral reflectance yielded the best classification performance, as expected, since dimensionality reduction through VI calculation or UMAP transformation inevitably resulted in some information loss. However, VI-based classification (Table 3) maintained high accuracy, indicating that key stress-related spectral features were effectively captured, supporting previous findings that highlight VI as a computationally efficient and

scalable alternative for remote sensing-based stress detection [103]. Reducing dataset complexity through VI selection or UMAP transformation minimizes data dimensionality and significantly reduces processing time, making ML classification more efficient and scalable for large-scale applications [104]. The selection of appropriate VI plays a crucial role in classification accuracy. While this study used all selected VI (Table 2) for model training, optimizing VI selection per growth stage (T1, T2, and T3) could improve accuracy by focusing only on the most relevant indices. This targeted approach would enhance model efficiency while maintaining high reliability, reinforcing the need for tailored feature selection in RS-based stress detection in precision agriculture.

These findings confirm the hypotheses that spectral reflectance differences and VI correlations can accurately classify aphid-infested wheat canopies above ET levels. Furthermore, ML models demonstrated high classification accuracy, validating the hypothesis that hyperspectral data combined with ML techniques provide a robust approach for early aphid stress detection across multiple growth stages. Building on these validated hypotheses, this study also acknowledges certain limitations. The field-based hyperspectral measurements were manually operated, requiring on-site data collection, which limited the temporal resolution. Additionally, all measurements were conducted under clear, sunny conditions with no wind, potentially restricting applicability under variable weather conditions. Despite these limitations, this study demonstrates significant advancements. By integrating hyperspectral data with ML models, it presents a scalable approach for aphid stress detection, supporting real-time pest monitoring and decision making. The economic threshold-based classification approach could simplify intervention decisions, enhancing site-specific pest management. Additionally, conducting the research under realistic field conditions and at the canopy level increases the applicability of the findings to practical agricultural settings. A particularly noteworthy advancement is the high classification accuracy achieved with the generalized dataset, which combines data from all three growth stages. This demonstrates the model's robustness, indicating that the developed models can accurately classify aphid infestations without requiring precise knowledge of the exact growth stage. This generalization capability enhances the practical applicability of the method, making it suitable for field-scale deployment where phenological stages may vary spatially and temporally. Furthermore, this study evaluates ML classification accuracy using different sets of inputs, including full-spectrum measurements, the VI, and UMAP-transformed data. This approach brings additional value in terms of implementation analysis, as in the case of drone or satellite measurements, only specific VI might be available for analysis. This adaptability makes the proposed method highly versatile for various remote sensing platforms, ensuring scalability and practical applicability.

The findings from this study have significant implications for aphid management in cereals, particularly in advancing precision agriculture practices. Drones or satellites could be used to map large areas and generate regional and local risk assessments, prompting targeted scouting efforts [30]. Vegetation indices, already widely implemented in remote sensing-based monitoring systems, could facilitate early aphid detection, ensuring timely treatment decisions when infestations exceed economic threshold levels. However, further research is needed to distinguish aphid-induced stress from other biotic and abiotic stressors to prevent misclassification and optimize remote sensing-based decision support systems for precision pest management.

5. Conclusions

This study highlights the potential of hyperspectral remote sensing (RS) and machine learning (ML) for the early detection and classification of aphid infestations in winter wheat, enabling more precise and targeted pest management. Aphid-induced stress resulted in distinct spectral reflectance changes, particularly in the VIS and NIR regions, with the most pronounced differences observed at T2, aligning with peak stress impact during the flowering stage. Vegetation indices, particularly MTVI (r_{pb} = −0.77 to −0.82) and TVI (r_{pb} = −0.75 to −0.66), exhibited strong and stable correlations with aphid stress, reinforcing their reliability for mapping infestation hotspots and supporting site-specific pest control strategies. Machine learning models demonstrated high classification performance (F1-scores of 0.89–0.99), with SVM and RF consistently outperforming other models. The full spectral dataset yielded the highest classification accuracy, while VI-based classification provided a computationally efficient alternative with minimal performance reduction. UMAP visualization effectively clustered healthy and infested plants, particularly at T2, supporting the feasibility of dimensionality reduction for large-scale aphid detection. These findings confirm that the hypotheses proposed in this study are supported:

- Canopy-level spectral reflectance effectively distinguishes between aphid-infested wheat canopies above the economic threshold (ET) and healthy canopies across different growth stages.
- MTVI and TVI are the most reliable vegetation indices for detecting aphid-induced stress in winter wheat.
- ML classification models, particularly SVM and RF, accurately distinguish between healthy and aphid-infested wheat canopies, enabling data-driven treatment decisions.

The integration of hyperspectral imaging with ML offers a scalable and automated approach for real-time aphid stress detection and mapping, enhancing precision agriculture practices. Drones or satellite-based monitoring could provide early warnings by identifying areas where aphid populations exceed the ET, prompting timely scouting and intervention. These findings emphasize the importance of combining RS with ML to improve field-scale pest monitoring, reduce unnecessary pesticide applications, and support more sustainable crop protection strategies.

Supplementary Materials: The following supporting information can be downloaded at: https://doi.org/10.5281/zenodo.14968259.

Author Contributions: Conceptualization, S.S., H.N. and D.L.; methodology, S.S., D.L. and M.Z.; software, H.N.; validation, D.L., M.Z., I.P.Ž., M.M. and V.L.; formal analysis, S.S. and H.N.; investigation, S.S.; resources, V.L., M.M. and M.Z.; data curation, D.L.; writing—original draft preparation, S.S.; writing—review and editing, S.S., H.N., D.L. and I.P.Ž.; visualization, H.N.; supervision, D.L.; project administration, M.Z., V.L. and S.S.; funding acquisition, V.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Regional Development Fund through the project: Advanced and predictive agriculture for resilience to climate change (AgroSPARC) (KK.05.1.1.02.0031).

Data Availability Statement: The full spectral reflectance dataset used in this study is available as Supplementary Materials and can be accessed at https://doi.org/10.5281/zenodo.14968258 (accessed on 4 March 2025).

Acknowledgments: The authors gratefully acknowledge the support of the European Union through the European Regional Development Fund within the Operational Programme Competi-

tiveness and Cohesion (OPCC) 2014–2020 for funding the project "Advanced and Predictive Agriculture for Resilience to Climate Change (AgroSPARC)" (KK.05.1.1.02.0031). We also extend our appreciation to the Environmental Protection and Energy Efficiency Fund of the Republic of Croatia for their support. Furthermore, we sincerely thank the Bc Institute for Breeding and Production of Field Crops for their invaluable contribution to this research.

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

Appendix A

Appendix A.1. ML Model Hyperparameters

Table A1. ML models' hyperparameter search space.

MODEL	HYPERPARAMETER	RANGE
Logistic regression	C	[1 × 10^{-3}, 50], uniform distribution in the log domain
	Tolerance	[1 × 10^{-6}, 1 × 10^{-3}], uniform distribution
	Penalty	[l1, l2, elasticnet]
K-nearest Neighbors	No. of neighbors	[1, 50], integer
	Weights	[uniform, distance]
	Metric	[Euclidean, Manhattan, Minkowski]
Support vector machine	Kernel	[linear, rbf, poly, sigmoid]
	C	[0.1, 50], float
	Gamma	[0.0001, 0.001, 0.01, 0.1, 1, scale, auto]
	Degree	[2, 3, 4, 5], if kernel = poly
Random Forest	No. of estimators	[1, 500], integer
	Max features	[auto, sqrt]
	Max depth	[10, 110], integer, step = 10
	Min samples split	[2, 10], integer, step = 2
	Min samples leaf	[1, 4], integer
Light gradient-boosting machine	Objective	Multiclass
	Boosting type	GBDT
	No. of leaves	[2, 256], integer
	Learning rate	[1 × 10^{-4}, 0.1], uniform distribution in the log domain
	No. of estimators	[10, 1000], integer
	Reg alpha	[1 × 10^{-8}, 10], uniform distribution in the log domain
	Reg lambda	[1 × 10^{-8}, 10], uniform distribution in the log domain
	Subsample	[0.5, 1], float, step = 0.01
	Colsample bytree	[0.5, 1], float, step = 0.01
	Min child samples	[5m 50], integer, step = 5
	Min child weight	[1 × 10^{-3}, 10], float, step = 1 × 10^{-3}
	Metric	Multi logloss
	Num of classes	4
	Max depth	−1
	Subsample freq	1
	Early stopping rounds	20
	Objective	Multiclass

* gray-shaded fields denote fixed hyperparameters.

Appendix A.2. Implementation Aspects

The processing and analysis of data, including algorithms presented in Section 2.2, Data Analysis, and Section 2.3, Machine Learning Analysis, was performed in Python 3.9, with the utilization of the corresponding open-source libraries:

- Scikit-learn: data processing, LR, KNN, SVM, RF [https://scikit-learn.org/stable/] (accessed on 21 November 2024).
- UMAP [https://umap-learn.readthedocs.io/en/latest/] (accessed on 16 November 2024).
- LightGBM [https://lightgbm.readthedocs.io/en/latest/] (accessed on 22 November 2024).
- Optuna [https://optuna.org/] (accessed on 22 November 2024).

The overall ML models' training, and validation procedure lasted for 16 h and 38 min on a machine with 4 cores, 3.5 GHz clock, and 16 GB of RAM.

References

1. European Commission. *EU Cereal Farms Report Based on 2017 FADN Data*; European Union: Brussels, Belgium, 2019.
2. Foyer, C.H.; Rasool, B.; Davey, J.W.; Hancock, R.D. Cross-tolerance to biotic and abiotic stresses in plants: A focus on resistance to aphid infestation. *J. Exp. Bot.* **2016**, *67*, 2025–2037.
3. Keerthana, A.; Keerthana, M.; Shireesh Kumar, M.P.; Bahuguna, R.N.; Singh, S.K.; Rai, D.; Reddy, M.S.S. Characterizing and assessing the wheat-aphid complex under varying temperature and humidity. *Cereal Res. Commun.* **2024**, *52*, 1613–1627.
4. Skendžić, S.; Zovko, M.; Živković, I.P.; Lešić, V.; Lemić, D. The impact of climate change on agricultural insect pests. *Insects* **2021**, *12*, 440.
5. Ferry, N.; Stavroulakis, S.; Guan, W.; Davison, G.M.; Bell, H.A.; Weaver, R.J.; Down, R.E.; Gatehouse, J.A.; Gatehouse, A.M.R. Molecular interactions between wheat and cereal aphid (*Sitobion avenae*): Analysis of changes to the wheat proteome. *Proteomics* **2011**, *11*, 1985–2002. https://doi.org/10.1002/pmic.200900801.
6. Deb, M.; Anderson, J.M. Development of a multiplexed PCR detection method for Barley and Cereal yellow dwarf viruses, Wheat spindle streak virus, Wheat streak mosaic virus *and* Soil-borne wheat mosaic virus. *J. Virol. Methods* **2008**, *148*, 17–24.
7. Parry, H.R. Cereal aphid movement: General principles and simulation modelling. *Mov. Ecol.* **2013**, *1*, 14. https://doi.org/10.1186/2051-3933-1-14.
8. Wang, X.; Li, W.; Yan, J.; Wang, Y.; Zhang, X.; Tan, X.; Chen, J. Developmental, reproduction, and feeding preferences of *Sitobion avenae* mediated by soil silicon application. *Plants* **2023**, *12*, 989. https://doi.org/10.3390/plants12050989.
9. Maceljski, M. *Poljoprivredna entomologija*; Zrinski: Čakovec, Croatia, 2002.
10. Igrc, J. Važnost i potreba suzbijanja lisnih uši (*Aphididae*) strnih žita. *Agron. Glas.* **1985**, *47*, 109–118. https://hrcak.srce.hr/146124.
11. Tomanović, Ž.; Kavallieratos, N.G.; Ye, Z.; Nika, E.P.; Petrović, A.; Vollhardt, I.M.G.; Vorburger, C. Cereal aphid parasitoids in Europe (*Hymenoptera*: *Braconidae*: *Aphidiinae*): Taxonomy, biodiversity, and ecology. *Insects* **2022**, *13*, 1142. https://doi.org/10.3390/insects13121142.
12. Prado, E.; Tjallingii, W.F. Behavioral evidence for local reduction of aphid-induced resistance. *J. Insect Sci.* **2007**, *7*, 48.
13. Vickerman, G.P.; Wratten, S.D. The biology and pest status of cereal aphids (*Hemiptera*: *Aphididae*) in Europe: A review. *Bull. Entomol. Res.* **1979**, *69*, 1–32.
14. Zhang, Y.; Fan, J.; Fu, Y.; Francis, F.; Chen, J. Plant-mediated interactions between two cereal aphid species: Promotion of aphid performance and attraction of more parasitoids by infestation of wheat with phytotoxic aphid *Schizaphis graminum*. *J. Agric. Food Chem.* **2019**, *67*, 2763–2773. https://doi.org/10.1021/acs.jafc.8b06150.
15. Pegadaraju, V.; Knepper, C.; Reese, J.; Shah, J. Premature leaf senescence modulated by the *Arabidopsis PHYTOALEXIN DEFICIENT4* gene is associated with defense against the phloem-feeding green peach aphid. *Plant Physiol.* **2005**, *139*, 1927–1934.
16. Hafeez, F.; Abbas, M.; Zia, K.; Ali, S.; Farooq, M.; Arshad, M.; Ansari, M.J. Prevalence and management of aphids (*Hemiptera*: *Aphididae*) in different wheat genotypes and their impact on yield and related traits. *PLoS ONE* **2021**, *16*, e0257952.
17. Niehoff, B.; Stäblein, J. Vergleichende Untersuchungen zum Schadpotential der Getreideblattlausarten *Metopolophium dirhodum* (Wlk.) und *Sitobion avenae* (F.) in Winterweizen. *J. Appl. Entomol.* **1998**, *122*, 223–229.
18. Honek, A.; Martinkova, Z.; Saska, P.; Dixon, A.F. Aphids (*Homoptera*: *Aphididae*) on winter wheat: Predicting maximum abundance of *Metopolophium dirhodum*. *J. Econ. Entomol.* **2018**, *111*, 1751–1759.

19. Knodel, J.J.; Beauzay, P.; Boetel, M.; Prochaska, T.; Lubenow, L. *North Dakota Field Crop Insect Management Guide*; NDSU Extension: Fargo, ND, USA, 2018; Publication E-1143.
20. Larsson, H. A crop loss model and economic thresholds for the grain aphid, *Sitobion avenae* (F.), in winter wheat in southern Sweden. *Crop Prot.* **2005**, *24*, 397–405.
21. Voss, T.S.; Kieckhefer, R.W.; Fuller, B.W.; McLeod, M.J.; Beck, D.A. Yield losses in maturing spring wheat caused by cereal aphids (*Homoptera*: *Aphididae*) under laboratory conditions. *J. Econ. Entomol.* **1997**, *90*, 1346–1350.
22. Johnston, R.L.; Bishop, G.W. Economic injury levels and economic thresholds for cereal aphids (*Homoptera*: *Aphididae*) on spring-planted wheat. *J. Econ. Entomol.* **1987**, *80*, 478–482.
23. Marston, Z.P.; Cira, T.M.; Knight, J.F.; Mulla, D.; Alves, T.M.; Hodgson, E.W.; Koch, R.L. Linear support vector machine classification of plant stress from soybean aphid (*Hemiptera*: *Aphididae*) using hyperspectral reflectance. *J. Econ. Entomol.* **2022**, *115*, 1557–1563.
24. Tanguy, S.; Dedryver, C.A. Reduced *BYDV–PAV* transmission by the grain aphid in a *Triticum monococcum* line. *Eur. J. Plant Pathol.* **2009**, *123*, 281–289.
25. Carvalho, F.P. Agriculture, pesticides, food security and food safety. *Environ. Sci. Policy* **2006**, *9*, 685–692.
26. Luo, J.; Huang, W.; Zhao, J.; Zhang, J.; Zhao, C.; Ma, R. Detecting aphid density of winter wheat leaf using hyperspectral measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. **2013**, *6*, 690–698.
27. Aeberli, A.; Robson, A.; Phinn, S.; Lamb, D.W.; Johansen, K. A comparison of analytical approaches for the spectral discrimination and characterization of mite infestations on banana plants. *Remote Sens.* **2022**, *14*, 5467. https://doi.org/10.3390/rs14215467.
28. Peignier, S.; Lacotte, V.; Duport, M.G.; Baa-Puyoulet, P.; Simon, J.C.; Calevro, F.; da Silva, P. Detection of aphids on hyperspectral images using one-class SVM and Laplacian of Gaussians. *Remote Sens.* **2023**, *15*, 2103.
29. Duveiller, G.; Defourny, P. A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sens. Environ.* **2010**, *114*, 2637–2650.
30. Iost Filho, F.H.; Heldens, W.B.; Kong, Z.; de Lange, E.S. Drones: Innovative technology for use in precision pest management. *J. Econ. Entomol.* **2020**, *113*, 1–25.
31. Ahmad, U.; Alvino, A.; Marino, S. A review of crop water stress assessment using remote sensing. *Remote Sens.* **2021**, *13*, 4155.
32. Omia, E.; Bae, H.; Park, E.; Kim, M.S.; Baek, I.; Kabenge, I.; Cho, B.K. Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sens.* **2023**, *15*, 354.
33. Javaid, M.; Haleem, A.; Khan, I.H.; Suman, R. Understanding the potential applications of Artificial Intelligence in Agriculture Sector. *Adv. Agrochem.* **2023**, *2*, 15–30.
34. Sharma, K.; Shivandu, S.K. Integrating artificial intelligence and internet of things (IoT) for enhanced crop monitoring and management in precision agriculture. *Sens. Int.* **2024**, *100292*.
35. Jones, H.G.; Vaughan, R.A. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*; Oxford University Press: Oxford, UK, 2010.
36. Li, H.; Cui, J.; Zhang, X.; Han, Y.; Cao, L. Dimensionality Reduction and Classification of Hyperspectral Remote Sensing Image Feature Extraction. *Remote Sens.* **2022**, *14*, 4579. https://doi.org/10.3390/rs14184579.
37. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259.
38. Ang, K.L.M.; Seng, J.K.P. Big data and machine learning with hyperspectral information in agriculture. *IEEE Access* **2021**, *9*, 36699–36718.
39. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
40. Myasnikov, E. Using UMAP for dimensionality reduction of hyperspectral data. In Proceedings of the 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, Russia, 6–9 October 2020; pp. 1–5.
41. Kothari, S.; Schweiger, A.K. Plant spectra as integrative measures of plant phenotypes. *J. Ecol.* **2022**, *110*, 2536–2554.
42. Mulla, D.J. Twenty-five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371.
43. Stanton, C.; Starek, M.J.; Elliott, N.; Brewer, M.; Maeda, M.M.; Chu, T. Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. *J. Appl. Remote Sens.* **2017**, *11*, 026035.
44. Yang, Z.; Rao, M.N.; Elliott, N.C.; Kindler, S.D.; Popham, T.W. Differentiating stress induced by greenbugs and Russian wheat aphids in wheat using remote sensing. *Comput. Electron. Agric.* **2009**, *67*, 64–70.

45. Yuan, L.; Huang, Y.; Loraamm, R.W.; Nie, C.; Wang, J.; Zhang, J. Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. *Field Crops Res.* **2014**, *156*, 199–207.
46. Mirik, M.; Michels Jr, G.J.; Kassymzhanova-Mirik, S.; Elliott, N.C.; Bowling, R. Hyperspectral spectrometry as a means to differentiate uninfested and infested winter wheat by greenbug (*Hemiptera*: *Aphididae*). *J. Econ. Entomol.* **2006**, *99*, 1682–1690.
47. Mirik, M.; Michels Jr, G.J.; Kassymzhanova-Mirik, S.; Elliott, N.C.; Catana, V.; Jones, D.B.; Bowling, R. Using digital image analysis and spectral reflectance data to quantify damage by greenbug (*Hemiptera*: *Aphididae*) in winter wheat. *Comput. Electron. Agric.* **2006**, *51*, 86–98.
48. Mirik, M.; Ansley, R.J.; Michels, G.J.; Elliott, N.C. Spectral vegetation indices selected for quantifying Russian wheat aphid (*Diuraphis noxia*) feeding damage in wheat (*Triticum aestivum* L.). *Precis. Agric.* **2012**, *13*, 501–516.
49. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* **2020**, *12*, 3136.
50. Osco, L.P.; Ramos, A.P.M.; Moriya, É.A.S.; de Souza, M.; Junior, J.M.; Matsubara, E.T.; Creste, J.E. Improvement of leaf nitrogen content inference in Valencia-orange trees applying spectral analysis algorithms in UAV mounted-sensor images. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *83*, 101907.
51. Ramos, A.P.M.; Gomes, F.D.G.; Pinheiro, M.M.F.; Furuya, D.E.G.; Gonçalvez, W.N.; Junior, J.M.; Osco, L.P. Detecting the attack of the fall armyworm (*Spodoptera frugiperda*) in cotton plants with machine learning and spectral measurements. *Precis. Agric.* **2022**, *23*, 470–491.
52. Martins, G.D.; Galo, M.D.L.B.T.; Vieira, B.S. Detecting and mapping root-knot nematode infection in coffee crop using remote sensing measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5395–5403.
53. Rammeloo, C.; Baumgartner, A. Spectroradiometer calibration for radiance transfer measurements. *Sensors* **2023**, *23*, 2339.
54. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260.
55. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817.
56. Behmann, J.; Mahlein, A.K.; Rumpf, T.; Römer, C.; Plümer, L. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precis. Agric.* **2015**, *16*, 239–260.
57. Skendžić, S.; Zovko, M.; Lešić, V.; Pajač Živković, I.; Lemić, D. Detection and evaluation of environmental stress in winter wheat using remote and proximal sensing methods and vegetation indices—A review. *Diversity* **2023**, *15*, 481.
58. Bronson, K.F.; Booker, J.D.; Keeling, J.W.; Boman, R.K.; Wheeler, T.A.; Lascano, R.J.; Nichols, R.L. Cotton canopy reflectance at landscape scale as affected by nitrogen fertilization. *Agron. J.* **2005**, *97*, 654–660.
59. Babar, M.A.; Reynolds, M.P.; Van Ginkel, M.; Klatt, A.R.; Raun, W.R.; Stone, M.L. Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* **2006**, *46*, 1046–1057.
60. Bowman, B.C.; Chen, J.; Zhang, J.; Wheeler, J.; Wang, Y.; Zhao, W.; Bonman, J.M. Evaluating grain yield in spring wheat with canopy spectral reflectance. *Crop Sci.* **2015**, *55*, 1881–1890.
61. Oliveira, L.F.; Scharf, P.C.; Vories, E.D.; Drummond, S.T.; Dunn, D.; Stevens, W.G.; Jones, A.S. Calibrating canopy reflectance sensors to predict optimal mid-season nitrogen rate for cotton. *Soil Sci. Soc. Am. J.* **2013**, *77*, 173–183.
62. Skendzic S, Novak H, Lemić D, Pajač Živković I, Zovko M, Maričević M. Hyperspectral Canopy Reflectance Data for Aphid-Infested Winter Wheat Classification [Data set]. Zenodo; 2025. Available online: https://doi.org/10.5281/zenodo.14968258 (accessed on 4 March 2025).
63. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.A.; Kwok, I.W.; Ng, L.G.; Newell, E.W. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **2019**, *37*, 38–44.
64. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65.
65. Venkateswarlu, B.; Shanker, A.K.; Shanker, C.; Maheswari, M. (Eds.) *Crop Stress and Its Management: Perspectives and Strategies*; Springer Science & Business Media: New York, NY, USA, 2011.
66. Prabhakar, M.; Prasad, Y.G.; Rao, M.N. Remote sensing of biotic stress in crop plants and its applications for pest management. In *Crop Stress and Its Management: Perspectives and Strategies*; Venkateswarlu, B., Shanker, A.K., Shanker, C., Maheswari, M., Eds.; Springer: Cham, The Netherlands, 2011; pp. 517–545.
67. Kashyap, B.; Kumar, R. Sensing methodologies in agriculture for monitoring biotic stress in plants due to pathogens and pests. *Inventions* **2021**, *6*, 29.
68. Chen, T.; Zeng, R.; Guo, W.; Hou, X.; Lan, Y.; Zhang, L. Detection of stress in cotton (*Gossypium hirsutum* L.) caused by aphids using leaf level hyperspectral measurements. *Sensors* **2018**, *18*, 2798.

69. Din, M.; Zheng, W.; Rashid, M.; Wang, S.; Shi, Z. Evaluating hyperspectral vegetation indices for leaf area index estimation of *Oryza sativa* L. at diverse phenological stages. *Front. Plant Sci.* **2017**, *8*, 820.
70. Elazab, A.; Bort, J.; Zhou, B.; Serret, M.D.; Nieto-Taladriz, M.T.; Araus, J.L. The combined use of vegetation indices and stable isotopes to predict durum wheat grain yield under contrasting water conditions. *Agric. Water Manag.* **2015**, *158*, 196–208.
71. Zhou, Y.; Lao, C.; Yang, Y.; Zhang, Z.; Chen, H.; Chen, Y.; Chen, J.; Ning, J.;Yang, N. Diagnosis of winter-wheat water stress based on UAV-borne multispectral image texture and vegetation indices. *Agric. Water Manag.* **2021**, *256*, 107076.
72. Zhu, K.; Sun, Z.; Zhao, F.; Yang, T.; Tian, Z.; Lai, J.; Zhu, W.; Long, B. Relating hyperspectral vegetation indices with soil salinity at different depths for the diagnosis of winter wheat salt stress. *Remote Sens.* **2021**, *13*, 250.
73. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.
74. Gitelson, A.A.; Merzlyak, M.N. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space Res.* **1998**, *22*, 689–692. https://doi.org/10.1016/S0273-1177(97)01133-2.
75. de Souza, E.G.; Scharf, P.C.; Sudduth, K.A. Sun position and cloud effects on reflectance and vegetation indices of corn. *Agron. J.* **2010**, *102*, 734–744. https://doi.org/10.2134/agronj2009.0206.
76. Gao, B.C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266.
77. Zhao, D.; Reddy, K.R.; Kakani, V.G.; Read, J.J.; Koti, S. Selection of optimum reflectance ratios for estimating leaf nitrogen and chlorophyll concentrations of field-grown cotton. *Agron. J.* **2005**, *97*, 89–98.
78. Kureel, N.; Sarup, J.; Matin, S.; Goswami, S.; Kureel, K. Modelling vegetation health and stress using hyperspectral remote sensing data. *Model. Earth Syst. Environ.* **2022**, *8*, 1–16.
79. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352.
80. Springer, K.R.; Wang, R.; Gamon, J.A. Parallel seasonal patterns of photosynthesis, fluorescence, and reflectance indices in boreal trees. *Remote Sens.* **2017**, *9*, 691. https://doi.org/10.3390/rs9070691.
81. Punalekar, S.M.; Thomson, A.; Verhoef, A.; Humphries, D.J.; Reynolds, C.K. Assessing suitability of Sentinel-2 bands for monitoring of nutrient concentration of pastures with a range of species compositions. *Agronomy* **2021**, *11*, 1661.
82. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309.
83. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* **2002**, *81*, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4.
84. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
85. Wang, G.; Zeng, F.; Song, P.; Sun, B.; Wang, Q.; Wang, J. Effects of reduced chlorophyll content on photosystem functions and photosynthetic electron transport rate in rice leaves. *J. Plant Physiol.* **2022**, *272*, 153669.
86. Fereres, A.; Gutierrez, C.; Del Estal, P.; Castanera, P. Impact of the English grain aphid, *Sitobion avenae* (F.) (*Homoptera*: *Aphididae*), on the yield of wheat plants subjected to water deficits. *Environ. Entomol.* **1988**, *17*, 596–602.
87. Silva, A.M.; Sampaio, M.V.; de Oliveira, R.S.; Korndorfer, A.P.; Ferreira, S.E.; Polastro, G.C.; Dias, P.A.S. Antibiosis and non-preference of *Sitobion avenae* (F.) (*Hemiptera*: *Aphididae*) on leaves and ears of commercial cultivars of wheat (*Triticum aestivum*). *Neotrop. Entomol.* **2013**, *42*, 304–310.
88. Zarco-Tejada, P.J.; Pushnik, J.C.; Dobrowski, S.; Ustin, S.L. Steady-state chlorophyll *a* fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sens. Environ.* **2003**, *84*, 283–294.
89. Alves, T.M.; Macrae, I.V.; Koch, R.L. Soybean aphid (*Hemiptera*: *Aphididae*) affects soybean spectral reflectance. *J. Econ. Entomol.* **2015**, *108*, 2655–2664.
90. Alves, T.M.; Moon, R.D.; MacRae, I.V.; Koch, R.L. Optimizing band selection for spectral detection of *Aphis glycines* Matsumura in soybean. *Pest Manag. Sci.* **2019**, *75*, 942–949.
91. Marston, Z.P.; Cira, T.M.; Hodgson, E.W.; Knight, J.F.; MacRae, I.V.; Koch, R.L. Detection of stress induced by soybean aphid (*Hemiptera*: *Aphididae*) using multispectral imagery from unmanned aerial vehicles. *J. Econ. Entomol.* **2020**, *113*, 779–786.
92. Knipling, E.B. Physical and physiological basis for reflectance of visible and near infrared radiation from vegetation. *Remote Sens. Environ.* **1970**, *1*, 155–159.

93. Jacquemoud, S.; Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* **1990**, *34*, 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z.
94. Croft, H.; Chen, J.M. Leaf pigment content. In *Comprehensive Remote Sensing*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 117–142. https://doi.org/10.1016/B978-0-12-409548-9.10547-0.
95. Riedell, W.E.; Blackmer, T.M. Leaf reflectance spectra of cereal aphid-damaged wheat. *Crop Sci.* **1999**, *39*, 1835–1840.
96. Yang, C.M.; Cheng, C.H.; Chen, R.K. Changes in spectral characteristics of rice canopy infested with brown planthopper and leaffolder. *Crop Sci.* **2007**, *47*, 329–335.
97. Zovko, M.; Žibrat, U.; Knapič, M.; Kovačić, M.B.; Romić, D. Hyperspectral remote sensing of grapevine drought stress. *Precis. Agric.* **2019**, *20*, 335–347.
98. Farooq, M.; Hussain, M.; Siddique, K.H. Drought stress in wheat during flowering and grain-filling periods. *Crit. Rev. Plant Sci.* **2014**, *33*, 331–349.
99. Smith, A.M.; Bourgeois, G.; Teillet, P.M.; Freemantle, J.; Nadeau, C. A comparison of NDVI and MTVI2 for estimating LAI using CHRIS imagery: A case study in wheat. *Can. J. Remote Sens.* **2008**, *34*, 539–548.
100. Jiang, P.; Zhou, X.; Liu, T.; Guo, X.; Ma, D.; Zhang, C.; Liu, S. Prediction dynamics in cotton aphid using unmanned aerial vehicle multispectral images and vegetation indices. *IEEE Access* **2023**, *11*, 5908–5918.
101. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: Cham, The Netherlands, 2001.
102. Kamran, A.; Iqbal, M.; Spaner, D. Flowering time in wheat (*Triticum aestivum* L.): A key factor for global adaptability. *Euphytica* **2014**, *197*, 1–26.
103. Marino, S.; Alvino, A. Vegetation indices data clustering for dynamic monitoring and classification of wheat yield crop traits. *Remote Sens.* **2021**, *13*, 541.
104. Peralta, D.; Saeys, Y. Robust unsupervised dimensionality reduction based on feature clustering for single-cell imaging data. *Appl. Soft Comput.* **2020**, *93*, 106421.