

Article

Hyperspectral Sensing and Machine Learning for Early Detection of Cereal Leaf Beetle Damage in Wheat: Insights for Precision Pest Management

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 Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia; vinko.lesic@fer.unizg.hr

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 (S.S.); hrvoje.novak@fer.unizg.hr (H.N.); Tel.: +385-998-563-990 (S.S.); +385-997-348-453 (H.N.)

Abstract

The cereal leaf beetle (CLB; *Oulema melanopus* L., Coleoptera: Chrysomelidae) is a serious pest of wheat, capable of causing yield losses of up to 40% through photosynthetic impairment. Early detection and severity assessment are essential for effective and sustainable pest management. This study evaluates the potential of hyperspectral remote sensing (RS) combined with machine learning (ML) for non-invasive detection of CLB-induced stress in winter wheat. Spectral reflectance was measured using a full-range spectroradiometer (350–2500 nm) from flag leaves categorized into four damage levels (healthy, slightly, moderately, and severely damaged). Three input datasets were used for ML classification: full spectral reflectance, a set of 13 vegetation indices (VIs), and outputs of dimensionality reduction technique. CLB stress increased reflectance in the visible range (400–700 nm) and reduced it in the near-infrared (700–1400 nm), consistent with chlorophyll degradation and mesophyll damage. Several VIs, including RIGreen, NDVI750, GNDVI, and NDVI, correlated strongly with damage severity ($\tau = 0.78$ – 0.81). Among the six ML models tested, Support Vector Machine (SVM) achieved the highest classification accuracy of 90.0% (precision = 0.90, recall = 0.90, F1 = 0.90) across the four severity classes, and achieved 91.9% accuracy at the early-detection threshold. As far as the currently available literature indicates, this study provides one of the earliest quantitative assessments of CLB damage severity based on full-spectrum leaf-level hyperspectral reflectance integrated with ML classification. These findings were obtained under controlled, leaf-level measurement conditions and therefore represent a proof-of-concept; future validation using UAV and satellite platforms is needed to assess performance under operational field variability. Overall, our findings highlight the potential of hyperspectral RS and ML for precision pest monitoring, supporting threshold-based decision-making and more sustainable insecticide use.

Keywords: winter wheat; cereal leaf beetle; *Oulema melanopus*; hyperspectral sensing; vegetation indices; machine learning; precision pest management



Academic Editors: Andrzej Borusiewicz, Janusz Zarajczyk and Maciej Kubon

Received: 2 November 2025

Revised: 25 November 2025

Accepted: 26 November 2025

Published: 29 November 2025

Citation: Skendžić, S.; Novak, H.; Zovko, M.; Pajač Živković, I.; Lešić, V.; Maričević, M.; Lemić, D. Hyperspectral Sensing and Machine Learning for Early Detection of Cereal Leaf Beetle Damage in Wheat: Insights for Precision Pest Management. *Agriculture* **2025**, *15*, 2482. <https://doi.org/10.3390/agriculture15232482>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wheat (*Triticum aestivum* L.) is a globally important crop, serving as a major source of calories and protein [1,2]. However, climate change is intensifying pest pressure, exacerbating biotic stress [3,4]. Among key wheat pests, the Cereal leaf beetle (CLB), *Oulema melanopus* L. (Coleoptera: Chrysomelidae) has shown increasing outbreak severity under warming conditions and remains particularly damaging. Larvae cause substantial leaf tissue loss, reducing chlorophyll content and photosynthesis, leading to up to 80% assimilation loss in severe cases [5–10]. Both larvae and adults feed on the upper mesophyll layer, stripping tissue down to the cuticle and creating a distinctive damage pattern [11]. Damage peaks during flowering and grain filling, impacting grain development and yield [10]. Beyond direct feeding damage, CLB infestations can accelerate senescence, reduce tillering, and exacerbate water stress [10,12].

CLB, native to Europe, has become a major invasive pest in North America, causing yield losses of up to 40% in untreated fields [12–17]. In Europe yield losses have been reported at 3% per larva in relation to tiller damage in Germany [10,18] and between 0.3 and 0.5 t/ha in Poland [19]. Managing CLB damage in wheat depends on economic thresholds, which vary by region [20]. In the USA, early thresholds of one larva per stem [12,21] often led to excessive defoliation [17,22]. Current recommendations suggest intervention at 25 eggs or larvae per 100 tillers [23] and phenology-based thresholds of three eggs or larvae per plant before boot stage (BBCH 39) and one larva per flag leaf beyond BBCH 40 [24]. In Europe, chemical control remains the dominant management strategy despite lower CLB impact [10,19,25]. Thresholds vary: one to two larvae per stem at boot stage in Poland [19] and around three larvae per stem in Switzerland [10]. Timely intervention is crucial to minimize yield losses while preserving natural enemies [26]. This highlights the need for well-planned control measures, especially since a single larva can cause up to 10% flag leaf defoliation, while severe damage can lead to yield losses of up to 23% [14,27–29].

Pest control relies heavily on chemicals, often leading to overuse and pest resistance [30]. However, remote sensing (RS) and machine learning (ML) offer new possibilities for early detection and targeted management; unlike traditional visual scouting, which is subjective and limited in scope, RS and ML enable consistent, large-scale detection of early pest damage [31,32]. To maintain agricultural yields, farmers are turning to precision agriculture applications coupled with RS [33]. RS is based on the principle that vegetation, soil, and other objects reflect electromagnetic energy at specific wavelengths depending on their chemical composition, physical properties, and surface characteristics [34]. As a non-invasive monitoring technique, RS enables the assessment of plant health and stress without physical contact, making it highly valuable for precision agriculture. The most important wavelengths for vegetation studies include the visible (400–700 nm), near-infrared (750–1300 nm) and short-wave infrared (1300–2500 nm) ranges [34]. Multi- and hyperspectral sensors measure the spectral reflectance of plants and enable the calculation of vegetation indices (VI) as reliable indicators of plant stress, which have practical applications for site-specific management [35]. For effective remote or proximal sensing of a particular plant stress parameter, it is important to determine the spectral reflectance and VI that exhibit sensitivity to its influence on plant response [36,37]. These metrics can be effectively used to detect hot spots in fields where stressed plants indicate potential problems. By pinpointing specific problem areas, interventions can be strategically focused, enabling proactive and targeted measures to mitigate the impact of damage in fields [38].

While hyperspectral sensing provides highly detailed crop information, its large-scale use remains constrained by cost, workflow complexity, weather windows, and accessibility for smallholder farmers [32,35].

RS data has been widely used to investigate various abiotic and biotic stress factors in wheat crops. Notable examples include the detection of aphid damage in wheat [38–41], the identification of fungal diseases [42–45], the assessment of nutrient status and deficiency [46–48] and the monitoring of drought stress [49–51]. However, none of these studies specifically investigated stress in the form of leaf tissue loss—morphologically and distinct form of damage that directly removes photosynthetic tissue and therefore requires different management actions than other stress factors. Understanding how RS can capture this type of damage through spectral reflectance changes could enhance early detection and improve targeted management strategies.

Machine learning (ML) has become an essential tool in RS for detecting crop stress and monitoring pest impacts [52]. ML approaches can process large, multitemporal datasets and uncover subtle spectral patterns associated with stress responses by evaluating multiple variables simultaneously [53,54]. Pre-processing techniques such as Uniform Manifold Approximation and Projection (UMAP) or Principal Component Analysis (PCA) are commonly used to reduce data dimensionality and improve computational performance while retaining agronomically relevant information [41,55]. Supervised algorithms, including support vector machines (SVM), random forests (RF), and gradient boosting models (GBM), have demonstrated strong performance in agricultural stress classification, particularly when stress symptoms are visually subtle or confounded by environmental conditions [56]. SVM has been especially effective in crop stress detection applications due to its ability to model non-linear boundaries under limited training samples [41,55]. These methods are suitable for differentiating pest-induced damage, such as defoliation caused by the CLB, from other biotic and abiotic stressors. To enhance interpretability, explainable ML tools—such as Shapley Additive exPlanations (SHAP)—are increasingly used to quantify the contribution of individual spectral bands, VI, or temporal features to model outputs [57]. This supports agronomic interpretation by highlighting spectral signatures most indicative of pest-induced stress, enabling improved decision-making in wheat protection and integrated pest management.

Despite extensive remote sensing research on wheat physiology, disease stress and abiotic stress, to our knowledge no published study has quantitatively evaluated CLB damage severity using hyperspectral reflectance integrated with ML classifiers.

The main aim of this study is to demonstrate the advanced potential of hyperspectral RS and ML for early detection and targeted management of CLB infestations in winter wheat.

The specific objectives of this paper include: (1) to investigate the changes in spectral reflectance of wheat plants in response to varying levels of CLB damage to flag leaves; (2) to identify the VI that correlate most strongly with the severity of CLB infestation; and (3) to develop ML algorithms to classify the severity of damage to wheat plants based on spectral data.

2. Materials and Methods

The analytical approach in this study includes data acquisition, processing, and ML-based classification of plant damage levels. As shown in Figure 1, the framework consists of three key stages:

- I. data acquisition;
- II. data analysis;
- III. ML modeling.

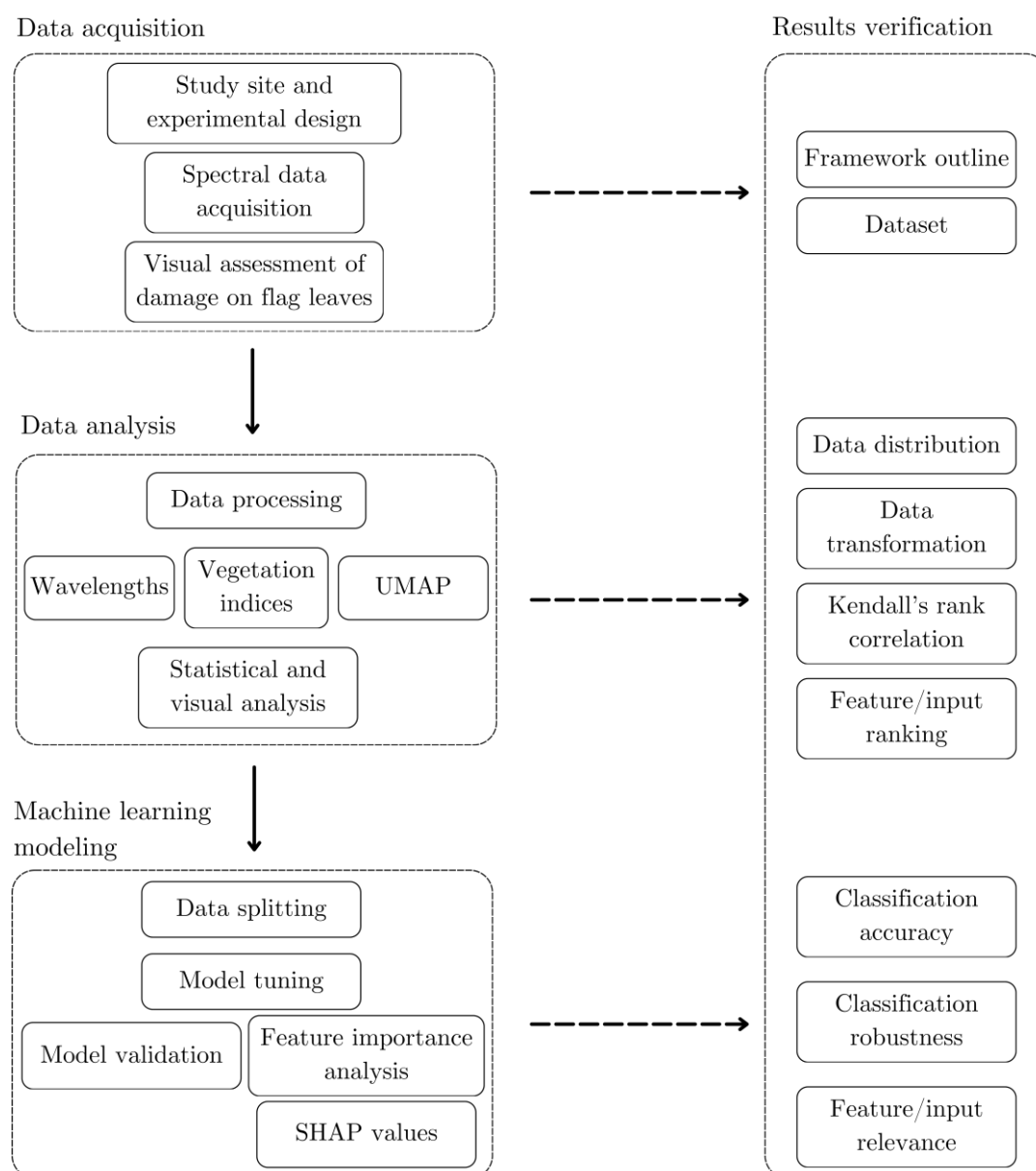


Figure 1. Methodological workflow of the study, from data acquisition to data analysis and machine learning modeling. Solid arrows show sequential processing steps, while dashed arrows link each stage to the corresponding results-verification procedures.

2.1. Data Acquisition

2.1.1. Study Site and Experimental Design

The study was conducted at the experimental fields of Bc Institute for Breeding and Production of Field Crops (45°44′49.1″ N, 15°56′13.4″ E, Zagreb, Croatia) (Figure 2). The institute specializes in seed production and crop breeding programs. The winter wheat variety used was Bc Anica, a common and high-yielding soft wheat variety in this region of Europe. The experiment was conducted on a 240 m² wheat field, sown in rows with a typical spacing of 12 cm. Standard agronomic practices, including fungicide application and fertilization, were applied to maintain consistent growing conditions, with the wheat relying solely on rainfall for water. The study period extended from early April to mid-June 2022 and covered the critical phenophases for CLB infestation on wheat: stem elongation (from BBCH 39), booting (BBCH 40–49), heading (BBCH 50–59), flowering (BBCH 60–69) and milky grain development (up to BBCH 75). Visual assessments and spectral data

acquisition were conducted once during each BBCH phase listed above, covering the entire experimental field.

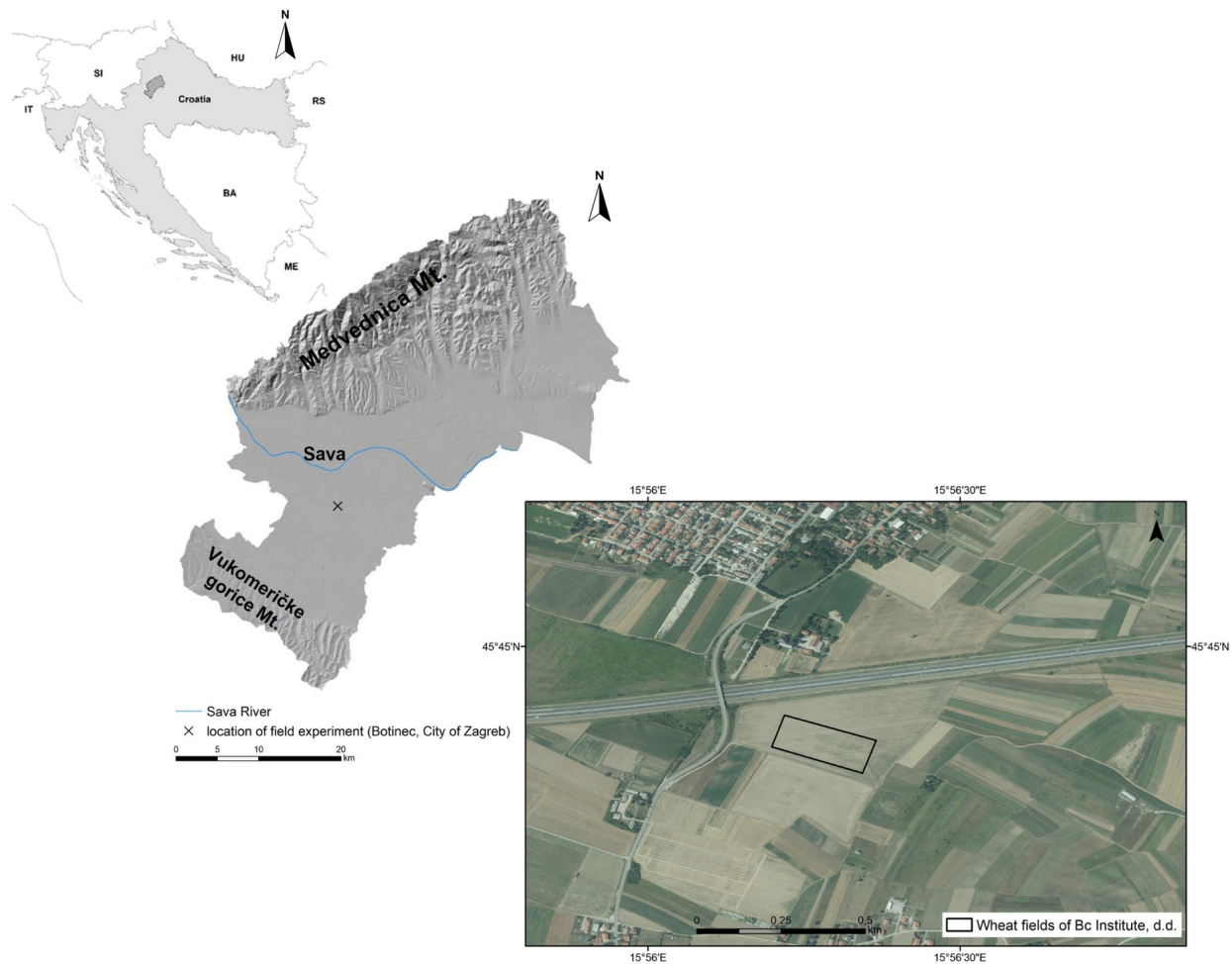


Figure 2. Study site location in Zagreb, Croatia. The upper map shows the regional context, while the lower panel presents a satellite view of the experimental wheat field ($45^{\circ}44'49.1''$ N, $15^{\circ}56'13.4''$ E), where CLB assessments and hyperspectral measurements were conducted.

2.1.2. Visual Assessment of Damage on Flag Leaves

Visual assessments of flag leaf damage caused by CLB were conducted concurrently with the spectral measurements by an experienced entomologist. In the entire experimental field, the plants were inspected and the CLB damage reading was performed according to the standard method. The damage rate in winter wheat was defined as the percentage of flag leaves exhibiting tissue reduction. The damage categories recorded during in situ measurements reflected the most common patterns observed in the field. The rate of damage was estimated by calculating the average percentage of leaf tissue loss across all sampled plants, with a focus on the most frequently occurring damage levels to ensure accurate assessment.

Flag leaf samples were classified into four distinct damage categories:

- I. healthy leaf samples with no visible symptoms, representing 0% damage;
- II. slightly damaged leaves, with 10–15% leaf tissue loss; corresponding to the treatment threshold of 1 larva per flag leaf;
- III. moderately damaged leaves, exhibiting 15–30% leaf tissue loss, corresponding to the treatment threshold of 2–3 larvae per flag leaf (aligns with central European

economic thresholds, where farmers can often tolerate yield losses from damage levels below this threshold);

- IV. severely damaged leaves, showing extensive feeding, with 30–60% leaf tissue loss.

This damage scale aligns with research indicating that a single CLB larva can consume around 10% of a flag leaf surface [14,27,29] providing practical damage thresholds for field management.

2.1.3. Spectral Data Acquisition

Spectral measurements were conducted in situ on flag leaves of winter wheat throughout the BBCH 39–75 phenological phases using a Spectral Evolution® SR-2500 (Haverhill, MA, USA) portable spectroradiometer. The device can measure spectral reflectance of samples and covers the range from 350 to 2500 nm. The spectroradiometer captures data across 2151 spectral bands and automatically resamples the spectra at a 1 nm resolution for export. The spectral resolution varies across the range, with 5 nm in the 350–1000 nm range and 22 nm in the 1000–2500 nm range. A contact probe with an integrated halogen light source was used for precise measurements. Before each scan, any insect fragments or debris were carefully removed from the leaf surface to prevent any interference. Calibration was performed every 10 scans with a white BaSO₄ reference panel (99% Spectralon; SphereOptics GmbH, Herrsching, Germany). To ensure a uniform measurement, the contact probe of the spectroradiometer was positioned directly and vertically above the adaxial side of the flag leaf, with black paper placed under the leaf to avoid light interference. In this way, environmental factors such as lighting conditions and leaf orientation were minimized. This method greatly reduced the influence of external light or shadow, which improved the reliability and consistency of the reflectance data [58]. Spectral measurements were taken at three different points on each flag leaf to enable accurate assessment of infestation severity. The device was connected to a portable computer and the collected data was stored using Darwin SP V1.5 software (Spectral Evolution, Haverhill, MA, USA). This setup allowed for efficient data management and analysis of the collected spectral reflectance.

2.2. Data Analysis

2.2.1. Data Processing

To extract relevant patterns correlating with damage levels, the analysis was conducted through consecutive steps, detailed in the following sections.

Measurements were taken at three spots on each flag leaf to account for within-leaf variability. These three spectral signatures were averaged into a single representative spectrum per leaf using the Darwin SP V1.5 software. Prior to averaging, faulty or inconsistent measurements, as well as spectra that could not be assigned to predefined categories, were removed to ensure data quality. After preprocessing and averaging, the final dataset consisted of 210 unified spectral reflectance measurements, distributed across the four damage categories as follows:

1. Healthy plants: 52;
2. Slightly damaged plants: 52;
3. Moderately damaged plants: 46;
4. Severely damaged plants: 60.

The number of samples in each category was chosen to achieve a relatively balanced distribution across all damage levels. The target variable was encoded in an ordinal manner, with healthy plants labeled as 1 and severely damaged plants labeled as 4.

2.2.2. Data Segmentation

To account for varying data availability and information content, the dataset was segmented into: (i) spectral reflectance, (ii) calculated VI, and (iii) UMAP-derived outputs. Since spectral reflectance acquisition requires expert knowledge, costly equipment, and on-site measurements, VI-based approaches (ii) were explored as an alternative using publicly available RS data (e.g., Copernicus, Landsat). UMAP (iii) was utilized for visual analysis and its relevance to ML modeling, offering improved accuracy and reduced computational demand.

Spectral Reflectance Data

The raw spectral reflectance data, within the wavelength range from 350 to 2500 nm, were considered as the first subset of available data, consisting of 2151 individual reflectance measurements per sample. Reflectance was measured as a percentage for each wavelength, ranging from 0 to 100%. The full dataset is openly available on Zenodo and is provided in the Data Availability Statement [59].

Vegetation Indices (VI)

The selection of VI was based on a comprehensive review of the latest scientific literature (Table 1). VI were carefully chosen to monitor plant responses to pest-induced stress, focusing on physiological parameters such as chlorophyll content and cell structure which are known to change under biotic stress conditions [60–62]. Priority was given to indices that have demonstrated high sensitivity to these factors, ensuring a detailed assessment of CLB damage severity. In addition, several supplementary indices were included to explore potential stress indicators, such as changes in water content, which may reflect subtle or secondary plant responses to pest damage. The 13 VI listed in Table 1, defined using reflectance (R) at specific wavelengths (nm), were employed to analyze the spectral data.

Table 1. List of spectral vegetation indices used in the analysis, defined using reflectance (R) at specific wavelengths (nm).

Vegetation Index	Formula	Source
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{R_{800} - R_{670}}{R_{800} + R_{670}}$	[63]
Red Edge Normalized Difference Vegetation Index (NDVI750)	$NDVI_{750} = \frac{R_{750} - R_{705}}{R_{750} + R_{705}}$	[64]
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{R_{800} - R_{550}}{R_{800} + R_{550}}$	[65]
Carotenoid Reflectance Index 1 (CRI1)	$CRI1 = \frac{1}{R_{510}} - \frac{1}{R_{550}}$	[66]
Carotenoid Reflectance Index (CRI2)	$CRI2 = \frac{1}{R_{510}} - \frac{1}{R_{700}}$	[67]
Structural Independent Pigment Index (SIPI)	$SIPI = \frac{R_{800} - R_{445}}{R_{800} + R_{680}}$	[68]
Chlorophyll Reflectance Index green (RIgreen)	$RI_{green} = \left(\frac{1}{R_{520} - R_{585}} - \frac{1}{R_{750} - R_{800}} \right) \cdot R_{750} - R_{800}$	[69]
Water Band Index (WBI)	$WBI = \frac{R_{900}}{R_{970}}$	[70]
Photochemical Reflectance Index (PRI)	$PRI = \frac{R_{531} - R_{570}}{R_{531} + R_{570}}$	[71]
Moisture Stress Index (MSI)	$MSI = \frac{R_{1599}}{R_{819}}$	[72]
Normalized Difference Water Index (NDWI)	$NDWI = \frac{R_{860} - R_{1240}}{R_{860} + R_{1240}}$	[50]
Red Edge Position (REP)	$REP = 700 + 40 \cdot \frac{\frac{R_{670} + R_{780}}{2} - R_{700}}{R_{740} - R_{700}}$	[73]
Chlorophyll/Carotenoid Index (CCI)	$CCI = \frac{R_{532} - R_{630}}{R_{532} + R_{630}}$	[74]

Uniform Manifold Approximation (UMAP)

To optimize ML-based modeling of plant damage levels, dimensionality reduction was applied to the initial dataset of 2151 spectral measurements per sample. Uniform Manifold Approximation and Projection (UMAP) [75] was selected due to its non-linear transformation capabilities, allowing for the detection of complex patterns [41].

2.2.3. Visual and Statistical Analysis

Visual and statistical analyses were conducted using visualization techniques and summary statistics to identify patterns and anomalies. Kendall's τ was used for correlation analysis, as input variables were continuous and the target variable ordinal [76]:

$$\tau = \frac{C - D}{C + D},$$

where C is the number of concordant pairs and D is the number of discordant pairs.

2.3. Machine Learning (ML) Analysis

The ML analysis focused on differentiating plant damage levels using various algorithms. It was conducted through a series of consecutive steps, detailed in the following sections.

2.3.1. Data Splitting and Processing

The main goal of data splitting was to ensure that the training and testing of the ML models was performed on separate datasets, i.e., that the models were tested on previously unseen data, with no data leakage. The splitting of the overall available dataset, including all 210 data samples, was performed in the same manner, regardless of the number of features utilized as model inputs (subsets i–iii from Section 2.2.2). The training and testing data was randomly separated in a stratified manner with a 70–30 ratio. Stratification ensured that the ratio of different damage levels within the overall dataset was maintained in both the training and testing datasets. To ensure accurate validation of the ML models on a rather small dataset size, a nested cross-validation was implemented (Figure 3). This strategy implied several repetitions of training and testing ML models on different folds of the data (outer loop), while further splitting the training dataset into several folds for tuning the models' hyperparameters and preventing overfitting (inner loop). The outer loop consisted of 10 folds with 70–30 stratified train–test splits, with the results presented as an average over the folds with accompanying variance. The inner fold consisted of 5 folds with 80–20 train–validation stratified splits, with results averaged over the folds, which were utilized only to determine the best set of model hyperparameters on the training set.

In addition to the cross-validation strategy, the data for all input combinations were standardized, i.e., the mean was set to 0 and the standard deviation to 1. Although some of the implemented ML models did not require data standardization, the same unified preprocessing procedure was applied in all cases.

To account for the slight class imbalance in stress severity levels, class-balanced weights (inverse-frequency weighting) were applied during model training for algorithms that support cost-sensitive learning (e.g., SVM and gradient boosting models).

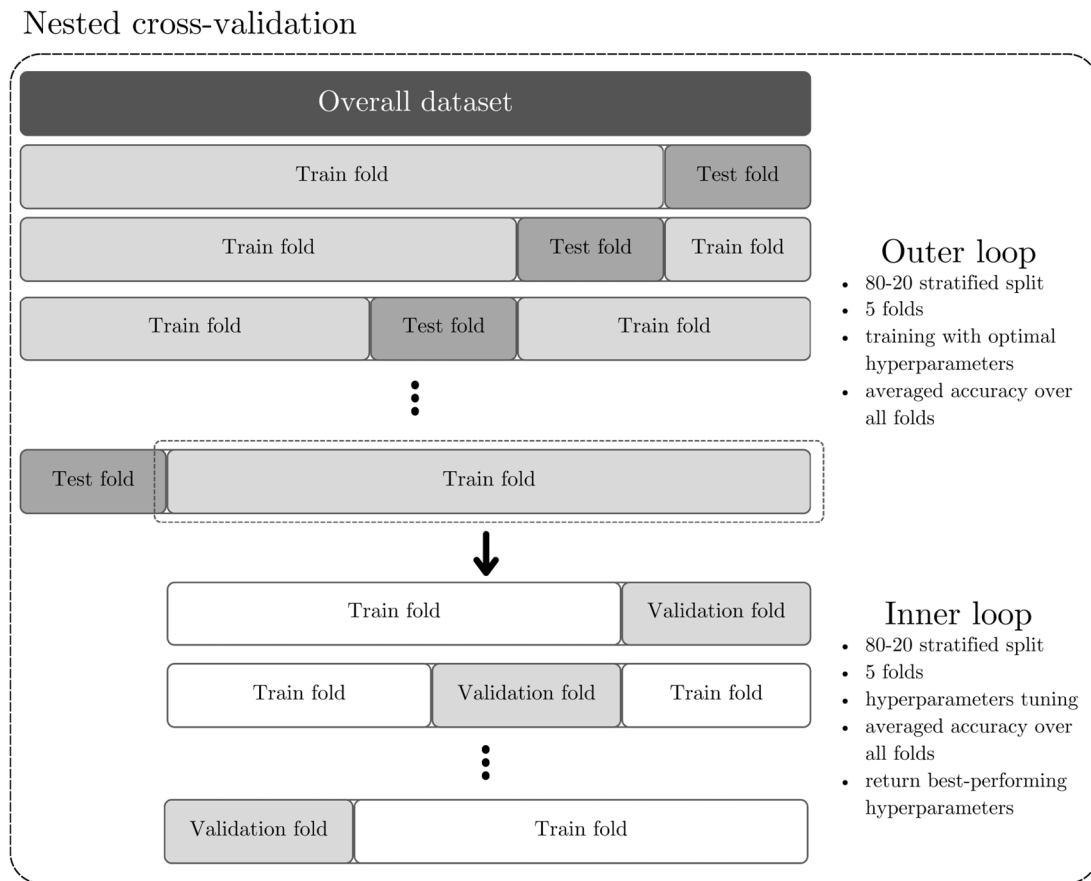


Figure 3. Nested cross-validation workflow. The outer loop applies repeated 70–30 stratified train–test splits (10 folds), while the inner loop divides each training set into 80–20 train–validation folds (5 folds) for hyperparameter tuning. Arrows indicate the sequential flow, and the three dots (“...”) mark the continuation of the procedure through all n-fold iterations.

2.3.2. Machine Learning Models and Models Training

For the ML-based modelling of the posed classification problem, the following selection of models was chosen:

- Logistic regression (LR);
- K Nearest Neighbours (KNN);
- Support Vector Machine (SVM);
- Random Forest (RF);
- LightGBM [77];
- XGBoost [78].

The selected list of models covered a range from simple to more complex. Chosen approaches encompassed a variety of structures (from linear to decision trees ensembles), a range of model and model tuning complexities, in order to obtain a general sense of the ML application contributions towards damage severity classification.

Model training was performed using a nested cross-validation strategy (Figure 3, inner loop). For the tuning of respective model hyperparameters, a Tree-structured Parzen Estimator implemented within the Optuna Python library was utilized [79]. Full details on the methodology are provided in Appendix A, while a complete list of individual model hyperparameters is included in Appendix B (Tables A1 and A2). Performance metrics of the overall model tuning and inference times are provided in Appendix C (Tables A3 and A4).

2.3.3. Models Validation

Validation of the models' accuracy was performed over 10 different stratified data folds, as presented in Figure 3 (outer loop). A selection of most common classification metrics was chosen for the analysis of models' performance:

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP denotes True Positive classifications, TN the True Negative classifications, FP the False Positive classification, and FN the False Negative classifications. Final performance results were presented as averaged metrics over the 10 folds, together with a standard deviation of the obtained metric values to indicate the robustness of individual model performance. Results were presented for individual model performances over each of the 3 separate input combinations (as presented in Section 2.2.2).

In addition to the general multi-class classification across all damage severity levels, model validation was also conducted at two pest management-relevant thresholds: (i) slightly damaged plants (≈ 1 larva per flag leaf; early detection threshold), and (ii) moderately damaged plants ($\approx 2\text{--}3$ larvae per flag leaf; treatment threshold). This enabled assessment of model performance under both early-warning and intervention-oriented decision contexts.

2.3.4. Feature Importance Analysis

To assess the contribution of spectral reflectance and VI, a feature importance analysis was conducted on the best-performing ML model. Global Shapley additive explanations (i.e., SHAP values) [80] were used to quantify feature importance, applying a game-theoretic approach to rank raw spectral wavelengths and VI.

3. Results and Discussion

3.1. Visual and Statistical Analysis

The analysis was performed separately on individual subsets of the data, i.e., (i) spectral reflectance, (ii) set of calculated VI, and (iii) UMAP transformed data.

3.1.1. Response of Leaf Reflectance Spectra to CLB Damage

The spectral reflectance curves for different levels of CLB-induced damage (Figure 4) were shown to display clear patterns throughout VIS, NIR and SWIR spectrum, allowing a distinct separation between healthy and three levels of damaged flag leaves. The following interpretations regarding chlorophyll content, mesophyll structure, and water status are based on observed spectral patterns and supporting literature, as no direct physiological or biochemical measurements were conducted in this study.

In the VIS range (400–700 nm), particularly around 550 nm (green peak), healthy plants exhibited almost 10% lower reflectance compared to severely CLB-affected plants, a pattern reflecting their higher chlorophyll content and more efficient absorption of light for photosynthesis [81,82]. This pattern is interpreted as reflecting lower chlorophyll content in damaged plants, as suggested by previous studies [64,83–86]. In the red region around 650 nm, these differences became even more pronounced: healthy plants absorbed a considerable amount of red light, while damaged plants, especially those with severe damage, reflected more light due to further chlorophyll degradation [84]. This trend

aligns with a previous study [7] which reported a decrease in photosynthetic pigments, including total chlorophyll content, in winter wheat as CLB infestation increased. In addition to chlorophyll, reflectance changes in the blue region (~440–470 nm) may also be linked to alterations in carotenoid content and the carotenoid-to-chlorophyll ratio under stress, as carotenoids strongly absorb blue light and play key roles in photoprotection and antioxidative defense [84–88].

In the NIR range (750–1300 nm), the pattern was reversed: healthy flag leaves exhibited high reflectance due to their intact mesophyll structure, which effectively scatters NIR radiation within the leaf [89,90]. In contrast, damaged leaves reflected less NIR radiation, consistent with mesophyll degradation and water loss that reduce scattering capacity and assimilation efficiency [91,92]. Herbivory damage caused by CLB is known to compromise structural integrity and leaf water content [10], with mesophyll tissue disruption further limiting photosynthetic capacity. Similar trends were observed by Xu et al. [93], who reported markedly lower NIR reflectance in tomato leaves infested with leaf miners compared to healthy controls.

In the SWIR regions, a clear trend was observed between 1550 and 1800 nm, with the most pronounced peak at 1650 nm. At this wavelength, healthy plants exhibited higher reflectance, likely related to their greater water content and structural integrity [10,94–96]. This made 1650 nm a particularly sensitive wavelength for detecting water stress and CLB-induced leaf damage. Spectral reflectance at around 1650 nm also underpins several VI used to estimate leaf water content [94–96]. However, other regions of the SWIR spectrum showed less consistent trends. In addition to water absorption, SWIR reflectance is also affected by proteins, cellulose, lignin [92,97,98], and other biochemical changes in leaf composition [98,99]. Although the spectral patterns described above provide clear separability among damage levels, they reflect leaf-level measurements collected under controlled and uniform conditions. This design ensured that differences in reflectance were attributable to CLB-induced tissue loss rather than variety-, site- or season-driven variability. As noted earlier, the interpretation of pigment- and water-related changes remains inferential, albeit consistent with established literature. While such controlled measurements are appropriate for isolating the spectral signature of CLB damage, they represent a necessary first step before canopy-level validation under more variable field conditions.

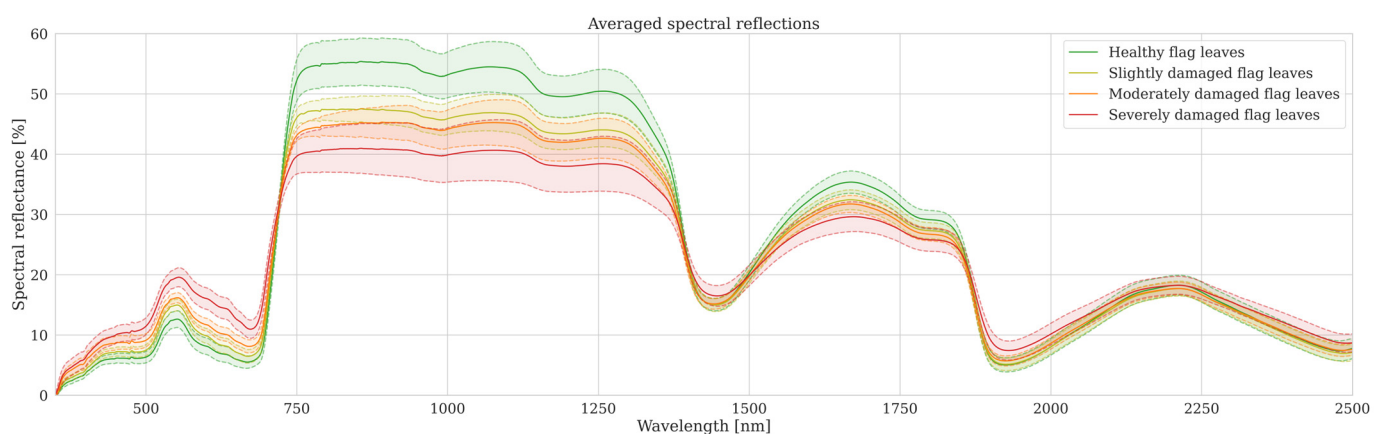


Figure 4. Mean spectral reflectance curves (350–2500 nm) for four CLB damage severity levels (healthy, slightly damaged, moderately damaged, and severely damaged). Solid lines represent mean reflectance, while shaded areas indicate the 95.45% confidence interval for each damage class.

3.1.2. Vegetation Indices (VI)

The VI listed in Table 1 were analyzed. First, box plots of the VI distribution are presented in Figure 5, where for ease of understanding and simpler visualization, all VI

values were z-standardized. All VIs exhibited a decreasing trend with increasing CLB damage, except for SIPI and MSI. Both indices increased progressively from healthy to moderately damaged leaves and showed an additional jump in the severe damage class. This upward deviation in the severe category contrasts sharply with the monotonic decline observed in the other indices and represents a clear trend shift in their response to CLB damage. Additionally, SIPI and NDWI displayed a trend shift, where severe damage levels resulted in slightly higher values compared to the rest of the data.

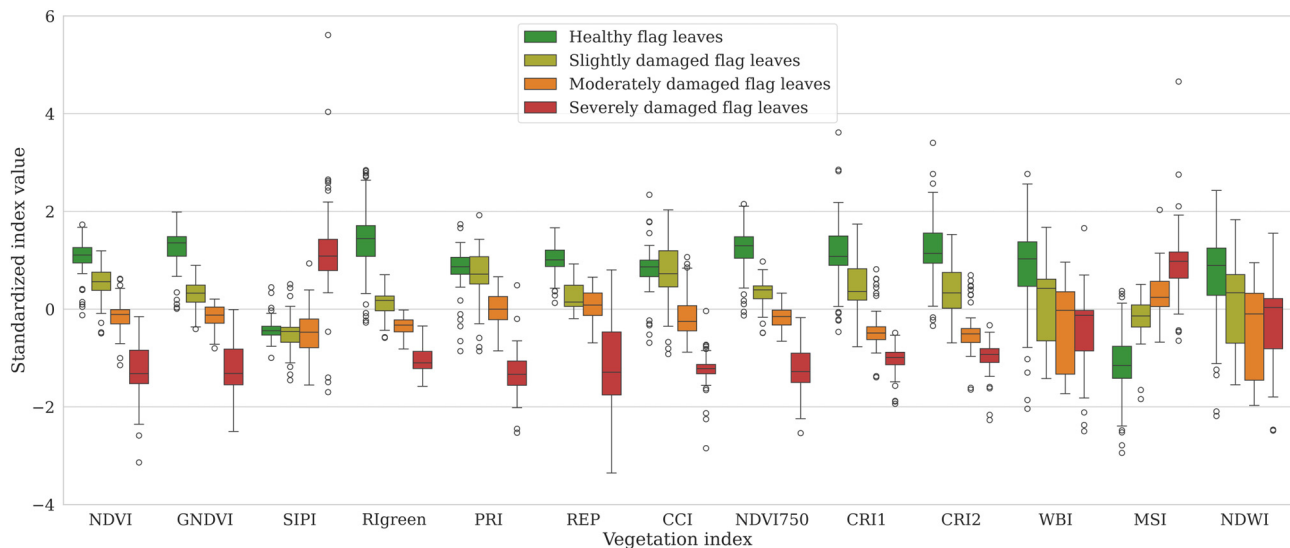


Figure 5. Distribution of vegetation indices (VIs) for four CLB damage severity levels (healthy, slightly damaged, moderately damaged, severely damaged). All VI values were z-standardized to enable comparison across indices with different numerical ranges. Boxplots represent the median, interquartile range, and outliers.

The results were further validated by the Kendall's Rank correlation calculations, presented in the heatmap matrix in Figure 6, which illustrates the correlation between the severity of CLB damage and each of the VI tested.

VI with the highest Kendall correlation coefficients were RIGreen ($\tau = -0.81$), NDVI750 ($\tau = -0.80$), GNDVI ($\tau = -0.80$), and NDVI ($\tau = -0.78$), highlighting their strong relationship with CLB-induced damage levels on flag leaves. The RIGreen stands out due to its more complex formula, which combines multiple wavelength ranges (as seen in Table 1). This index relies heavily on the green range (around 520 nm and 585 nm), along with contributions from the blue (520 nm) and NIR (750–800 nm) regions. Its strong performance can be attributed to its ability to capture variations in the green region, which is highly sensitive to chlorophyll content. Based on the box plot in Figure 5, the RIGreen shows the best separation between healthy leaves (green) and slightly damaged leaves (yellow) making it particularly useful, aiding early CLB detection. Identifying infestations at the onset of larval feeding enables timely intervention, preventing severe damage as larvae progress to highly destructive late instars, which can escalate within days [14]. Additionally, since this damage coincides with the economic threshold of one larva per flag leaf (or 10% flag leaf damage by CLB), the RIGreen could be a valuable tool for estimating the optimal moment for treatment. For the discrimination of healthy plants from moderately damaged plants (15–30% damage on flag leaves), NDVI750 performed the best. This is particularly relevant when insecticide treatments against CLB larvae become necessary, aligning with the economic threshold of two to three larvae per plant. Other indices, such as GNDVI and NDVI, which are widely used [100–103] can also be employed for this purpose. These indices are suitable for mapping infestation hotspots in fields through satellite or drone

imagery [104–106] as CLB infestations typically occur in patches [29,107,108]. By integrating these tools into precision agriculture, growers can generate prescription maps to target specific areas with CLB infestation for insecticide treatments, optimizing resource use and minimizing unnecessary chemical application. Similar strategies have been reviewed [109], highlighting the benefits of precision systems for targeted pest control, which aligns with the findings of this study.

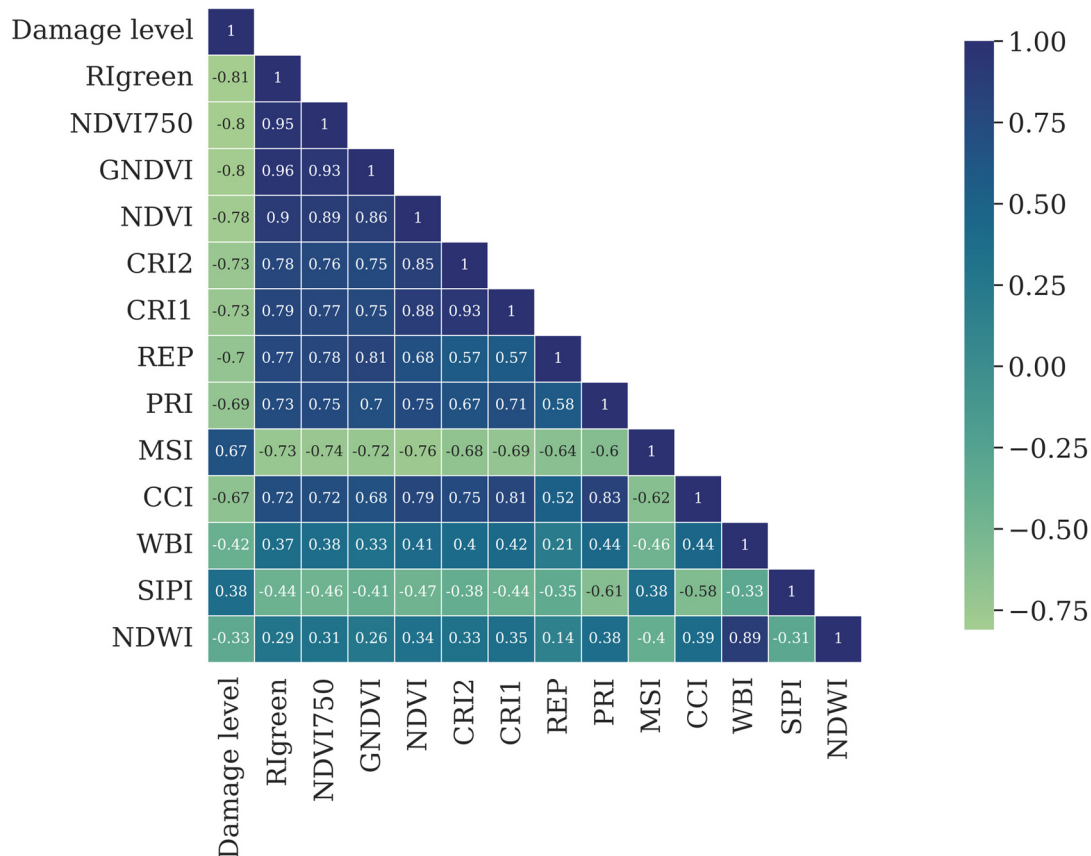


Figure 6. Kendall's τ correlation matrix between vegetation indices and CLB damage severity. Indices are ordered by the strength of their correlation with damage level.

While CLB infestations can affect the water content in wheat, the results for water-related VI in this study, based on Kendall correlation, were mixed. MSI showed a relatively good correlation with plant damage ($\tau = 0.67$), indicating that it could still be useful in assessing water stress as secondary symptom of CLB infested plants. In contrast, WBI ($\tau = -0.42$) and NDWI ($\tau = -0.33$) displayed lower correlation values, suggesting that these indices may not be as effective for detection of damage by CLB.

Several studies have demonstrated the utility of various VI for detecting pest damage. For instance, PRI and SIPI were effective for detecting fungal infections in winter wheat, as they respond to fungal tissue on the leaf surface [110]. Similarly, Das et al. [111] showed that classical indices such as NDVI, RVI, GI, PRI and LMVI1 can discriminate yellow mosaic virus-infected soybean leaves, highlighting the broader applicability of VI for detecting biotic stress across crop species. In contrast, our study focused on CLB-induced leaf tissue loss, which explains why indices like NDVI750, RIgreen and GNDVI sensitive to chlorophyll and structural changes, were more suitable. The key difference lies in the physiological response of the plant: fungal infections alter the leaf surface by introducing fungal structures, changing the reflectance properties of the leaf. In contrast, CLB larvae cause direct tissue loss, creating characteristic white streaks on the leaf as they strip away the outer layers [11] which reduces both the leaf area and its ability to reflect light. Similarly,

aphid infestations, characterized by piercing–sucking feeding, lead to distinct spectral changes [40]. The Aphid Index (AI) was the most effective for monitoring aphid damage, as it leverages red edge and NIR regions to detect internal stress responses without significant tissue loss. This highlights how the type of damage, whether surface alteration or tissue loss, determines which VI are most suitable for detecting different pest damage types.

3.1.3. Uniform Manifold Approximation and Projection (UMAP) Transformation

UMAP dimensionality reduction was performed using optimized hyperparameters ($n_neighbors = 15$, $min_dist = 0.1$), which were selected through a nested cross-validation procedure (5 outer folds, 3 inner folds) combined with a constrained, literature-informed grid search. Within each inner loop, candidate embeddings were evaluated based on silhouette analysis of the obtained data transformation, ensuring that selected parameter values maximized preservation of class-relevant structure while mitigating overfitting given the limited sample size (~200 samples).

The final two-dimensional embeddings (Figure 7) retained meaningful separation between severity categories, with clear clustering in the extreme classes (healthy and severe) and moderate overlap in intermediate stages, reflecting gradual physiological progression of damage rather than discrete categorical boundaries. Although UMAP is unsupervised and does not explicitly optimize for class separability, the embeddings captured relevant spectral variation associated with pest-induced stress while reducing dimensionality from 2151 wavelengths to two latent axes. This confirms that UMAP can effectively compress high-dimensional hyperspectral data while still preserving biologically interpretable gradients of wheat damage severity.

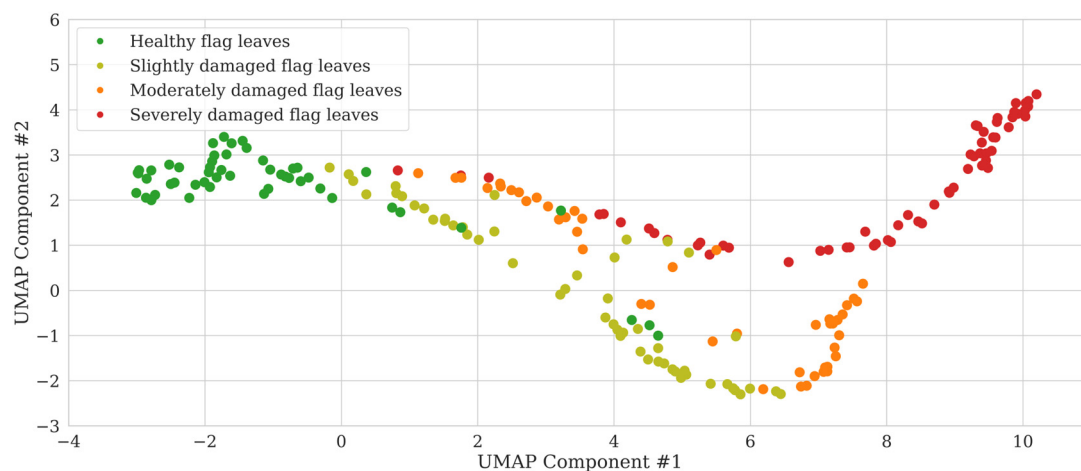


Figure 7. Uniform Manifold Approximation and Projection (UMAP) transformation of hyperspectral data for preassigned damage levels in CLB infested wheat flag leaves.

Finally obtained UMAP transformations demonstrated strong Kendall’s rank correlations, with Component 1 showing a strong correlation of 0.829, and Component 2 a mild correlation of 0.315. These results additionally confirm that UMAP components were closely aligned with the damage levels, rendering it a suitable input dataset for the ML-based approaches.

3.2. Machine Learning (ML) Modelling

For the ML modelling and analysis of the data, 6 different ML models were trained and validated, utilizing the nested cross-validation strategy depicted in Figure 3. The process was repeated for three different sets of inputs: (i) spectral reflectance data (wavelengths), (ii) VI and (iii) UMAP data transformation.

3.2.1. ML Models Validation

The obtained validation results are listed in Table 2, individually for each model as well as for each set of inputs. Bolded values represented the best-performing model for each set of inputs, with the results additionally averaged to obtain a “global” performance of the ML models across the input sets. Overall accuracy results showed good classification performance, with average validation accuracy ranging from 81.43–90.63%. The variance of the performance remained within reasonable levels of <7%, which was expected due to the small size of the overall dataset. Precision and recall results also demonstrated good performance of the models, with false positive (precision) and false negative (recall) values, along with variance, remaining within reasonable amounts.

Table 2. Machine learning models classification validation results. Bolded values represent the best-performing model for each set of inputs, with the results additionally averaged to obtain a global performance of the ML models across the input sets.

		LR	KNN	SVM	RF	LGBM	XGBM	Average
Wavelengths	Accuracy	88.25 ± 2.50	86.98 ± 4.95	90 ± 2.37	89.52 ± 3.68	89.05 ± 2.84	87.94 ± 4.92	88.62 ± 3.54
	Precision	0.89 ± 0.03	0.87 ± 0.05	0.90 ± 0.02	0.90 ± 0.03	0.89 ± 0.03	0.88 ± 0.05	0.88 ± 0.04
	Recall	0.88 ± 0.03	0.87 ± 0.05	0.90 ± 0.02	0.89 ± 0.04	0.89 ± 0.03	0.88 ± 0.05	0.89 ± 0.04
	F1-score	0.88 ± 0.03	0.87 ± 0.05	0.90 ± 0.02	0.89 ± 0.04	0.89 ± 0.03	0.88 ± 0.05	0.89 ± 0.04
VI	Accuracy	87.78 ± 3.09	87.62 ± 2.68	90.63 ± 3.54	86.51 ± 4.18	85.71 ± 3.67	86.83 ± 3.43	87.51 ± 3.43
	Precision	0.88 ± 0.03	0.88 ± 0.02	0.91 ± 0.03	0.87 ± 0.05	0.86 ± 0.04	0.87 ± 0.04	0.88 ± 0.04
	Recall	0.87 ± 0.03	0.87 ± 0.03	0.90 ± 0.04	0.86 ± 0.04	0.85 ± 0.04	0.86 ± 0.03	0.87 ± 0.04
	F1-score	0.87 ± 0.03	0.87 ± 0.03	0.90 ± 0.04	0.86 ± 0.04	0.85 ± 0.04	0.85 ± 0.04	0.87 ± 0.04
UMAP	Accuracy	84.29 ± 3.92	85.55 ± 4.99	88.41 ± 5.55	85.71 ± 5.55	82.38 ± 6.75	81.43 ± 3.89	84.63 ± 5.10
	Precision	0.85 ± 0.04	0.86 ± 0.05	0.89 ± 0.05	0.87 ± 0.05	0.84 ± 0.06	0.82 ± 0.04	0.86 ± 0.05
	Recall	0.84 ± 0.04	0.85 ± 0.05	0.88 ± 0.05	0.85 ± 0.06	0.82 ± 0.07	0.81 ± 0.04	0.84 ± 0.05
	F1-score	0.84 ± 0.04	0.85 ± 0.05	0.88 ± 0.05	0.86 ± 0.05	0.82 ± 0.07	0.81 ± 0.04	0.84 ± 0.05

Among the ML models tested, SVM consistently outperformed others. While advanced tree-based models such as LGBM and XGBM were included, SVM excelled due to its robust mathematical foundations, making it well suited for small, complex datasets [56,112]. In contrast, tree-based ensemble models tend to overfit and require careful hyperparameter tuning, as they typically perform best when larger sample sizes allow reliable partitioning of feature space and sufficient ensemble diversity [56].

The best overall performance was achieved using the full spectral dataset, as some information was lost when transformed into VI or UMAP values. However, the VI-based inputs showed no significant performance drop, highlighting their potential for remote sensing-based CLB stress detection in the field. When contextualized with previous hyperspectral pest- and disease-detection studies, the performance of our models aligns well with published accuracy ranges. A similar study [33] used a Random Forest (RF) classifier to detect fall armyworm (*Spodoptera frugiperda*) damage in cotton and achieved an F-measure of 0.913 by day eight. Near-perfect detection (>99%) has been reported for binary leafminer infestation using SVM classifiers [113] and Susić et al. [114] achieved 90–100% accuracy when distinguishing nematode-infested from drought-stressed tomato leaves. Yellow rust in wheat has similarly been detected with >99% accuracy using MLP-derived spectral features [115]. Moderate performance has also been documented: for example, PLS-DA classification of Potato Y-virus-infected leaves produced a mean validation $\kappa \approx 0.73$ [116]. Additionally, SVM has previously been identified as the best-performing model for classifying soybean rust infection stages [117], reinforcing its robustness for multi-level crop damage classification. By contrast, multi-class arthropod stress detection is substantially more challenging. Nansen et al. [118] reported only 39–68% accuracy when separating whitefly-, mite-, and thrips-infested gerbera plants, with an overall accuracy of 76%. Within this context,

our four-class CLB severity classification achieving 88–91% accuracy falls well within—or above—the expected performance range for multi-level biotic stress discrimination.

Building on this overall performance evaluation, we further examined how well the models operate at agronomically meaningful thresholds relevant for CLB management in a one-versus-all manner: (II) slightly damaged (early detection/economic threshold of approximately 1 larva per flag leaf) (Figure 8) and (III) moderately damaged (treatment threshold of approximately 2–3 larvae per flag leaf) (Figure 9). Across both thresholds, full spectral reflectance (wvls) delivered the most stable and highest accuracy, while VI achieved near-equivalent performance, typically within a few percentage points, supporting their use as scalable inputs for UAV/satellite workflows with only a modest loss of discriminative power [119,120]. In contrast, UMAP-derived inputs were more variable and generally less accurate, consistent with reports that dimensionality reduction can erode class separability in small multi-class hyperspectral datasets [121].



Figure 8. Classifier performance at the early-detection threshold (slightly damaged plants) across spectral inputs (wvls, VI, UMAP) using 10-fold nested cross validation; Accuracy, Precision, Recall, F1-score (mean ± SD).

At the slightly damaged stage (Figure 8), the spectral signal is weaker and overlaps with background variation; therefore, maximising sensitivity (recall) is essential to avoid missing incipient cases, consistent with early-detection studies where pre-symptomatic signals require sensitive features or classifiers [122,123]. By contrast, at the moderately damaged stage (Figure 9), clearer class boundaries yield higher accuracy and precision, so a balanced emphasis on accuracy and F1 is appropriate for treatment decisions. These trends align with reports that classical learners—particularly SVM—perform robustly on high-dimensional spectra, and that VI can serve as practical surrogates for wide-area monitoring [36,37,119,124].

3.2.2. ML Models Feature Importance Analysis

Feature importance analysis was performed for the highest-performing model from Table 2, the SVM model, using the overall dataset that included all four damage severity levels and both types of spectral inputs: vegetation indices (VI) and full reflectance wavelengths. For simplicity, only the top ten most relevant features were presented in the analysis. SVM feature importance obtained based on the SHAP values calculation is presented in Figure 10, both for the wavelengths (nm) and VI datasets.

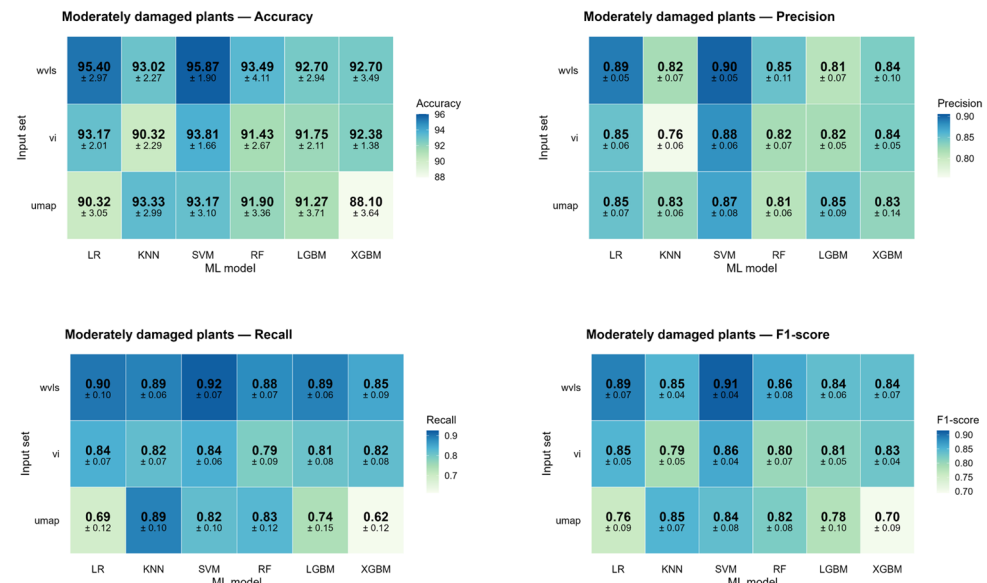


Figure 9. Classifier performance at the treatment threshold (moderately damaged plants) across spectral inputs (wvls, VI, UMAP) using 10-fold nested cross validation; Accuracy, Precision, Recall, F1-score (mean ± SD).

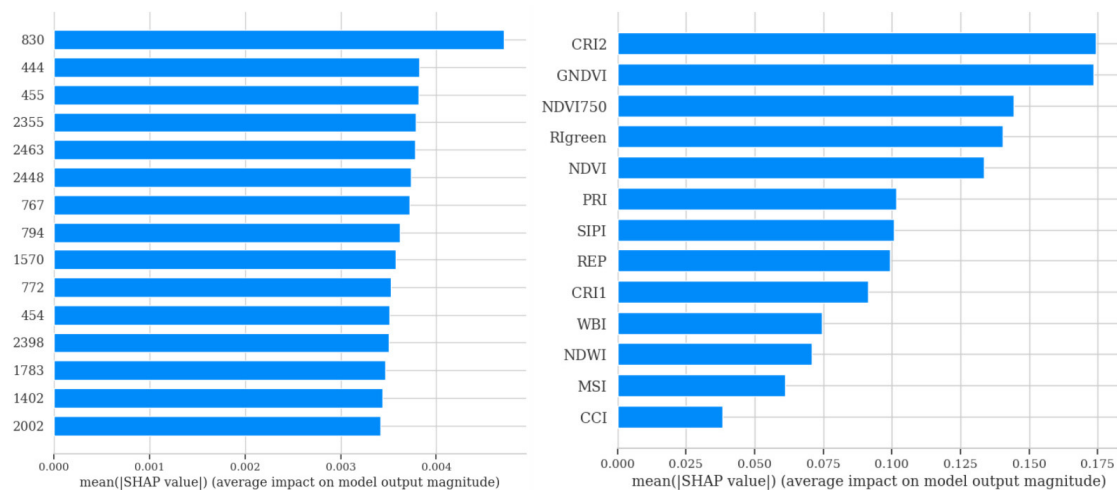


Figure 10. Feature importance quantization for the wavelengths (left) and vegetation indices (right) inputs set of the SVM model, based on SHAP values calculation.

The obtained wavelength importance results correlated well with the analysis presented in Section 3.1.1, particularly Figure 4, which highlighted the differences between the spectral reflectance of different damage levels. It is interesting to note that the SVM algorithm, due to its inherent robustness, uniformly utilized reflectance throughout the entire measured spectrum, with relevance given to specific, more correlated bands. Among the top-ranked features, SHAP highlighted narrow bands in the blue region (444, 454, and 455 nm), which are mainly sensitive to chlorophyll absorption but also to carotenoids that absorb strongly in this part of the spectrum and contribute to photoprotection and stress response [81,85]. In addition, the 830 nm band, located in the red-edge/NIR transition, was the most important single feature, reflecting its dual sensitivity to chlorophyll concentration and changes in leaf structure, making it a robust indicator of early stress effects on photosynthetic efficiency and canopy integrity [125]. Alongside these, bands in the broader NIR and SWIR regions also emerged as highly relevant. The NIR range was important for assessing internal leaf structure, while the SWIR range effectively captured water-related

stress and changes in leaf composition, consistent with previous reports linking these regions to plant physiological status [89–92,94–99].

A similar match was observed between the SHAP values of VI and Kendall's Rank coefficient (see Figure 6), where NDVI, GNDVI, NDVI750, RIGreen and CRI2 were present in the top five VI of both ranking procedures. This was expected as these indices are known for their ability to represent plant health in terms of greenness and vigor. Notably, the CRI2 stood out with the highest SHAP value, indicating its significant impact on the model output. These findings align with those of a previous study [7], which observed increased carotenoid content in wheat genotypes infested with CLB. This suggests that carotenoids, important for photoprotection and antioxidant defense [126] play a role in the plant's stress response to herbivory—supporting the use of CRI2 for detecting such changes. In addition, CRI1, PRI and REP were prominently featured in the SHAP feature importance analysis, indicating their significant contribution to the classification accuracy of the SVM model. These indices were frequently used by the model, emphasizing their role in discriminating damage levels. Although SIPI was highlighted in the analysis, its lower Kendall's rank coefficient and the fact that it primarily distinguishes severely damaged plants (Figure 5) that are already beyond economic thresholds, limit its practical value.

In precision agriculture, analyzing high-dimensional spectral data, such as that from hyperspectral sensors, often requires dimensionality reduction techniques to simplify the data while retaining the most relevant information. Our study has shown that despite the challenges of early pest detection, a refined method combining spectral reflectance, derived VI and UMAP-transformed data has been developed to classify different levels of CLB damage using ML models. The SVM model achieved similar accuracy (88–90%) regardless of whether the full spectral dataset or the reduced data (VI and UMAP) were used, highlighting the effectiveness of this approach. This method proves particularly valuable for decision-making and early intervention, as it allows accurate pest management when economic thresholds are reached, ensuring timely and effective control.

Although leaf-level measurements provide clear spectral separation among CLB damage classes, their transferability to canopy or UAV scales is influenced by factors such as soil background, LAI, leaf angle distribution and multiple scattering [127]. Earlier studies demonstrated that different canopy layers contribute unevenly to the measured signal [127] and that soil and plant architectural variability can alter canopy-level biochemical retrievals [128] supporting the view that leaf-to-canopy transitions are not universally linear [129]. However, wheat represents a favourable case for achieving successful leaf-to-canopy transfer. Its canopy is comparatively shallow, and the flag leaf—both physiologically dominant and structurally exposed—accounts for much of the intercepted radiation and assimilate supply [130]. Because CLB larvae remove tissue directly from this uppermost leaf, the resulting spectral signal is less diluted by lower canopy layers, increasing the likelihood that leaf-level CLB signatures remain detectable at canopy scale, particularly when using proximal or UAV hyperspectral sensors. Radiative transfer models such as PROSAIL, which couple leaf optical properties with canopy architecture, provide a robust framework for exploring how leaf-level changes propagate to canopy reflectance under realistic sun–sensor geometries [131]. While full radiative-transfer modelling was beyond the scope of this study, future work could use such simulations, together with large-scale spectral observations, to more rigorously assess the detectability of CLB damage at canopy level under field heterogeneity.

4. Conclusions

Wheat, as a staple crop, is highly susceptible to various biotic stress factors, including insect pests, which can significantly reduce crop yields. Recent advances in multispectral

and hyperspectral sensors have enabled more precise monitoring of plant stress using spectral reflectance and VI as numerical indicators. As far as current literature indicates, this is among the first studies to investigate the effects of CLB infestation on the spectral reflectance of wheat and provides new insights into how pest-induced stress can be detected non-destructively. A framework was developed to model the spectral response of CLB damage at four severity levels, aligned with economic thresholds for insecticide application. In addition to this full multi-level classification, the framework was also evaluated at two agronomically relevant thresholds (slightly and moderately damaged plants), strengthening its decision-making relevance. Using a full-range spectroradiometer (350–2500 nm) and ML techniques, results confirmed that hyperspectral sensing effectively captures CLB-induced physiological changes. UMAP was used together with ML models such as SVM to efficiently manage large spectral datasets and classify damage levels with high accuracy. Among the models tested, the SVM classifier achieved the strongest performance, reaching 90% overall accuracy across the four severity classes and over 91% accuracy at the early-detection threshold, demonstrating its suitability for operational severity assessment. This approach illustrates how big data in agriculture can be effectively processed using advanced computational techniques. Feature importance analysis identified key VI (GNDVI, NDVI750, RIGreen, NDVI and CRI2), which were closely associated with pigment and structural changes in flag leaves. One of the most important contributions of this research is the potential to simplify and improve pest monitoring practices. Although proximal sensing was the focus of this study, the methodology could be adapted to systems capable of scanning larger areas, such as drone- or satellite-based sensors, bearing in mind that satellite platforms provide only discrete VIS–NIR and selected SWIR channels (e.g., Sentinel-2: 1610 and 2190 nm; Landsat 8/9: 1570–1650 and 2110–2290 nm) rather than full 350–2500 nm hyperspectral coverage. As a result, only part of the diagnostic spectral information identified here would be transferable to satellite applications, whereas UAV-based sensing can retain much of the spectral detail required for accurate damage discrimination. Given the scale of wheat cultivation, this approach offers a scalable solution for field monitoring and early pest detection. Differentiating healthy from slightly/moderately damaged plants enables timely intervention and more precise pest management, including targeted pesticide use. Additionally, the spectral differentiation between damage levels suggests that this approach could be applied to map infestation hotspots, providing valuable insights for generating prescription maps. This work integrates spectral reflectance, VI, and ML modelling for both global and threshold-based classifications, achieving high accuracy and identifying key indices relevant to plant stress, demonstrating scalability to UAV or satellite monitoring. It should also be noted that other abiotic or biotic stresses (e.g., drought, fungal pathogens, nutrient deficiencies) can induce partially overlapping spectral responses, and future work should evaluate model specificity under multi-stress field conditions. Future studies integrating canopy-level or large-scale sensing will be essential to validate these leaf-derived signatures under real field conditions. While results reflect specific experimental conditions, they serve as a proof-of-concept for early detection and agronomically relevant, threshold-based pest management.

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).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The hyperspectral reflectance data supporting the findings of this study are openly available in Zenodo [59]. (accessed on 25 November 2025).

Acknowledgments: We sincerely thank the Bc Institute for Breeding and Production of Field Crops for their invaluable contribution to this research. We also extend our appreciation to the Environmental Protection and Energy Efficiency Fund of the Republic of Croatia for their support.

Conflicts of Interest: Author Marko Maričević is employed by the Bc Institute for Breeding and Production of Field Crops. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CLB	Cereal Leaf Beetle
ML	Machine learning
RS	Remote sensing
VI	Vegetation index/indices
LR	Linear regression
KNN	k-nearest neighbor
RF	Random Forest
LGBM	Light Gradient-Boosting Machine
XGBM	Extreme Gradient Boosting Machine
LAI	Leaf area index

Appendix A. 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 utilization of the corresponding open-source libraries:

Scikit-learn: data processing, PCA, LR, KNN, SVM, RF (version 1.2.2 <https://scikit-learn.org/stable/> (accessed on 11 September 2025)),

uMAP (<https://umap-learn.readthedocs.io/en/latest/> (accessed on 12 September 2024) version 0.1.1),

LightGBM (<https://lightgbm.readthedocs.io/en/latest/> (accessed on 19 September 2025) version 4.1.0),

XGBoost (https://xgboost.readthedocs.io/en/stable/python/python_intro.html (accessed on 20 September 2025) version 1.6.2),

Optuna (<https://optuna.org/> (accessed on 20 September 2025) version 3.2.0),

SHAP (<https://shap.readthedocs.io/en/latest/> (accessed on 8 October 2024) version 0.44.1).

Appendix B

Table A1. Model hyperparameter search space.

Model	Hyperparameter	Range
UMAP	No. of neighbors	[5, 10, 15, 20, 30]
	Minimum distance	[0.0, 0.01, 0.05, 0.1, 0.2, 0.5]
Logistic regression	C	$[1 \times 10^{-3}, 50]$, uniform distribution in the log domain
	Tolerance	$[1 \times 10^{-6}, 1 \times 10^{-3}]$, 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	$[1 \times 10^{-4}, 0.1]$, uniform distribution in the log domain
	No. of estimators	[10, 1000], integer
	Reg alpha	$[1 \times 10^{-8}, 10]$, uniform distribution in the log domain
	Reg lambda	$[1 \times 10^{-8}, 10]$, 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	[5, 50], integer, step = 5
	Min child weight	$[1 \times 10^{-3}, 10]$, float, step = 1×10^{-3}
	Metric	Multi logloss
	Num of classes	4
	Max depth	−1
	Subsample freq	1
	Early stopping rounds	20
XGBoost	Objective	Multi:softmax
	Booster	gbtree
	No. of leaves	[2, 256], integer
	Learning rate	$[1 \times 10^{-4}, 0.1]$, uniform distribution in the log domain
	No. of estimators	[10, 1000], integer

Table A1. *Cont.*

Model	Hyperparameter	Range
XGBoost	Alpha	$[1 \times 10^{-8}, 10]$, uniform distribution in the log domain
	Lambda	$[1 \times 10^{-8}, 10]$, 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	$[5, 50]$, integer, step = 5
	Min child weight	$[1 \times 10^{-3}, 10]$, float, step = 1×10^{-3}
	Metric	Multi logloss
	Num of classes	4
	Subsample freq	1
	Early stopping rounds	20

gray-shaded fields denote fixed hyperparameters.

Appendix C

Table A2. Model tuning performance times—wavelength inputs set.

Model	Single Trial Time (s)	Total CV Tuning Time (s)	Inference Time per Sample (ms)
LR	1.2879 ± 0.0643	64.39 ± 3.22	0.0155 ± 0.0015
KNN	0.3867 ± 0.0455	19.34 ± 2.28	0.0376 ± 0.0109
SVM	0.3224 ± 0.0398	16.11 ± 1.99	0.0237 ± 0.0058
RF	21.0088 ± 6.9670	1050.44 ± 348.35	0.0427 ± 0.0101
LGBM	1.0313 ± 0.1558	51.56 ± 7.79	0.0924 ± 0.0081
XGBM	8.4305 ± 2.9326	421.52 ± 146.63	0.0215 ± 0.0052

Table A3. Model tuning performance times—vegetation indices inputs set.

Model	Single Trial Time (s)	Total CV Tuning Time (s)	Inference Time per Sample (ms)
LR	0.1013 ± 0.0100	5.06 ± 0.50	0.0020 ± 0.0021
KNN	0.0570 ± 0.0153	2.85 ± 0.76	0.0030 ± 0.0032
SVM	0.0598 ± 0.0158	2.99 ± 0.79	0.0032 ± 0.0059
RF	1.8527 ± 0.3578	92.64 ± 17.89	0.0176 ± 0.0102
LGBM	0.2271 ± 0.0214	11.36 ± 1.07	0.0074 ± 0.0049
XGBM	1.2151 ± 0.3844	60.75 ± 19.22	0.0019 ± 0.0022

Table A4. Model tuning performance times—UMAP-transformed inputs set.

Model	Single Trial Time (s)	Total CV Tuning Time (s)	Inference Time per Sample (ms)
LR	0.0747 ± 0.0146	3.73 ± 0.73	0.0018 ± 0.0028
KNN	0.0592 ± 0.0215	2.96 ± 1.07	0.0018 ± 0.0023
SVM	0.0570 ± 0.0183	2.85 ± 0.91	0.0011 ± 0.0008
RF	2.2010 ± 0.2254	110.05 ± 11.27	0.0301 ± 0.0140
LGBM	0.2268 ± 0.0153	11.34 ± 0.76	0.0071 ± 0.0108
XGBM	1.0016 ± 0.3155	50.58 ± 15.77	0.0034 ± 0.0031

The overall ML models training and validation procedure lasted for 5 h and 30 min on a machine with 4 cores, 3.5 GHz clock and 16 GB of RAM.

References

1. Semenov, M.A.; Stratonovitch, P.; Alghabari, F.; Gooding, M.J. Adapting wheat in Europe for climate change. *J. Cereal Sci.* **2014**, *59*, 245–256. [\[CrossRef\]](#)
2. Zhao, K.; Tao, Y.; Liu, M.; Yang, D.; Zhu, M.; Ding, J.; Zhu, X.; Guo, W.; Zhou, G.; Li, C. Does temporary heat stress or low temperature stress similarly affect yield, starch, and protein of winter wheat grain during grain filling? *J. Cereal Sci.* **2022**, *103*, 103408. [\[CrossRef\]](#)
3. Matzrafi, M. Climate change exacerbates pest damage through reduced pesticide efficacy. *Pest Manag. Sci.* **2019**, *75*, 9–13. [\[CrossRef\]](#) [\[PubMed\]](#)
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. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Hatchett, J.H.; Starks, K.J.; Webster, J.A. Insect and mite pests of wheat. In *Wheat and Wheat Improvement*, 2nd ed.; Heyne, E.G., Ed.; Agronomy Monograph 13; ASA, CSSA, SSSA: Madison, WI, USA, 1987; pp. 625–675.
6. Lukács, H.; Jócsák, I.; Somfalvi-Tóth, K.; Keszthelyi, S. Physiological responses manifested by some conventional stress parameters and biophoton emission in winter wheat as a consequence of cereal leaf beetle infestation. *Front. Plant Sci.* **2022**, *13*, 839855. [\[CrossRef\]](#)
7. Parthiban, M.; Jan, F.; Mantoo, M.A.; Kaur, S.; Rustgi, S.; Wani, F.J.; Sofi, P.A.; Najeeb, S.; Sharma, M.; Sofi, N.R. Antioxidant and secondary metabolite responses in wheat under cereal leaf beetle (*Oulema melanopus* L.) infestation. *Physiol. Plant.* **2024**, *176*, e14158. [\[CrossRef\]](#)
8. Kogan, M.; Gerling, D.; Maddox, J.V. Enhancement of biological control in annual agricultural environments. In *Handbook of Biological Control*; Bellows, T., Fisher, T., Eds.; Academic Press: New York, NY, USA, 1999; pp. 789–818.
9. Campbell, J.M.; Sarazin, M.J.; Lyons, D.B. *Canadian Beetles (Coleoptera) Injurious to Crops, Ornamentals, Stored Products, and Buildings*; Agriculture Canada: Ottawa, ON, Canada, 1989.
10. Steinger, T.; Klötzli, F.; Ramseier, H. Experimental assessment of the economic injury level of the cereal leaf beetle (Coleoptera: Chrysomelidae) in winter wheat. *J. Econ. Entomol.* **2020**, *113*, 1823–1830. [\[CrossRef\]](#)
11. Batistič, L.; Vučajnk, F.; Vidrih, M.; Horvat, A.; Košir, I.J.; Šilc, U.; Bohinc, T.; Trdan, S. Evaluating locally sourced inert dusts as insecticides against the cereal leaf beetle (*Oulema melanopus* [L.]; Coleoptera: Chrysomelidae): A combined laboratory and field study. *Cogent Food Agric.* **2024**, *10*, 2415393. [\[CrossRef\]](#)
12. Haynes, D.L.; Gage, S.H. The cereal leaf beetle in North America. *Annu. Rev. Entomol.* **1981**, *26*, 259–287. [\[CrossRef\]](#)
13. Olfert, O.; Weiss, R.M.; Woods, S.; Philip, H.; Dosdall, L. Potential distribution and relative abundance of an invasive cereal crop pest, *Oulema melanopus* (Coleoptera: Chrysomelidae), in Canada. *Can. Entomol.* **2004**, *136*, 277–287. [\[CrossRef\]](#)
14. Philips, C.R.; Herbert, D.A.; Kuhar, T.P.; Reisig, D.D.; Thomason, W.E.; Malone, S. Fifty years of cereal leaf beetle in the US: An update on its biology, management, and current research. *J. Integr. Pest Manag.* **2011**, *2*, C1–C5. [\[CrossRef\]](#)
15. Lajos, K.; Császár, O.; Sárospataki, M.; Samu, F.; Tóth, F. Linear woody landscape elements may help to mitigate leaf surface loss caused by the cereal leaf beetle. *Landsc. Ecol.* **2020**, *35*, 2225–2238. [\[CrossRef\]](#)
16. Vijver, E.; Landschoot, S.; Roie, M.; Temmerman, F.; Dillen, J.; Ceuleners, K.; Wauters, A.; Haesaert, G. Inter- and intrafield distribution of cereal leaf beetle species (Coleoptera: Chrysomelidae) in Belgian winter wheat. *Environ. Entomol.* **2019**, *48*, 276–283. [\[CrossRef\]](#)
17. Buntin, G.D.; Flanders, K.L.; Slaughter, R.W.; Delamar, Z.D. Damage loss assessment and control of the cereal leaf beetle (Coleoptera: Chrysomelidae) in winter wheat. *J. Econ. Entomol.* **2004**, *97*, 374–382. [\[CrossRef\]](#)
18. Kirch, G. *Auftreten und Bekämpfung Phytophager Insekten an Getreide und Raps in Schleswig-Holstein*; Institut für Phytopathologie und Angewandte Zoologie, Justus-Liebig-Universität Gießen: Gießen, Germany, 2006.
19. Wenda-Piesik, A.; Kazek, M.; Piesik, D. Cereal leaf beetles (*Oulema* spp., Coleoptera: Chrysomelidae) control following various dates of wheat sowing and insecticidal treatments. *Int. J. Pest Manag.* **2018**, *64*, 157–165. [\[CrossRef\]](#)
20. Thiel, L.; Mergenthaler, M.; Wutke, M.; Haberland-Korr, V. Use of insect pest thresholds in oilseed rape and cereals: Is it worth it? *Pest Manag. Sci.* **2024**, *80*, 2353–2361. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Webster, J.A.; Smith, D.H.; Lee, C. Reduction in yield of spring wheat caused by cereal leaf beetles. *J. Econ. Entomol.* **1972**, *65*, 832–835. [\[CrossRef\]](#)
22. Ihrig, R.A.; Herbert, D.A.; Van Duyn, J.W.; Bradley, J.R. Relationship between cereal leaf beetle (Coleoptera: Chrysomelidae) egg and fourth-instar populations and impact of fourth-instar defoliation of winter wheat yields in North Carolina and Virginia. *J. Econ. Entomol.* **2001**, *94*, 634–639. [\[CrossRef\]](#)

23. Herbert, D.A.; Van Duyn, J.W. *Cereal Leaf Beetle: Biology and Management*; Virginia Cooperative Extension Publication: Blacksburg, VA, USA, 2009.
24. Glogoza, P. *North Dakota Small Grain Insects: Cereal Leaf Beetle*; North Dakota State University Extension: Fargo, ND, USA, 2002.
25. Tanasković, S.; Madić, M.; Đurović, D.; Knežević, D.; Vukajlović, F. Susceptibility of cereal leaf beetle (*Oulema melanopa* L.) in winter wheat to various foliar insecticides in western Serbia region. *Rom. Agric. Res.* **2012**, *29*, 361–365.
26. Reisig, D.D.; Bacheler, J.S.; Herbert, D.A.; Kuhar, T.P.; Malone, S.; Philips, C.R.; Weisz, R. Efficacy and value of prophylactic vs. integrated pest management approaches for management of cereal leaf beetle (Coleoptera: Chrysomelidae) in wheat and ramifications for adoption by growers. *J. Econ. Entomol.* **2012**, *105*, 1612–1619. [[CrossRef](#)]
27. Heyer, W. Biologie und Schadwirkung der Getreidehähnchen *Lema* (*Oulema* spp.) in der industriemäßigen Getreideproduktion. *Nachrbl. Pflanzensch. DDR* **1977**, *31*, 167–169.
28. Gallun, R.L.; Everly, R.T.; Yamazaki, W.T. Yield and milling quality of Monon wheat damaged by feeding of cereal leaf beetle. *J. Econ. Entomol.* **1967**, *60*, 356–359. [[CrossRef](#)]
29. Samu, F.; Szita, É.; Simon, J.; Cséplő, M.; Botos, E.; Pertics, B.; Růžicková, J.; Gerstenbrand, R.; Rakszegi, M.; Elek, Z.; et al. Cereal leaf beetle (*Oulema* spp.) damage reduces yield and is more severe when natural enemy action is prevented. *Crop Prot.* **2024**, *185*, 106893. [[CrossRef](#)]
30. 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. [[CrossRef](#)]
31. Kelly, M.; Guo, Q. Integrated agricultural pest management through remote sensing and spatial analyses. In *General Concepts in Integrated Pest and Disease Management*; Oliver, M.B., Raney, A.A., Bryant, J., Eds.; Springer: Dordrecht, The Netherlands, 2007; pp. 191–207.
32. Iost Filho, F.H.; Heldens, W.B.; Kong, Z.; Lange, E.S. Drones: Innovative technology for use in precision pest management. *J. Econ. Entomol.* **2020**, *113*, 1–25. [[CrossRef](#)] [[PubMed](#)]
33. Ramos, A.P.M.; Gomes, F.D.G.; Pinheiro, M.M.F.; Furuya, D.E.G.; Gonçalves, 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. [[CrossRef](#)]
34. Jones, H.G.; Vaughan, R.A. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*; Oxford University Press: Oxford, UK, 2010.
35. Mulla, D.J. Twenty-five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [[CrossRef](#)]
36. 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. [[CrossRef](#)]
37. Pinto, J.; Powell, S.; Peterson, R.; Rosalen, D.; Fernandes, O. Detection of defoliation injury in peanut with hyperspectral proximal remote sensing. *Remote Sens.* **2020**, *12*, 3828. [[CrossRef](#)]
38. 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. [[CrossRef](#)]
39. 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. [[CrossRef](#)]
40. 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. [[CrossRef](#)]
41. 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. [[CrossRef](#)]
42. Huang, W.; Lamb, D.W.; Niu, Z.; Zhang, Y.; Liu, L.; Wang, J. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. *Precis. Agric.* **2007**, *8*, 187–197. [[CrossRef](#)]
43. Ashourloo, D.; Mobasheri, M.R.; Huete, A. Developing two spectral disease indices for detection of wheat leaf rust (*Puccinia triticina*). *Remote Sens.* **2014**, *6*, 4723–4740. [[CrossRef](#)]
44. Yu, K.; Anderegg, J.; Mikaberidze, A.; Karisto, P.; Mascher, F.; McDonald, B.A.; Walter, A.; Hund, A. Hyperspectral canopy sensing of wheat septoria tritici blotch disease. *Front. Plant Sci.* **2018**, *9*, 1195. [[CrossRef](#)]
45. Kremneva, O.Y.; Danilov, R.Y.; Sereda, I.I.; Tutubalina, O.V.; Pachkin, A.A.; Zimin, M.V. Spectral characteristics of winter wheat varieties depending on the development degree of *Pyrenophora tritici-repentis*. *Precis. Agric.* **2023**, *24*, 830–852. [[CrossRef](#)]
46. Zhu, Y.; Yao, X.; Tian, Y.; Liu, X.; Cao, W. Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 1–10. [[CrossRef](#)]
47. Jia, L.; Yu, Z.; Li, F.; Gnyp, M.; Koppe, W.; Bareth, G.; Zhang, F. Nitrogen status estimation of winter wheat by using an IKONOS satellite image in the North China Plain. In *Computer and Computing Technologies in Agriculture V*; Li, D., Chen, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 174–184.

48. Yao, X.; Ren, H.; Cao, Z.; Tian, Y.; Cao, W.; Zhu, Y.; Cheng, T.; Tian, Y.; Han, J. Detecting leaf nitrogen content in wheat with canopy hyperspectrum under different soil backgrounds. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *32*, 114–124. [\[CrossRef\]](#)
49. Liu, L.; Zhang, S.; Zhang, B. Evaluation of hyperspectral indices for retrieval of canopy equivalent water thickness and gravimetric water content. *Int. J. Remote Sens.* **2016**, *37*, 3384–3399. [\[CrossRef\]](#)
50. Sun, H.; Feng, M.; Xiao, L.; Yang, W.; Wang, C.; Jia, X.; Zhao, Y.; Zhao, C.; Muhammad, S.K.; Li, D. Assessment of plant water status in winter wheat (*Triticum aestivum* L.) based on canopy spectral indices. *PLoS ONE* **2019**, *14*, e0216890. [\[CrossRef\]](#)
51. Ren, S.; Guo, B.; Wang, Z.; Wang, J.; Fang, Q.; Wang, J. Optimized spectral index models for accurately retrieving soil moisture (SM) of winter wheat under water stress. *Agric. Water Manag.* **2022**, *261*, 107333. [\[CrossRef\]](#)
52. Wang, S.; Xu, D.; Liang, H.; Bai, Y.; Li, X.; Zhou, J.; Su, C.; Wei, W. Advances in Deep Learning Applications for Plant Disease and Pest Detection: A Review. *Remote Sens.* **2025**, *17*, 698. [\[CrossRef\]](#)
53. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [\[CrossRef\]](#) [\[PubMed\]](#)
54. 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. [\[CrossRef\]](#)
55. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.-A.; Kwok, I.W.H.; Ng, L.G.; Ginhoux, F.; Newell, E.W. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **2019**, *37*, 38–44. [\[CrossRef\]](#)
56. Aman, M.; Ullah, K.Z.; Javed, K.; Sattar, A.M.; Aamir, A.; Nida, J.; Ziqi, H.; Fuzhong, L. A comprehensive review of crop stress detection: Destructive, non-destructive, and ML-based approaches. *Front. Plant Sci.* **2025**, *16*, 1638675. [\[CrossRef\]](#)
57. Zhao, Y.; Xu, D.; Li, S.; Tang, K.; Yu, H.; Yan, R.; Ren, H.; Zhang, Y.; Wang, J.; Liu, L. Comparative analysis of feature importance algorithms for grassland aboveground biomass and nutrient prediction using hyperspectral data. *Agriculture* **2024**, *14*, 389. [\[CrossRef\]](#)
58. Potůčková, M.; Červená, L.; Kupková, L.; Lhotáková, Z.; Lukeš, P.; Hanuš, J.; Novotný, J.; Albrechtová, J. Comparison of reflectance measurements acquired with a contact probe and an integration sphere: Implications for the spectral properties of vegetation at a leaf level. *Sensors* **2016**, *16*, 1801. [\[CrossRef\]](#)
59. Skendžić, S.; Novak, H.; Pajač Živković, I.; Zovko, M.; Lešić, V.; Maričević, M.; Lemić, D. Hyperspectral Reflectance Dataset (350–2500 nm) of Wheat Flag Leaves Under Four Levels of Cereal Leaf Beetle Damage [Data set]. Zenodo. 2025. Available online: <https://zenodo.org/records/17507507> (accessed on 25 November 2025).
60. Glenn, E.P.; Huete, A.R.; Nagler, P.L.; Nelson, S.G. Relationship between remotely sensed vegetation indices, canopy attributes, and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors* **2008**, *8*, 2136–2160. [\[CrossRef\]](#)
61. Fiorani, F.; Rascher, U.; Jahnke, S.; Schurr, U. Imaging plants dynamics in heterogeneous environments. *Curr. Opin. Biotechnol.* **2012**, *23*, 227–235. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Karnieli, A.; Bayarjargal, Y.; Bayasgalan, M.; Mandakh, B.; Dugarjav, C.; Burgheimer, J.; Khudulmur, S.; Bazha, S.N.; Gunin, P.D. Do vegetation indices provide a reliable indication of vegetation degradation? A case study in the Mongolian pastures. *Int. J. Remote Sens.* **2013**, *34*, 6243–6262. [\[CrossRef\]](#)
63. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [\[CrossRef\]](#)
64. Gitelson, A.A.; Merzlyak, M.N. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space Res.* **1998**, *22*, 689–692. [\[CrossRef\]](#)
65. 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. [\[CrossRef\]](#)
66. Demarchi, L.; Kania, A.; Ciężkowski, W.; Piórkowski, H.; Oświecimska-Piasko, Z.; Chormański, J. Recursive feature elimination and random forest classification of Natura 2000 grasslands in lowland river valleys of Poland based on airborne hyperspectral and LiDAR data fusion. *Remote Sens.* **2020**, *12*, 1842. [\[CrossRef\]](#)
67. Iozia, L.M.; Varone, L. Short range shifts in plant physiological responses to induced water stress: Experimental evidence of intraspecific trait variability differentiating neighbouring Mediterranean plant populations. *Plant Stress* **2024**, *13*, 100556. [\[CrossRef\]](#)
68. Peñuelas, J.; Filella, I.; Gamon, J.A. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. *New Phytol.* **1995**, *131*, 291–296. [\[CrossRef\]](#)
69. España Boquera, M.L.; Lobit, P.; Castellanos Morales, V. Leaf chlorophyll content estimation in the monarch butterfly biosphere reserve. *Rev. Fitotec. Mex.* **2010**, *33*, 175–181. [\[CrossRef\]](#)
70. Peñuelas, J.; Filella, I.; Biel, C.; Serrano, L.; Savé, R. The reflectance at the 950–970 nm region as an indicator of plant water status. *Int. J. Remote Sens.* **1993**, *14*, 1887–1905. [\[CrossRef\]](#)
71. Gamon, J.; Serrano, L.; Surfus, J.S. The photochemical reflectance index: An optical indicator of photosynthetic radiation-use efficiency across species, functional types, and nutrient levels. *Oecologia* **1997**, *112*, 492–501. [\[CrossRef\]](#)

72. Hunt, E.R., Jr.; Rock, B.N. Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sens. Environ.* **1989**, *30*, 43–54.
73. Ali, A.; Imran, M. Evaluating the potential of red edge position (REP) of hyperspectral remote sensing data for real-time estimation of LAI and chlorophyll content of kinnow mandarin (*Citrus reticulata*) fruit orchards. *Sci. Hortic.* **2020**, *267*, 109326. [\[CrossRef\]](#)
74. 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. [\[CrossRef\]](#)
75. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
76. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. [\[CrossRef\]](#)
77. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017.
78. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016)*, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
79. 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 (KDD 2019)*, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
80. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
81. Datt, B. Remote Sensing of Chlorophyll *a*, Chlorophyll *b*, Chlorophyll *a+b*, and Total Carotenoid Content in Eucalyptus Leaves. *Remote Sens. Environ.* **1998**, *66*, 111–121. [\[CrossRef\]](#)
82. Sun, J.; Yang, L.; Yang, X.; Wei, J.; Li, L.; Guo, E.; Kong, Y. Using Spectral Reflectance to Estimate the Leaf Chlorophyll Content of Maize Inoculated with Arbuscular Mycorrhizal Fungi Under Water Stress. *Front. Plant Sci.* **2021**, *12*, 646173. [\[CrossRef\]](#)
83. Hatfield, P.L.; Pinter, P.J., Jr. Remote sensing for crop protection. *Crop Prot.* **1993**, *12*, 403–413. [\[CrossRef\]](#)
84. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [\[CrossRef\]](#)
85. Zhao, Y.-R.; Li, X.; Yu, K.-Q.; Cheng, F.; He, Y. Hyperspectral Imaging for Determining Pigment Contents in Cucumber Leaves in Response to Angular Leaf Spot Disease. *Sci. Rep.* **2016**, *6*, 27790. [\[CrossRef\]](#)
86. Sonobe, R.; Wang, Q. Nondestructive assessments of carotenoids content of broadleaved plant species using hyperspectral indices. *Comput. Electron. Agric.* **2018**, *145*, 18–26. [\[CrossRef\]](#)
87. Carter, G.A.; Spiering, B.A. Optical properties of intact leaves for estimating chlorophyll concentration. *J. Environ. Qual.* **2002**, *31*, 1424–1432. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Ahmed, M.R.; Yasmin, J.; Mo, C.; Lee, H.; Kim, M.S.; Hong, S.J.; Cho, B.K. Outdoor applications of hyperspectral imaging technology for monitoring agricultural crops: A review. *J. Biosyst. Eng.* **2016**, *41*, 396–407. [\[CrossRef\]](#)
89. Jacquemoud, S.; Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* **1990**, *34*, 75–91. [\[CrossRef\]](#)
90. Croft, H.; Chen, J.M. Leaf pigment content. In *Comprehensive Remote Sensing*; Liang, S., Ed.; Elsevier: Oxford, UK, 2018; pp. 117–142.
91. Slaton, M.R.; Hunt, E.R., Jr.; Smith, W.K. Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am. J. Bot.* **2001**, *88*, 278–284. [\[CrossRef\]](#)
92. Zovko, M.; Žibrat, U.; Knapič, M.; Kovačić, M.B.; Romić, D. Hyperspectral remote sensing of grapevine drought stress. *Precis. Agric.* **2019**, *20*, 335–347. [\[CrossRef\]](#)
93. Xu, H.R.; Ying, Y.B.; Fu, X.P.; Zhu, S.P. Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. *Biosyst. Eng.* **2007**, *96*, 447–454. [\[CrossRef\]](#)
94. Tian, Q.J.; Tong, Q.X.; Pu, R.L.; Guo, X.; Zhao, C. Spectroscopic determination of wheat water status using 1650–1850 nm spectral absorption features. *Int. J. Remote Sens.* **2001**, *22*, 2329–2338. [\[CrossRef\]](#)
95. Hunt, E.R., Jr.; Yilmaz, M.T. Remote sensing of vegetation water content using shortwave infrared reflectances. In *Remote Sensing and Modeling of Ecosystems for Sustainability IV*; Gao, W., Ustin, S.L., Eds.; SPIE: Bellingham, WA, USA, 2007; Volume 6679, p. 667902.
96. Hunt, E.R., Jr.; Ustin, S.L.; Riaño, D. Remote sensing of leaf, canopy, and vegetation water contents for satellite environmental data records. In *Satellite-Based Applications on Climate Change*; Qu, J.J., Powell, A.M., Sivakumar, M.V.K., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 335–357.
97. Behmann, J.; Steinrücken, J.; Plümer, L. Detection of early plant stress responses in hyperspectral images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 98–111. [\[CrossRef\]](#)
98. Junttila, S.; Hölttä, T.; Saarinen, N.; Kankare, V.; Yrttimaa, T.; Hyypä, J.; Vastaranta, M. Close-range hyperspectral spectroscopy reveals leaf water content dynamics. *Remote Sens. Environ.* **2022**, *277*, 113071. [\[CrossRef\]](#)
99. Ashraf, M. Inducing drought tolerance in plants: Recent advances. *Biotechnol. Adv.* **2010**, *28*, 169–183. [\[CrossRef\]](#) [\[PubMed\]](#)

100. Zúñiga Espinoza, C.; Khot, L.R.; Sankaran, S.; Jacoby, P.W. High-resolution multispectral and thermal remote sensing-based water stress assessment in subsurface irrigated grapevines. *Remote Sens.* **2017**, *9*, 961. [\[CrossRef\]](#)
101. Sandmann, M.; Grosch, R.; Graefe, J. The use of features from fluorescence, thermography, and NDVI imaging to detect biotic stress in lettuce. *Plant Dis.* **2018**, *102*, 1101–1107. [\[CrossRef\]](#)
102. Thapa, S.; Rudd, J.C.; Xue, Q.; Bhandari, M.; Reddy, S.K.; Jessup, K.E.; Liu, S.; Devkota, R.N.; Baker, J.; Baker, S. Use of NDVI for characterizing winter wheat response to water stress in a semi-arid environment. *J. Crop Improv.* **2019**, *33*, 633–648. [\[CrossRef\]](#)
103. Pipatsitee, P.; Tisarum, R.; Taota, K.; Samphumphuang, T.; Eiumnoh, A.; Singh, H.P.; Cha-Um, S. Effectiveness of vegetation indices and UAV-multispectral imageries in assessing the response of hybrid maize (*Zea mays* L.) to water deficit stress under field environment. *Environ. Monit. Assess.* **2023**, *195*, 128. [\[CrossRef\]](#)
104. Shafri, H.Z.; Hamdan, N. Hyperspectral imagery for mapping disease infection in oil palm plantation using vegetation indices and red edge techniques. *Am. J. Appl. Sci.* **2009**, *6*, 1031–1035. [\[CrossRef\]](#)
105. Shi, Y.; Huang, W.; Luo, J.; Huang, L.; Zhou, X. Detection and discrimination of pests and diseases in winter wheat based on spectral indices and kernel discriminant analysis. *Comput. Electron. Agric.* **2017**, *141*, 171–180. [\[CrossRef\]](#)
106. Velusamy, P.; Rajendran, S.; Mahendran, R.K.; Naseer, S.; Shafiq, M.; Choi, J.G. Unmanned aerial vehicles (UAV) in precision agriculture: Applications and challenges. *Energies* **2021**, *15*, 217. [\[CrossRef\]](#)
107. Császár, O.; Tóth, F.; Lajos, K. Estimation of the expected maximal defoliation and yield loss caused by cereal leaf beetle (*Oulema melanopus* L.) larvae in winter wheat (*Triticum aestivum* L.). *Crop Prot.* **2021**, *145*, 105644. [\[CrossRef\]](#)
108. Reay-Jones, F.P. Spatial distribution of the cereal leaf beetle (Coleoptera: Chrysomelidae) in wheat. *Environ. Entomol.* **2010**, *39*, 1943–1952. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Zanin, A.R.A.; Neves, D.C.; Teodoro, L.P.R.; da Silva Júnior, C.A.; da Silva, S.P.; Teodoro, P.E.; Baio, F.H.R. Reduction of pesticide application via real-time precision spraying. *Sci. Rep.* **2022**, *12*, 5638. [\[CrossRef\]](#) [\[PubMed\]](#)
110. Shi, Y.; Huang, W.; Ye, H.; Ruan, C.; Xing, N.; Geng, Y.; Peng, D. Partial least square discriminant analysis based on normalized two-stage vegetation indices for mapping damage from rice diseases using PlanetScope datasets. *Sensors* **2018**, *18*, 1901. [\[CrossRef\]](#) [\[PubMed\]](#)
111. Das, D.K.; Pradhan, S.; Sehgal, V.K.; Sahoo, R.N.; Gupta, V.K.; Singh, R. Spectral Reflectance Characteristics of Healthy and Yellow Mosaic Virus Infected Soybean (*Glycine max* L.) Leaves in a Semiarid Environment. *J. Agrometeorol.* **2013**, *15*, 36–42. [\[CrossRef\]](#)
112. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
113. Nguyen, H.D.; Nansen, C. Hyperspectral remote sensing to detect leafminer-induced stress in bok choy and spinach according to fertilizer regime and timing. *Pest Manag. Sci.* **2020**, *76*, 2208–2216. [\[CrossRef\]](#)
114. Susič, N.; Žibrat, U.; Širca, S.; Strajnar, P.; Razinger, J.; Knapič, M.; Vončina, A.; Urek, G.; Gerič Stare, B. Discrimination between Abiotic and Biotic Drought Stress in Tomatoes Using Hyperspectral Imaging. *Sens. Actuators B Chem.* **2018**, *273*, 842–852. [\[CrossRef\]](#)
115. Moshou, D.; Bravo, C.; West, J.; Wahlen, S.; McCartney, A.; Ramon, H. Automatic detection of “yellow rust” in wheat using reflectance measurements and neural networks. *Comput. Electron. Agric.* **2004**, *44*, 173–188. [\[CrossRef\]](#)
116. Couture, J.J.; Singh, A.; Charkowski, A.O.; Groves, R.L.; Gray, S.M.; Bethke, P.C.; Townsend, P.A. Integrating spectroscopy with potato disease management. *Plant Dis.* **2018**, *102*, 2233–2240. [\[CrossRef\]](#)
117. Negrisola, M.M.; Negrisola, R.; da Silva, F.; Lopes, L.S.; Souza Júnior, F.S.D.; Velini, E.D.; Raetano, C.G. Soybean rust detection and disease severity classification by remote sensing. *Agron. J.* **2022**, *114*, 3246–3262. [\[CrossRef\]](#)
118. Nansen, C.; Murdock, M.; Purington, R.; Marshall, S. Early infestations by arthropod pests induce unique changes in plant compositional traits and leaf reflectance. *Pest Manag. Sci.* **2021**, *77*, 5158–5169. [\[CrossRef\]](#)
119. Segarra, J.; Araus, J.L.; Kefauver, S.C. Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102697. [\[CrossRef\]](#)
120. Kamenova, I.; Dimitrov, P. Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCOVER of winter wheat in Bulgaria. *Eur. J. Remote Sens.* **2020**, *54* (Suppl. S1), 89–108. [\[CrossRef\]](#)
121. Feng, F.; Li, W.; Du, Q.; Zhang, B. Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity. *Remote Sens.* **2017**, *9*, 323. [\[CrossRef\]](#)
122. Ou, Y.; Yan, J.; Liang, Z.; Zhang, B. Hyperspectral Imaging Combined with Deep Learning for the Early Detection of Strawberry Leaf Gray Mold Disease. *Agronomy* **2024**, *14*, 2694. [\[CrossRef\]](#)
123. Tomaszewski, M.; Nalepa, J.; Moliszewska, E.; Ruszczak, B.; Smykała, K. Early Detection of Solanum Lycopersicum Diseases from Temporally-Aggregated Hyperspectral Measurements Using Machine Learning. *Sci. Rep.* **2023**, *13*, 7671. [\[CrossRef\]](#)
124. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [\[CrossRef\]](#)
125. Gitelson, A.; Merzlyak, M.N. Quantitative Estimation of Chlorophyll-*a* Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252. [\[CrossRef\]](#)

126. Sun, T.; Rao, S.; Zhou, X.; Li, L. Plant carotenoids: Recent advances and future perspectives. *Mol. Hortic.* **2022**, *2*, 3. [[CrossRef](#)]
127. Wu, B.; Huang, W.; Ye, H.; Luo, P.; Ren, Y.; Kong, W. Using Multi-Angular Hyperspectral Data to Estimate the Vertical Distribution of Leaf Chlorophyll Content in Wheat. *Remote Sens.* **2021**, *13*, 1501. [[CrossRef](#)]
128. Darvishzadeh, R.; Skidmore, A.; Atzberger, C.; van Wieren, S. Estimation of Vegetation LAI from Hyperspectral Reflectance Data: Effects of Soil Type and Plant Architecture. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 358–373. [[CrossRef](#)]
129. Gara, T.W.; Skidmore, A.K.; Darvishzadeh, R.; Wang, T. Leaf to canopy upscaling approach affects the estimation of canopy traits. *GISci. Remote Sens.* **2019**, *56*, 554–575. [[CrossRef](#)]
130. Kumar, P.; Gill, H.S.; Singh, M.; Kaur, K.; Koupal, D.; Talukder, S.; Bernardo, A.; St Amand, P.; Bai, G.; Sehgal, S.K. Characterization of Flag Leaf Morphology Identifies a Major Genomic Region Controlling Flag Leaf Angle in the U.S. Winter Wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **2024**, *137*, 47. [[CrossRef](#)]
131. Verrelst, J.; Malenovsky, Z.; Van der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.-P.; Lewis, P.; North, P.; Moreno, J. Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. *Surv. Geophys.* **2019**, *40*, 589–629. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.