

Research article

<https://doi.org/10.31489/2026FEB3/137-151>

UDC 57.08:581:004.8

Received: 12.12.2025 | Accepted: 18.02.2026 | Published online: 30.09.2026

R.M. Ualiyeva¹, M.M. Kaverina^{2*}, A.V. Osipova³, A.E. Koppaev⁴,
M.A. Chaizabekova⁵, S.B. Zhangazin⁶

^{1, 2, 3, 4, 5}Toraighyrov University, Pavlodar, Kazakhstan;

⁶L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

*Corresponding author: k.ma96@mail.ru

Species identification of crops and weeds in wheat fields using hyperspectral data and machine learning algorithms

Based on the relationship established between the morphological and anatomical properties of plant tissues and the recognition performance of a machine learning algorithm, a comprehensive assessment of the limitations of ML algorithms, using Random Forest as an example, was conducted for the first time for the multiclass identification of crops and weeds in the agroecosystems of northeastern Kazakhstan. The study identifies the principal biological factors underlying the optical overlap of classes in the hyperspectral feature space. Data were collected using a hyperspectral sensor operating in the VIS–NIR range. Spectral profiles were analyzed by assessing variance distribution using principal component analysis and constructing reflectance curves. Classification of the extracted data was performed using the Random Forest algorithm. Spring wheat (*Triticum aestivum*) at different phenological stages and five dominant weed species (*Euphorbia virgata*, *Descurainia sophia*, *Vicia hirsuta*, *Fallopia convolvulus*, and *Echinochloa crus-galli*) were selected as the study objects. The algorithm demonstrated an overall classification accuracy of 72.94% on the test data for crop-weed discrimination. High identification accuracy was achieved for wheat (MCC = 0.994), attributable to its spectral stability and pronounced red-edge features (680–730 nm). Lower accuracy in the intra-class separation of weed species was associated with species-specific anatomical characteristics: the thick cuticle of *Euphorbia virgata* increased reflectance, the complex structure of *Vicia hirsuta* contributed to spectral dispersion, and similarities in mesophyll structure between *Echinochloa crus-galli* and other grasses resulted in optical overlap with crop features. Thus, the model enables reliable crop–weed discrimination, which is important for precision agriculture and targeted agrochemical application.

Keywords: hyperspectral sensing, spring wheat, weeds, machine learning, species identification, species differentiation, spectral signatures, precision agriculture.

Introduction

Wheat is an important crop that ensures food security [1], accounting for approximately one-fifth of the total current global food demand [2]. However, despite the global prevalence of its cultivation areas, these regions are characterised by a number of problems that impede the implementation of initiatives in this direction [1]. One of the key destructive factors affecting cultivation and leading to substantial economic losses is the infestation of crop stands with weeds [3], which is caused by the delay and slowdown of crop growth, leading to reduced productivity and harvest quality [4] due to competition with weeds for sunlight, essential nutrients, territory, and water resources [5, 6]. According to FAO data, the share of annual crop losses due to weeds amounts to approximately 10 % [7, 8]. Traditional control methods, based on the blanket application of herbicides to field areas infested with weed vegetation, are associated with high labour intensity, increased financial costs, soil degradation, and negative environmental impacts [4, 9].

In the context of precision agriculture, the SSWM (Site-Specific Weed Management) strategy is gaining relevance [3, 10, 11], involving the targeted application of protective agents only in locations where specific species of undesirable plants (weeds) are detected [12], which allows for a manifold reduction of the pesticide load on ecosystems [13]. The basis of this strategy is the automated and accurate identification of weeds against the background of crop vegetation [14], which remains a challenging task to this day due to the morphological similarity of plants, especially at early developmental stages.

Standard digital RGB cameras have significant limitations when classifying species with similar visual characteristics. A promising alternative to them is hyperspectral systems, which, unlike traditional RGB and

multispectral imaging [14], record the reflectance of objects in hundreds of narrow spectral bands, thereby enabling the acquisition of detailed, unique spectral signatures of plants that reflect not only external morphological characteristics but also structural features of tissues, as well as internal chemical composition (chlorophyll, anthocyanin, carotenoid content, etc.) [15]. This makes it possible to distinguish species of crops and weeds based on species identification models, detect stress conditions, and conduct phytosanitary monitoring of crops at various stages of plant development [16].

One of the key research directions is the identification of weed vegetation in wheat crops [17]. However, despite the capabilities of hyperspectral sensors, the spectral information extracted in the form of hypercubes is characterised by high redundancy and dimensionality, which requires the application of comprehensive raw data processing methods. Machine learning and deep learning methods have become the primary tools for analysing such data arrays. The application of hyperspectral images in conjunction with machine learning algorithms enables the effective differentiation of several weed species using PLS-DA, SVM, and multilayer neural networks (MLP), with the MLP model demonstrating the best results with classification accuracy reaching nearly 100 % [18]. The combination of UAV imaging and convolutional neural networks allows for the accurate localisation of infestation hotspots and the assessment of the impact of weeds on wheat yield and biomass [4]. This is based on the difference in spectral signatures of leaves of crop and weed plants [19]. Traditional classifiers such as SVM, RF, and LDA [3, 20] demonstrate high performance in classification tasks when appropriate preprocessing parameters and feature sets are selected [21]. The combination of UAV-based hyperspectral imaging and deep neural networks enables the prediction of wheat yield and plant phenotyping with high accuracy, while the application of CNNs [3, 14, 21], Transformer models, and other artificial intelligence algorithms significantly enhances the efficiency of hyperspectral cube processing and the automation of crop analysis [22].

Despite significant advances in this field, there is still insufficient data on the application of hyperspectral imaging for the identification of weed vegetation and its differentiation from cereal crops, with this issue being particularly relevant for the regions of Central Asia specialising in grain cultivation. In this regard, the aim of this study is to investigate the application of hyperspectral imaging technology, as well as the possibilities of integrating hyperspectral sensing and machine learning for the species-level identification of crop and weed vegetation. The results of this study may help reduce the environmental burden and increase the efficiency of agroecosystem monitoring through more precise application of plant protection products, aligning with SDG 2 “Zero Hunger”, particularly the tasks of implementing innovative agricultural technologies and ensuring sustainable agriculture.

Experimental

The objects of the study were samples of crop and weed vegetation. The study crop was spring wheat (*Triticum aestivum* L., 1753). The weed vegetation was represented by the following five dominant species: *Euphorbia virgata* Waldst. & Kit., 1803; *Descurainia sophia* (L.) Webb ex Prantl, 1887 (syn. *Sisymbrium sophia* L., 1753); *Vicia hirsuta* (L.) Gray, 1821 (syn. *Ervum hirsutum* L., 1753); *Fallopia convolvulus* (L.) Á. Löve, 1970 (syn. *Polygonum convolvulus* L., 1753); *Echinochloa crus-galli* (L.) P. Beauv., 1812 (syn. *Panicum crus-galli* L., 1753). Sampling of both crop and weed vegetation was conducted in the north-east of Pavlodar Region, Republic of Kazakhstan, during the main developmental stages of spring wheat according to the BBCH scale: 10–20, 25–35, 39–59, and 61–75. The weather and climatic conditions of the study areas were characterised by a sharply continental climate with an average daily temperature during the growing season of +17.5 °C, average precipitation levels ranging from 40 to 48 mm, NE and SW winds with a speed of about 4 m/s, dark chestnut and chestnut medium-loamy soils with a calcareous composition and alkaline reaction (average pH 7.5).

For the analysis and subsequent imaging, 120 plant samples were selected: *Triticum aestivum* — 20, *Euphorbia virgata* — 20, *Descurainia sophia* — 20, *Vicia hirsuta* — 20, *Fallopia convolvulus* — 20, *Echinochloa crus-galli* — 20. The collected samples were transported from the fields to the Laboratory of Biological Research at Toraighyrov University (Pavlodar, Kazakhstan) for spectral studies. The acquisition, preprocessing, and processing of raw spectral data were performed using a FigSpec FS-13 hyperspectral camera (CHNSpec Technology, Hangzhou, China) with a CMOS detector and Breeze software (Umeå, Sweden). Hyperspectral imaging was conducted in the visible and NIR range with a spatial step of 2.5 nm at 1.92 thousand pixels, each with a size of 5.86 µm. The raw three-dimensional hypercubes, for the extraction of informative data, were preliminarily subjected to reference calibration (Spectralon, 99 %; Labsphere, Inc., North Sutton, NH, USA), noise elimination, smoothing, normalisation, and centering. Multidimensional lay-

er-by-layer visualisation and spatial dimensionality reduction were performed based on a PCA model with the extraction of principal components using Variance Scatter and the spectral representation of pixels from key informative regions via Raw Spectrum. PC1 and PC2 were used for the analysis of feature variability (t[1] reflects the main spectral differences between features, while t[2] determines additional variable characteristics). The formation of labeling in the dataset for further training and the selection of working pixels were carried out manually, followed by the segmentation of plant samples into distinct regions. The extracted sample regions, totaling 538, were used to develop a working model based on the Random Forest machine learning algorithm, and were divided into two main sets: the training set included 358 sample regions—66.5 % of the entire dataset; the test set comprised 180 sample regions—33.5 % of the dataset. The training of the RF-based model was performed using cross-validation. The workflow of the algorithm, from the import of raw data to training, is presented in Figure 1.

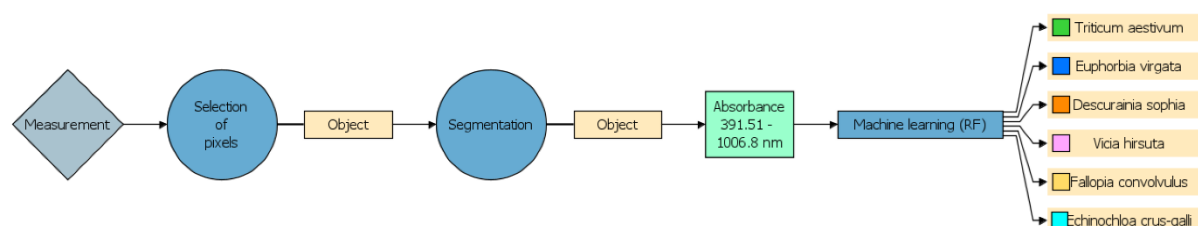


Figure 1. Algorithm workflow from raw data import to the application of machine learning

The extracted spectral reflectance data were processed using descriptive statistics and analysis of variance (ANOVA), with statistical significance set at $p < 0.05$. The obtained indicators for species identification and differentiation of weed and crop vegetation are presented in the format of a Confusion Matrix. The assessment of classification quality was conducted using metrics such as Precision, Recall, and F1-score, while Accuracy and MCC were calculated to evaluate the performance of the model and the obtained results, allowing for the determination of the machine learning model's operability and its applicability to individual species of the studied weed and crop vegetation.

Results and Discussion

Spectral analysis

The figures present hyperspectral images of wheat and weed vegetation of agroecosystems, Raw Spectrum plots with curves reflecting the spectral characteristics of plant regions, and Variance Scatter diagrams. The hyperspectral image is rendered in pseudo-colours that correspond to the level of light reflectance intensity. Warm colours (red, yellow) denote areas with high reflectance, while cool colours (cyan, blue) represent zones of low reflectance. The Raw Spectrum plot contains curves reflecting the spectral characteristics of a specific part of the hyperspectral image. The plot represents a basic presentation of spectral data, showing the dependence of the object's reflectance coefficient on wavelength (400–1000 nm), where this indicator is a key source of information about the physiological state of plant tissues. The Variance Scatter diagrams reflect the distribution of objects in the principal component space.

Figure 2 presents hyperspectral images, Raw Spectrum plots, and a Variance Scatter diagram of a young wheat plant (Sample 1) and a plant at the heading stage (Sample 2). The spectral curves of the Raw Spectrum plot for Sample 1 are in the range of 500–780 nm. It was established that the most intense curves (purple and cyan), corresponding to the brightest areas of the hyperspectral image (coleoptile and spike), reach a reflectance coefficient of about 50–53 % in the wavelength region of 620–650 nm. The minimum values are characteristic of the red curve, whose reflectance coefficient does not exceed 10–12 %. For the green and red curves, a pronounced red edge region and a sharp increase in reflectance between 680 and 730 nm are recorded after 680 nm, driven by the transition from strong chlorophyll absorption to high reflectance in the near-infrared region. Along with this, it was revealed that after 730–780 nm, the reflectance coefficient gradually decreases. For the wheat spike (Sample 2), the spectral curves have a similar shape but are distinguished by lower reflectance amplitude. The maximum reflectance coefficient values of the most intense areas reach 37.5 % in the 600–650 nm range. The red curve, corresponding to regions with minimum reflectance, is characterised by a coefficient of about 7.5 % at the first peak and 20 % at the second peak. At the same time, the red edge in the 680–730 nm interval is observed for all curves of the plot; however, the

curves of the spikelet glumes have lower reflectance within the second peak. The Variance Scatter diagrams reflect the distribution of spectral variability of the data in the principal component space. Accordingly, it was revealed that for the young wheat plant, the first component $t[1]$ explains 64.5 % of the total variance, and the second component $t[2]$ explains 14.6 %, which may indicate high spectral heterogeneity of tissues among the leaf, coleoptile, and root. For the wheat spike, the first principal component already explains 82.2 % of the total variance, and the second component only 5.21 %. The more compact distribution of points and the dominance of the first component indicate a more homogeneous spectral structure of the spike compared to the young plant.

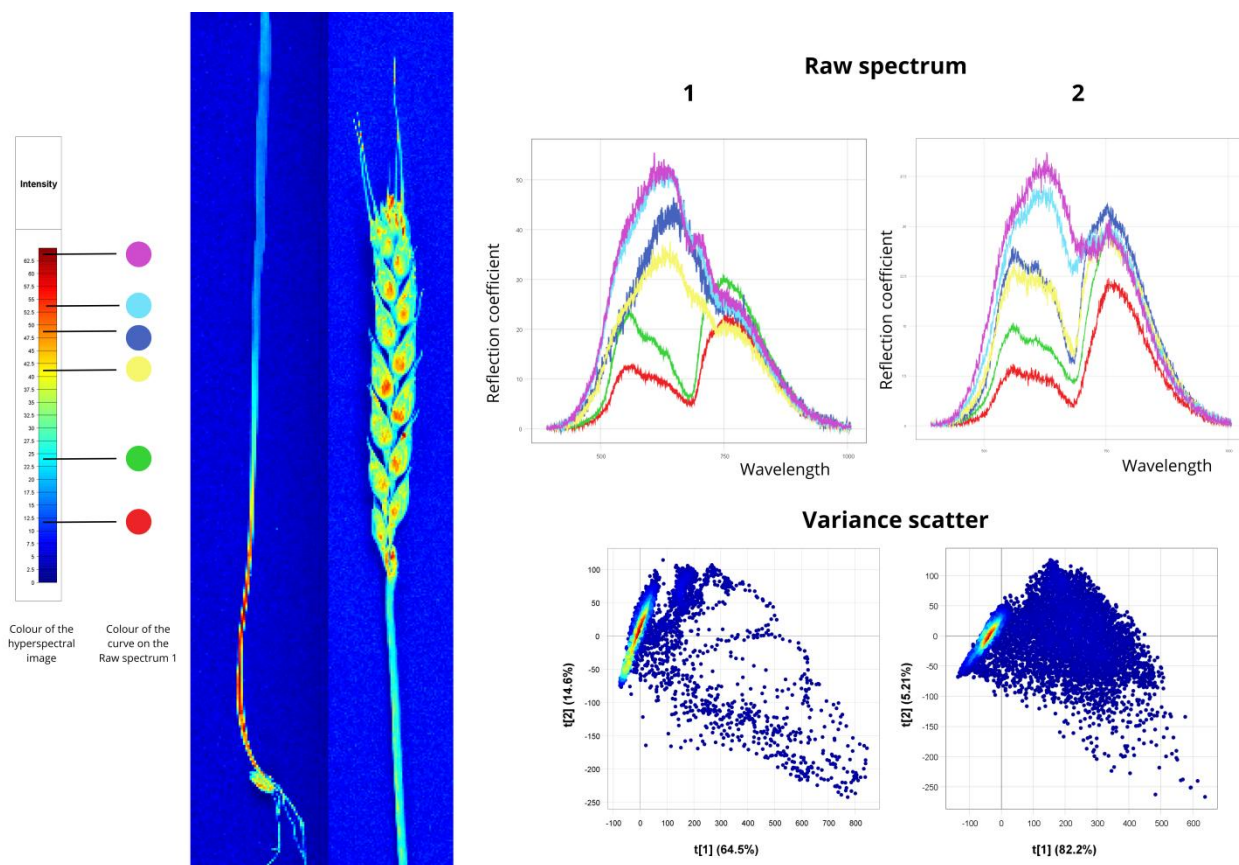


Figure 2. Wheat spectral profile

Figure 3 presents the weed vegetation and Variance Scatter diagrams, and Figure 4 presents the Raw Spectrum plots.

The Raw Spectrum plots reflect the spectral characteristics of different parts of the weed plants presented in the hyperspectral images. The spectral curves of Sample 1 demonstrate the highest reflectance values among all plants. The maximum reflectance coefficient of the purple curve reaches 75–77 % in the 610–650 nm range, which corresponds to areas of the root zone. The cyan and blue spectra have reflectance coefficients corresponding to 60–67 %. The yellow curve has its first peak recorded in the wavelength region of 500–730 nm, with a reflectance coefficient of 45 %, and its second peak noted in the region of 730–750 nm with a reflectance coefficient within 30–35 %, also belonging to the root part. The green curve reflects the leaf veins and stem of the weed and has a reflectance coefficient of 28 %. The red spectrum displays the spectral characteristics of the leaves with a reflectance coefficient not exceeding 14–15 %. For Sample 2, the maximum reflectance of the purple curve reaches 25–35 %, and the reflectance peak is noted in the wavelength region of about 620–650 nm, which corresponds to the root zone of the weed. The cyan, blue, and yellow curves are distinguished by a reflectance coefficient in the range of 20–27 % and belong to the stem and leaves. The red curve has a reflectance coefficient of 7 % and corresponds to individual shaded areas of the weed leaves. In the green spectral region, reflectance reaches up to 10 %, followed by a sharp increase to 25 % after 700 nm. The presence of the red edge in the region of 700 nm was also revealed for all spectral curves except the purple one.

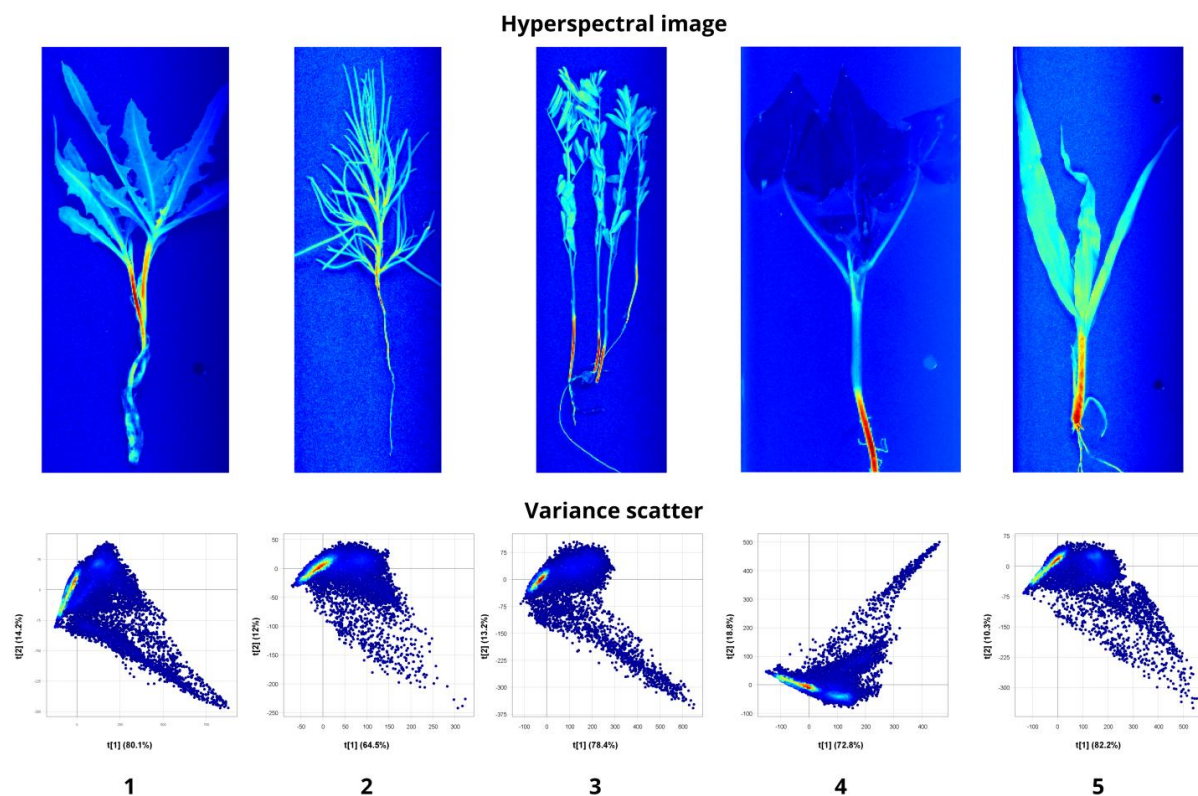


Figure 3. Hyperspectral images and Variance Scatter of weeds
(1 — *Euphorbia virgata*, 2 — *Descurainia sophia*, 3 — *Vicia hirsuta*,
4 — *Fallopia convolvulus*, 5 — *Echinochloa crus-galli*)

For Sample 3, the maximum value of the purple spectrum reaches 60 % in the region of 620–650 nm. The cyan, blue, and yellow curves have reflectance coefficients of 40–53 %, corresponding to root areas. The red and green spectra display the leaves and stems, having reflectance coefficients of 10–15 % at the first peak and 20–30 % at the second peak, and a clearly pronounced red edge is observed in the 680–730 nm interval.

For Sample 4, the maximum reflectance coefficient of the purple curve is 60 %, the cyan is 50 %, the blue corresponds to 30–35 %, and the yellow to 20–30 %, which relate to the root and near-root parts. The green spectrum has a reflectance coefficient of 5 %, and the red curve is characterised by a minimum value of 5 % in the visible range. In the 670–700 nm region, a deep reflectance minimum is observed, followed by a sharp red edge in the 700–740 nm range, with the reflectance coefficient reaching 25 % for the green and 15 % for the red curve, where the former spectrum belongs to stem areas and the latter to leaves.

All curves of Sample 5 have two peaks due to the formation of the red edge in the 700–740 nm range, after which reflectance gradually decreases in the near-infrared range beyond 780 nm. The first peak is in the 500–730 nm wavelength range, and the second peak is in the 730–780 nm region. The reflectance coefficient of the purple curve reaches 51 % at the first peak and 40 % at the second, the cyan 37.5 % and 35 %, the blue 30 % and 35 %, the yellow 27 % and 30 %, the green 15 % and 25 %, and the red 10 % and 15 %, respectively. The purple, cyan, and blue curves display the near-root part, the yellow and green curves display the leaves, and the red curve corresponds to individual shaded areas of the leaves.

The formation of a dense cluster of points in the region of low values of the second component and an elongated spectral tail, reflecting the heterogeneity of plant tissues and differences between leaves, stems, and root zones, is characteristic of all samples. The highest value of the second component is observed in *Vicia hirsuta* (18.8 %), indicating the heterogeneity of the plant's spectral characteristics. At the same time, in *Descurainia sophia*, the first component explains the lowest percentage of variance among all samples (64.5 %), and its variance diagram shows a wide distribution of points and a less pronounced cluster compactness. The fifth sample (*Echinochloa crus-galli*) exhibited the lowest value of the second component and the greatest dominance of the first (82.2 %), indicating the relative homogeneity of the spectral data of this weed species.

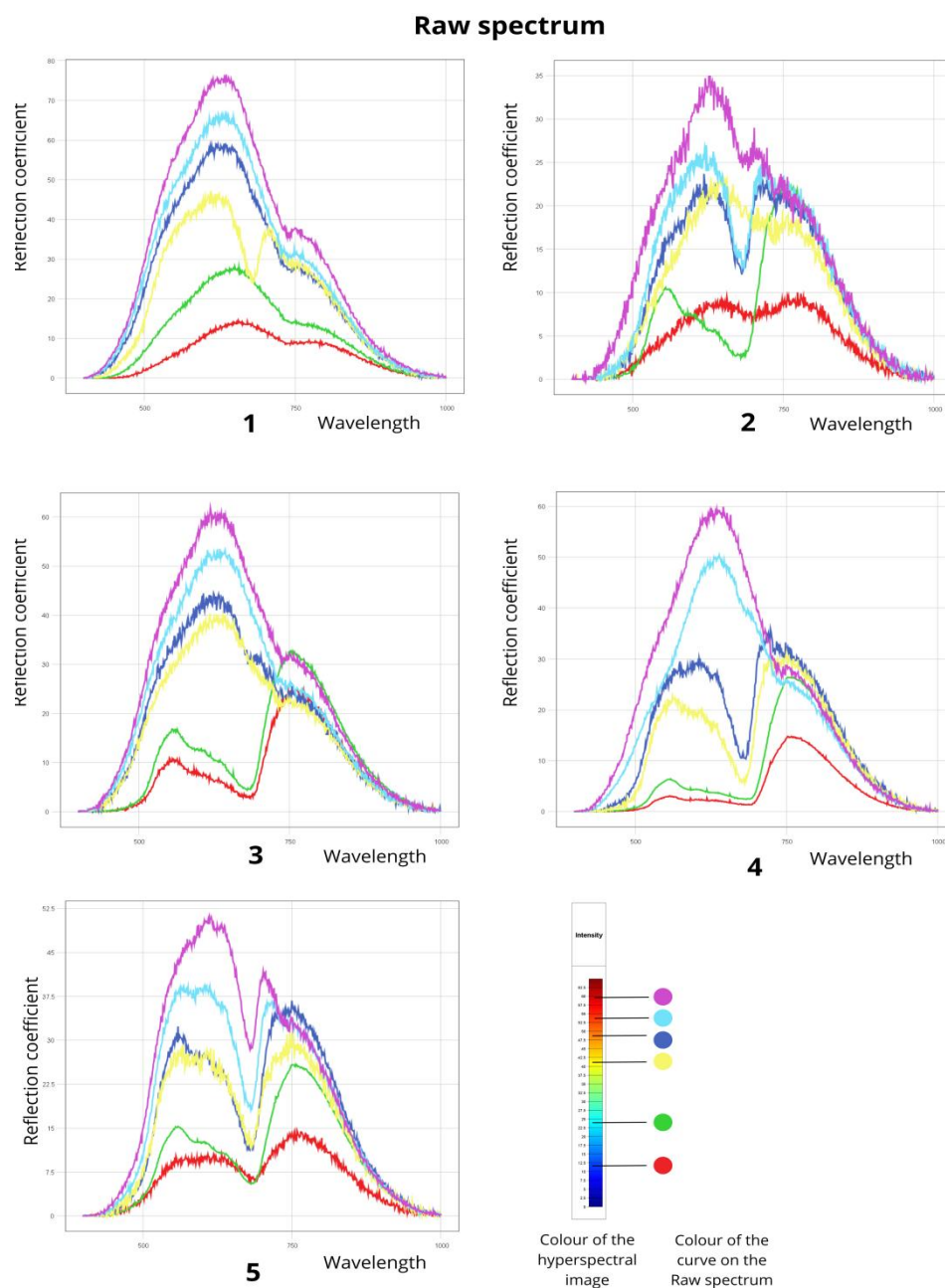


Figure 4. Raw spectra of weeds
(1 — *Euphorbia virgata*, 2 — *Descurainia sophia*, 3 — *Vicia hirsuta*,
4 — *Fallopia convolvulus*, 5 — *Echinochloa crus-galli*)

The obtained spectral differences between wheat and weed vegetation are driven by a combination of biophysical properties of plant tissues, including chlorophyll content, mesophyll anatomy, cuticle thickness, cell wall structure, the number of intercellular spaces, and the level of tissue water retention. The most significant differences were observed in the red-edge and NIR ranges, associated with the physiological state of plants and the structural features of leaf tissue. The spectral curves of the green parts of the studied objects demonstrate the reflectance pattern of photosynthetic vegetation, with low reflectance in the blue and red regions of the spectrum due to chlorophyll absorption, an increase in reflectance in the green region, and the formation of the red edge in the 680–730 nm range. The red edge is formed due to the combination of strong light absorption by chlorophyll in the red region of the spectrum and intense multiple scattering in the near-infrared range. The spectral curves of the root and near-root zones, due to the absence of photosynthetic activity and the presence of other pigments, do not exhibit a red-edge feature and reflect light throughout the wavelength range from 500 to 780 nm.

The higher reflectance values of the young wheat plant compared to the spike are explained by the anatomical features of the vegetative organs. The leaves and coleoptile of the young plant contain developed spongy parenchyma and a large amount of intercellular space, which enhances internal light scattering and leads to increased reflectance in the NIR region. In addition, young tissues are characterised by high chlorophyll content and an active photosynthetic apparatus, which contributes to the formation of a more pronounced red edge. In contrast, the tissues of the spike have a denser structure, a smaller proportion of intercellular spaces, and a higher content of mechanical tissues, which reduces the intensity of internal scattering and lowers the reflectance coefficient. Similar patterns have been described in studies of the spectral structure of cereal crops, where it has been shown that the transition from vegetative to generative organs is accompanied by a decrease in red edge amplitude and a reduction in reflectance in the near-infrared range due to changes in tissue structure and decreased chlorophyll content [23, 24]. The pronounced red edge in young wheat leaves and individual weed plants indicates a high content of photosynthetically active pigments and a developed internal cellular structure. According to research data, the position and steepness of the red edge are closely related to chlorophyll concentration, biomass, and plant water status. An increase in chlorophyll content causes a shift of the red edge toward longer wavelengths (red shift), whereas a decrease in physiological activity and water stress leads to a reduction in amplitude and a smoothing of the red edge [25–28].

The differences among weed species are also associated with their morphological and anatomical characteristics. *Euphorbia virgata* was characterised by the highest reflectance values, especially in the near-root zone. This may be explained by the presence of thick, succulent tissues, a developed system of vascular elements, and the presence of amyloplasts. Spurge contains a significant amount of milky sap and possesses a comparatively thick cuticle, which enhances internal light scattering and increases the reflectance properties of the tissues. The broad leaves of *Euphorbia virgata* form a more uniform reflectance surface compared to narrow-leaved species. *Descurainia sophia* exhibited the lowest values of the first principal component and high spectral heterogeneity. This species has highly dissected, filiform leaves, resulting in pronounced variability in reflection angles and shadow zones. The thin, linear leaves reduce the volume of multiple internal scattering, and the heterogeneous leaf orientation leads to a broadening of the point cloud on the Variance Scatter diagram. *Vicia hirsuta* is characterised by the highest value of the second component, indicating pronounced intra-tissue heterogeneity of spectral characteristics. This species has a complex, branching shoot structure and combines thin leaves and stems with varying degrees of lignification and moisture content. This leads to differences in reflectance between individual plant organs and the formation of a wide spectral dispersion. In *Fallopia convolvulus*, the deep minimum in the 670–700 nm region and the subsequent sharp increase in reflectance are associated with intense chlorophyll absorption and a strong contrast between the leaf blade and the vascular tissues of the stem. The presence of a pronounced red edge indicates the preservation of an active photosynthetic apparatus. *Echinochloa crus-galli* demonstrated the most homogeneous spectral structure. Its broad leaves and relatively uniform tissue anatomy provide similar light reflectance conditions over most of the plant surface.

The main difference between weed vegetation and wheat is the higher spectral variability and a wider range of reflectance. The young wheat plant was characterised by a relatively smooth and stable spectral curve shape with a maximum reflectance coefficient of about 50–53 %, whereas in some weed species, especially *Euphorbia virgata* and *Vicia hirsuta*, reflectance reached 60–77 %.

Analysis of statistical indicators

In the course of the study, the reflectance characteristics of spring wheat and five weed species at different stages of ontogenesis were analysed, as well as the spectral contrast characteristics within the VNIR range. The reflectance indicators clearly demonstrate variability among the studied plant species, as well as within a species depending on the phenological phase. Wheat exhibited a pronounced decrease in reflectance during the transition from the young plant stage to the heading stage with mean and maximum reflectance coefficient (RC) values within 31.50 % and 53 % for the first variant and 22.50 % and 37.50 % for the second, respectively. Among the studied weeds, the maximum reflectance values were found in *Euphorbia virgata*, which applies both to the mean reflectance (45.50 %) of the plant tissues of this species and to the maximum (77 %). These characteristics, along with a delta reflectance of 63 %, represented the highest values among all weed species and across all studied wheat growth stages. However, alongside this, certain groups of weeds were identified whose spectral characteristics are approximately similar in their manifestation to the crop vegetation during certain phenophases. The closest indicators in terms of value are the reflectance-absorption characteristics of *Echinochloa crus-galli*, with a mean RC value within 30.50 %, which clearly demonstrates the similarity of the spectral characteristics of this weed species to the young wheat

plant (mean RC=31.50 %). A similarity was also established for such spectral reflectance characteristics as rate of change and CV: for the young spring wheat plant—27.56 % and 5.46 %, and for the weed—26.28 % and 5.21 %, respectively. *Vicia hirsuta* (mean RC=35 %) has values that differ somewhat more but are not greatly distant from those of young wheat. On the other hand, the reflectance patterns of *Descurainia sophia* (21 %) and *Fallopia convolvulus* (26.50 %) are highly similar to those of heading wheat (22.50 %). At the same time, the highest coefficient of variation was recorded for the latter weed and was equal to 22.52 %, which is an indicator of high intraspecific heterogeneity of reflectance properties in this species (Tab. 1).

Table 1

Reflectance characteristics of wheat and weeds

№	Sample		Reflectance (%)		Mean	Standard deviation	CV	Rate of change	ΔReflectance
			min	max					
1	<i>Triticum aestivum</i>	Young plant	10	53	31.50	5.46	17.34	27.56	43,00
		Ear-bearing wheat plant	7.5	37.5	22.50	3.81	16.93	19.23	30,00
2	<i>Euphorbia virgata</i>		14	77	45.50	8.00	17.58	40.38	63.00
3	<i>Descurainia sophia</i>		7	35	21.00	3.56	16.93	17.95	28.00
4	<i>Vicia hirsuta</i>		10	60	35.00	6.35	18.14	32.05	50.00
5	<i>Fallopia convolvulus</i>		3	50	26.50	5.97	22.52	30.13	47.00
6	<i>Echinochloa crus-galli</i>		10	51	30.50	5.21	17.07	26.28	41.00

The assessment of the spectral characteristics of the wavelength range within VIS-NIR demonstrated the identity of macroparameters for the two groups of studied vegetation (“Wheat” and “Weeds”), corresponding to the wavelength interval of 500–780 nm. It was also observed that both groups demonstrated a spectral bandwidth of 280 nm, a contrast ratio of 1.56, a relative bandwidth of 0.22, and a normalised contrast of 0.36 (Tab. 2).

Table 2

Spectral contrast features across the visible-to-near-infrared spectral region

№	Samples	Wavelength (nm)		Spectral bandwidth	Contrast ratio	Relative bandwidth	Normalised contrast
		$\lambda_{\min}$	$\lambda_{\max}$				
1	Wheat	500	780	280	1.56	0.22	0.36
2	Weeds	500	780	280	1.56	0.22	0.36

The main key task in precision smart farming is the reliable and accurate differentiation of crop and weed vegetation for the subsequent targeted application of herbicides. The obtained data on the statistical indicators of the spectra made it possible to identify a number of features and limitations associated with the automated identification and differentiation of plant species. First and foremost, an obvious phenological dependence of differentiation was revealed. In accordance with the reflectance characteristic indicators, it was established that using only reflectance amplitude to create a single static “crop/weed” segregation model is not feasible. Due to the fact that the reflectance indicators within the species (wheat) drop by nearly a third (from 31.50 % to 22.50 %) by the time of heading, this indicates that the model’s efficiency for differentiation depends specifically on the phenological phase of monitoring. It was established that at the early stages of vegetation, young wheat can be easily distinguished, for example, from *Descurainia sophia* (mean difference of about 10 %), but in relation to *Echinochloa crus-galli*, a “spectral mimicry” effect may arise (mean RC difference = 1 %), which may be related to this plant’s belonging to the category of grass weeds. However, by the heading stage, wheat, according to its spectral profile (22.50 %), begins to overlap with *Descurainia sophia* (21 %) and may have similar spectral features to *Fallopia convolvulus* (26.50 %), while *Vicia hirsuta* (35 %) and *Echinochloa crus-galli* (30.50 %) contrast quite clearly with the crop. On the other hand, our data indicates that *Euphorbia virgata* can be consistently identified across all wheat development stages, owing to its anomalously high reflectance (45.50 %).

It was established that for interspecific identification, the use of standard broadband vegetation indices or multispectral cameras with wide spectral bands is ineffective due to the identical nature of their fundamental spectral contrast features. This is related to the similarity of the cellular structure of the leaves

and the chlorophyll content in the tissues of the studied plant samples, which influences the shape of the spectral curve.

Machine learning

The identified differences in hyperspectral characteristics between wheat and weed vegetation, as well as among individual weed species, formed the basis for constructing the classification model. The confusion matrix of the classification model demonstrates varying efficiency in recognising wheat and individual weed species based on hyperspectral features. The training model showed 84.9 % correctly classified samples. The model performed best with wheat—96.9 % correctly classified samples. The worst results were obtained for *Echinochloa crus-galli*, whose classification accuracy amounted to 68.4 % (Fig. 5, 6).

Actual classes	Total	Triticum aestivum	Euphorbia virgata	Descurainia sophia	Vicia hirsuta	Fallopia convolvulus	Echinochloa crus-...	No class
Triticum aestivum	87 (48.3%)	86 (98.9%)						1 (1.15%)
Euphorbia virgata	18 (10%)		4 (22.2%)		1 (5.56%)		2 (11.1%)	11 (61.1%)
Descurainia sophia	26 (14.4%)			22 (84.6%)			2 (7.69%)	2 (7.69%)
Vicia hirsuta	15 (8.33%)				8 (53.3%)		2 (13.3%)	5 (33.3%)
Fallopia convolvulus	16 (8.89%)		2 (12.5%)	4 (25%)		0		10 (62.5%)
Echinochloa crus-galli	18 (10%)		1 (5.56%)	1 (5.56%)	4 (22.2%)		3 (16.7%)	9 (50%)
# Predicted	180 (100%)	86 (47.8%)	7 (3.89%)	27 (15%)	13 (7.22%)	0	9 (5%)	38 (21.1%)
Correctly	123 (68.3%)							
Incorrectly	57 (31.7%)							
Precision		100%	57.1%	81.5%	61.5%	0%	33.3%	
Recall		98.9%	22.2%	84.6%	53.3%	0%	16.7%	
F-score		99.4%	32%	83%	57.1%	0%	22.2%	

A

Actual classes	Total	Triticum aestivum	Euphorbia virgata	Descurainia sophia	Vicia hirsuta	Fallopia convolvulus	Echinochloa crus-...	No class
Triticum aestivum	65 (18.2%)	63 (96.9%)						2 (3.08%)
Euphorbia virgata	48 (13.4%)		36 (75%)		2 (4.17%)			10 (20.8%)
Descurainia sophia	100 (27.9%)		1 (1%)	95 (95%)				4 (4%)
Vicia hirsuta	72 (20.1%)			1 (1.39%)	60 (83.3%)		1 (1.39%)	10 (13.9%)
Fallopia convolvulus	35 (9.78%)			2 (5.71%)		24 (68.6%)		9 (25.7%)
Echinochloa crus-galli	38 (10.6%)		2 (5.26%)	1 (2.63%)			26 (68.4%)	9 (23.7%)
# Predicted	358 (100%)	63 (17.6%)	39 (10.9%)	99 (27.7%)	62 (17.3%)	24 (6.7%)	27 (7.54%)	44 (12.3%)
Correctly	304 (84.9%)							
Incorrectly	54 (15.1%)							
Precision		100%	92.3%	96%	96.8%	100%	96.3%	
Recall		96.9%	75%	95%	83.3%	68.6%	68.4%	
F-score		98.4%	82.8%	95.5%	89.6%	81.4%	80%	

B

Figure 5. Confusion Matrix (A — Test, B — Train)

Sample	Group /	Image	Category	Machine learning (RF)	Machine learning (RF)
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
9.1.6. Пш...	Group		Triticum aestivum		Triticum aestivum
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		-
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.2. Euph...	Group		Euphorbia virgata		Vicia hirsuta
1.10. Ech...	Group		Echinochloa crus-galli		Vicia hirsuta
1.10. Ech...	Group		Echinochloa crus-galli		Vicia hirsuta
1.10. Ech...	Group		Echinochloa crus-galli		Vicia hirsuta
3.5. Desc...	Group		Descurainia sophia		-
3.5. Desc...	Group		Descurainia sophia		-
3.5. Desc...	Group		Descurainia sophia		Triticum aestivum
3.5. Desc...	Group		Descurainia sophia		-
3.5. Desc...	Group		Descurainia sophia		Vicia hirsuta
3.5. Desc...	Group		Descurainia sophia		Vicia hirsuta
3.5. Desc...	Group		Descurainia sophia		Descurainia sophia

Figure 6. Classification results

The overall classification result of the test model was 68.3 % (123 correctly classified samples out of 180), while 31.7 % of samples were assigned by the model to incorrect classes. The highest classification accuracy was achieved for wheat, where 86 out of 87 samples were correctly identified, corresponding to a recall of 98.9 %. The virtually complete absence of false-positive results indicates the high spectral uniqueness of wheat compared to weed vegetation. This is likely associated with the relatively homogeneous anatomical structure of the leaves and coleoptile, the characteristic red edge shape, and the stable reflectance distribution in the near-infrared range. The best results among the weed plants were obtained for *Descurainia sophia*. The model correctly classified 22 out of 26 samples of this species, corresponding to a recall of 84.6 % and a precision of 81.5 %. The high classification accuracy may be due to the plant’s specific morphology, with thin, filiform leaves and relatively low reflectance coefficients compared to other weeds. For *Vicia hirsuta*, the classification quality in the test set proved to be moderate: recall was 53.3 %, and precision — 61.5 %. The majority of samples were erroneously assigned to the “No class” category (33.3 %) and partially confused with *Echinochloa crus-galli*. *Vicia hirsuta* is characterised by a branching shoot structure and the presence of leaves with varying orientations, which leads to wide spectral dispersion. Low classification quality was observed for *Euphorbia virgata* and *Echinochloa crus-galli*. For *Euphorbia virgata*, recall reached only 22.2 %, despite relatively high reflectance coefficients. Most samples were erroneously classified as “No class” (61.1 %) or confused with *Echinochloa crus-galli*. A likely reason is the high variability of reflectance among the near-root zone, leaves, and stem of the plant. *Euphorbia virgata* has thick, succulent tissues, a developed system of vascular elements, and a heterogeneous distribution of moisture, which causes a significant change in spectral characteristics within a single plant. Because of a limited number of training examples, the model was unable to extract stable spectral features of this species. *Echinochloa crus-*

galli was also characterised by low recall (16.7 %) and precision (33.3 %). Some samples were erroneously assigned to *Vicia hirsuta* and the “No class” category. The reason may be the similarity of the spectral characteristics of *Echinochloa crus-galli* to wheat and other grass plants. As a representative of the *Poaceae* family, this species possesses linear leaves, similar mesophyll anatomy, and a close red edge position, which complicates class separation in multidimensional spectral space. For *Fallopia convolvulus*, the model failed to correctly classify any samples in the training data. Recall and precision were 0 %. The majority of objects were erroneously classified as “No class” or confused with *Descurainia sophia*. Such a result may be associated with the high spectral dissimilarity among different parts of the weed, the low reflectance intensity of the leaves, and an insufficient amount of training data. The high proportion of objects in the “No class” category (21.1 % of all predictions) indicates insufficient separability of certain spectral features and the presence of overlap between classes (Fig. 5).

According to the generalised statistical indicators of the algorithm (Tab. 3), the model demonstrates high potential in solving the basic task of classification and differentiation of crop and weed vegetation, which is confirmed by the Macro Accuracy value of 0.82214. In contrast, the decrease in this indicator on the test dataset (Macro Accuracy Test = 0.63123), the Log Loss metric (0.66778), and the cross-validation accuracy (CV Macro Accuracy = 0.72645) clearly reflect the natural complexity of working with the optical characteristics of biological objects. Such a change in the metric indicates spectral overlap between certain plant species within the “Weeds” categorical group, which, in turn, may be resolved in the future by expanding the database of spectral signatures and data for analysis, enabling more confident feature generalisation.

Table 3

Algorithm statistics

Algorithm name	Macro Accuracy	Log Loss	Log Loss Reduction	Micro Accuracy	Macro Accuracy Test	CV Macro Accuracy
Random Forest	0.82214	0.66778	0.61251	0.83544	0.63123	0.72645

Table 4 presents the accuracy and correlation metrics (MCC) for the algorithm used, confirming that the algorithm effectively copes with separating the crop plant from the weed background, with an average accuracy of 72.94 % in the test set.

Table 4

Quality assessment metrics for machine learning sample sets

Metrics	<i>Triticum aestivum</i>	<i>Euphorbia virgata</i>	<i>Descurainia sophia</i>	<i>Vicia hirsuta</i>	<i>Fallopia convolvulus</i>	<i>Echinochloa crus-galli</i>
Random Forest (Train)						
Binary Accuracy	96.92	75.00	95.00	83.33	68.57	68.42
Balanced Accuracy	98.45	87.47	97.46	91.63	84.30	84.19
MCC	0.984	0.831	0.954	0.897	0.827	0.811
Random Forest (Test)						
Binary Accuracy	98.85	22.22	84.62	53.33	0.00	16.67
Balanced Accuracy	99.45	61.07	92.25	76.60	50.00	58.29
MCC	0.994	0.355	0.829	0.572	-	0.234

Wheat demonstrated exceptional recognition stability, where binary accuracy rose from 96.92 % at the training stage to 98.85 % on the test set, and Balanced Accuracy from 98.45 % to 99.45 %. At the same time, the MCC value in the test set was determined to be 0.994, which confirms the uniqueness of the spectral characteristics of the crop, as well as the reliability of the model’s performance in its identification. Regarding the recognition of specific weeds, it was also established that the model effectively distinguishes species with uniquely characteristic morphology. In particular, *Descurainia sophia* deserves special attention, maintaining high classification performance on the test data (Binary Accuracy=84.62, MCC=0.829), which demonstrates the algorithm’s ability to capture and identify stable patterns of optical characteristics, particularly for fine-leaved plants. For the intra-class differentiation of a number of weed plants (*Euphorbia virgata*,

Vicia hirsuta, *Fallopia convolvulus*, *Echinochloa crus-galli*), classification proved to be a more challenging task, which was demonstrated by a decline in the metric indicators on the test set. However, this is substantiated by biological features. In particular, the signs of anatomical affinity of grass weeds (*Echinochloa crus-galli*) with the studied crop obviously create difficulties for separating these classes in the basic spectral space. Alongside this, a reason may be the high internal heterogeneity, forming a wide dispersion of spectral values (multidirectional leaf orientation in *Vicia hirsuta*, uneven moisture distribution in the dense tissues of *Euphorbia virgata*). It was found that these variability features were actually insufficient to form a single universal set of rules for recognising these species within the group (Tab. 4).

Thus, the results show that hyperspectral analysis possesses high efficiency for recognising wheat and certain weed species; however, classification accuracy depends on the spectral homogeneity of the plant and the degree of overlap of its spectral characteristics with those of other species. Species with unique morphology and a stable spectral profile are classified most successfully, whereas plants with complex shoot morphology or similar tissue anatomy exhibit a higher level of classification errors.

Conclusion

The conducted analysis of spectral characteristics and machine learning results allowed for drawing principal conclusions regarding the prospects for automated species-level identification of crop and weed vegetation. It was established that hyperspectral sensing, in combination with the Random Forest ML algorithm, proved the high validity and success of crop vegetation segregation from the weed background, which addresses the main task of precision agriculture. The spectral stability of spring wheat enables the model to identify it with high accuracy (MCC=0.994). A dependence of recognition on the anatomical specificity of the plant and the phenological phase was also revealed; in particular, the morphological features of *Descurainia sophia* form stable patterns, allowing the algorithm to perform accurate identification. However, the spectral proximity of green biomass across the broad VIS-NIR range (500–780 nm) requires the expansion of the spectral database for the application of standard metrics. The attempt to separate complex weed species from one another, such as grass weeds or species with high tissue heterogeneity, manifests in a decrease in the predictive ability of the model, which is an obvious consequence of their intraspecific variability and, as a result, spectral overlaps. Moreover, the application of the RF algorithm, with an overall accuracy of 72.94 % on the test set, formed a reliable basis for performing basic “crop-weed” classification. To transition toward high-precision multiclass identification and intra-group weed differentiation, it is essential to extract narrow-band spectral markers, balance complex training datasets, and integrate textural features, thereby compensating for optical overlaps. In conclusion, the developed classification model demonstrates potential for integration into site-specific herbicide application systems by solving the task of reliable crop preservation, thereby providing a foundation for further optimisation of machine learning algorithms in intra-class weed differentiation.

Funding

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26197815 “Development of an intelligent monitoring system for wheat agrocenosis through the synergistic integration of hyperspectral imaging and LiDAR using computer vision technologies”).

Conflict of Interest

Authors declare no conflict of interest.

Author contribution

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript: **Ualiyeva R.M.** — conceptualization, formal analysis, supervision, editing; **Kaverina M.M.** — methodology, investigation, analysis, writing draft; **Osipova A.V.** — methodology, investigation, analysis, writing draft; **Koppaev A.E.** — research, data collection; **Chaizabekova M.A.** — research, data collection; **Zhangazin S.B.** — data curation, analysis.

References

- 1 Kettlewell, P., Byrne, R., & Jeffery, S. (2023). Wheat area expansion into northern higher latitudes and global food security. *Agriculture, Ecosystems & Environment*, 351, 108499. <https://doi.org/10.1016/j.agee.2023.108499>
- 2 Gooding, M.J. (2023). Wheat. *ICC Handbook of 21st Century Cereal Science and Technology*, 121–130. Academic Press <https://doi.org/10.1016/B978-0-323-95295-8.00027-7>
- 3 Mensah, B., Rai, N., Betitame, K., & Sun, X. (2024). Advances in weed identification using hyperspectral imaging: A comprehensive review of platform sensors and deep learning techniques. *Journal of Agriculture and Food Research*, 18, 101388. <https://doi.org/10.1016/j.jafr.2024.101388>
- 4 Liu, T., Zhao, Y., Wang, H., Wu, W., Yang, T., Zhang, W., Zhu, S., Sun, C., & Yao, Z. (2024). Harnessing UAVs and deep learning for accurate grass weed detection in wheat fields: a study on biomass and yield implications. *Plant Methods*, 20. <https://doi.org/10.1186/s13007-024-01272-6>
- 5 Zhang, Y., Wang, M., Zhao, D., Liu, C., & Liu, Z. (2023). Early weed identification based on deep learning: A review. *Smart Agricultural Technology*, 3, 100123. <https://doi.org/10.1016/j.atech.2022.100123>
- 6 Che'Ya, N.N., Dunwoody, E., & Gupta, M. (2021). Assessment of Weed Classification Using Hyperspectral Reflectance and Optimal Multispectral UAV Imagery. *Agronomy*, 11(7), 1435. <https://doi.org/10.3390/agronomy11071435>
- 7 FAO (Food and Agriculture Organization of the United Nations). www.fao.org. Retrieved from <https://www.fao.org/home/en/>
- 8 Danilov, R., Kremneva, O., & Pachkin, A. (2023). Identification of the Spectral Patterns of Cultivated Plants and Weeds: Hyperspectral Vegetation Indices. *Agronomy*, 13(3), 859. <https://doi.org/10.3390/agronomy13030859>
- 9 Pantazi, X. -E., Moshou, D., & Bravo, C. (2016). Active learning system for weed species recognition based on hyperspectral sensing. *Biosystems Engineering*, 146, 193–202. <https://doi.org/10.1016/j.biosystemseng.2016.01.014>
- 10 Rai, N., Zhang, Y., Ram, B.G., Schumacher, L., Yellavajjala, R.K., Bajwa, S., & Sun, X. (2023). Applications of deep learning in precision weed management: A review. *Computers and Electronics in Agriculture*, 206, 107698. <https://doi.org/10.1016/j.compag.2023.107698>
- 11 Ronay, I., Lati, R.N., & Kizel, F. (2024). Weed Species Identification: Acquisition, Feature Analysis, and Evaluation of a Hyperspectral and RGB Dataset with Labeled Data. *Remote Sensing*, 16(15), 2808. <https://doi.org/10.3390/rs16152808>
- 12 Luo, T., Zhao, J., Gu, Y., Zhang, S., Qiao, X., Tian, W., & Han, Y. (2023). Classification of weed seeds based on visual images and deep learning. *Information Processing in Agriculture*, 10(1), 40–51. <https://doi.org/10.1016/j.inpa.2021.10.002>
- 13 Vasileiou, M., Kyrgiakos, L.S., Kleisiari, C., Kleftodimos, G., Vlontzos, G., Belhouchette, H., & Pardalos, P.M. (2024). Transforming weed management in sustainable agriculture with artificial intelligence: A systematic literature review towards weed identification and deep learning. *Crop Protection*, 176, 106522. <https://doi.org/10.1016/j.cropro.2023.106522>
- 14 Ram, B.G., Oduor, P., Igathinathane, C., Howatt, K., & Sun, X. (2024). A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*, 222, 109037. <https://doi.org/10.1016/j.compag.2024.109037>
- 15 Mishra, P., Lohumi, S., Khan, H.A., & Nordon, A. (2020). Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. *Computers and Electronics in Agriculture*, 178, 105780. <https://doi.org/10.1016/j.compag.2020.105780>
- 16 Kiani, S., & Jafari, A. (2012). Crop Detection and Positioning in the Field Using Discriminant Analysis and Neural Networks Based on Shape Features. *Journal of Agricultural Science and Technology*, 14, 755–765. <https://doi.org/10.1001/1.16807073.2012.14.4.10.6>
- 17 Gutjahr, C., & Gerhards, R. (2010). Decision rules for site-specific weed management. *Precision Crop Protection — the Challenge and Use of Heterogeneity*, 223–239. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-9277-9_14
- 18 Li, Y., Al-Sarayreh, M., Irie, K., Hackell, D., Bourdot, G., Reis, M.M., & Ghamkhar, K. (2021). Identification of Weeds Based on Hyperspectral Imaging and Machine Learning. *Frontiers in Plant Science*, 11, 611622. <https://doi.org/10.3389/fpls.2020.611622>
- 19 Smith, A.M., & Blackshaw, R.E. (2003). Weed-Crop Discrimination Using Remote Sensing: A Detached Leaf Experiment. *Weed Technology*, 17(4), 811–820. <https://doi.org/10.1614/WT02-179>
- 20 Liu, B., Li, R., Li, H., You, G., Yan, S., & Tong, Q. (2019). Crop/Weed Discrimination Using a Field Imaging Spectrometer System. *Sensors*, 19(23), 5154. <https://doi.org/10.3390/s19235154>
- 21 Wieme, J., Mollazade, K., Malounas, I., Zude-Sasse, M., Zhao, M., Gowen, A., Argyropoulos, D., Fountas, S., & Van Beek, J. (2022). Application of hyperspectral imaging systems and artificial intelligence for quality assessment of fruit, vegetables and mushrooms: A review. *J Wieme, Biosystems Engineering*, 222, 156–176. <https://doi.org/10.1016/j.biosystemseng.2022.07.013>
- 22 Moghimi, A., Yang, C., & Anderson, J.A. (2020). Aerial hyperspectral imagery and deep neural networks for high-throughput yield phenotyping in wheat. *Computers and Electronics in Agriculture*, 172, 105299. <https://doi.org/10.1016/j.compag.2020.105299>
- 23 Cheng, Y., He, W., Zheng, F., Zhang, F., Bai, B., Sun, N., Wang, W., & Geng, H. (2025). Genome-wide association study of wheat chlorophyll dynamics under drought and irrigation using multispectral UAV phenotyping. *Frontiers in Plant Science*, 16, 1607862. <https://doi.org/10.3389/fpls.2025.1607862>

- 24 Railyan, V.Ya., & Korobov, R.M. (1993). Red edge structure of canopy reflectance spectra of triticale. *Remote Sensing of Environment*, 46(2), 173–182. [https://doi.org/10.1016/0034-4257\(93\)90093-D](https://doi.org/10.1016/0034-4257(93)90093-D)
- 25 Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, 15(7), 1459–1470. <https://doi.org/10.1080/01431169408954177>
- 26 Horler, D.N.H., Dockray, M., Barber, J., & Barringer, A.R. (1983). Red edge measurements for remotely sensing plant chlorophyll content. *Advances in Space Research*, 3(2), 273–277. [https://doi.org/10.1016/0273-1177\(83\)90130-8](https://doi.org/10.1016/0273-1177(83)90130-8)
- 27 Mutanga, O., & Skidmore, A.K. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), 34–42. <https://doi.org/10.1016/j.isprsjprs.2007.02.001>
- 28 Clevers, J.G., & Gitelson, A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344–351. <https://doi.org/10.1016/j.jag.2012.10.008>

Р.М. Уалиева, М.М. Каверина, А.В. Осипова, А.Е. Коппаев,
М.А. Чайзабекова, С.Б. Жангазин

Машиналық оқыту алгоритмдерін пайдалану арқылы гиперспектралды деректер негізінде бидай егістіктеріндегі дақыл және арамшөп өсімдіктерін түрлік анықтау

Алғаш рет Қазақстанның солтүстік-шығысындағы агроценоздар жағдайында дақыл өсімдіктер мен арамшөптерді көпжылдық жіктеу кезінде машиналық оқыту алгоритмдерінің (Random Forest мысалында) шектеулеріне өсімдік тіндерінің морфологиялық-анатомиялық қасиеттері мен тану тиімділігі арасындағы өзара байланысты негізге ала отырып кешенді бағалау жүргізілді. Жұмыста гиперспектралды белгілер кеңістігінде кластардың оптикалық қабаттасуына әкелетін биологиялық табиғаттағы негізгі факторлар негізделді. Деректер VIS-NIR диапазонында гиперспектралды сенсор көмегімен жиналды. Спектрлік профильдерді талдау бас компоненттер әдісі арқылы дисперсия құрылымын бағалау және шағылысу спектрлерін құру негізінде жүзеге асырылды. Алынған деректерді жіктеу Random Forest алгоритмі арқылы орындалды. Зерттеу нысандары ретінде әртүрлі фенофазалардағы жаздық бидай (*Triticum aestivum*) егістіктері және арамшөптердің бес басым түрі (*Euphorbia virgata*, *Descurainia sophia*, *Vicia hirsuta*, *Fallopia convolvulus*, *Echinochloa crus-galli*) таңдалды. Алгоритм «дақыл-арамшөп» деңгейінде объектілерді тиімді ажыратып, тест деректерінде 72,94 % жалпы жіктеу дәлдігін көрсетті. Бидайды анықтау дәлдігі жоғары деңгейде (MCC=0.994) болып, бұл оның спектрлік тұрақтылығы мен red edge (680-730 нм) аймағындағы айқын белгілермен түсіндіріледі. Арамшөптерді түрлік деңгейде ажырату дәлдігінің төмендеуі олардың анатомиялық ерекшеліктерімен байланысты екені анықталды: *Euphorbia virgata*-ның қалың кутикулярлық қабаттары жоғары шағылысу мәндерін тудырады; *Vicia hirsuta*-ның күрделі морфологиялық құрылымы спектрлік ауытқуды күшейтеді; ал *Echinochloa crus-galli* мезофилияның дәнді дақылдарға ұқсас құрылымы кейбір спектрлік белгілердің оптикалық қабаттасуына әкеледі. Осылайша, әзірленген модель мәдени өсімдіктер мен арамшөптерді сенімді ажыратып, дәлме-дәл егіншілік технологиялары мен агрохимикаттарды максатты қолдану жүйелерін енгізуге негіз қалайды.

Кілт сөздер: гиперспектралды зондау, жаздық бидай, арамшөптер, машиналық оқыту, түрлік анықтау, түрлерді ажырату, спектрлік қолтаңбалар, нақты егіншілік.

Р.М. Уалиева, М.М. Каверина, А.В. Осипова, А.Е. Коппаев,
М.А. Чайзабекова, С.Б. Жангазин

Видовая идентификация культурной и сорной растительности в посевах пшеницы по гиперспектральным данным с использованием алгоритмов машинного обучения

Впервые проведена комплексная оценка ограничений ML-алгоритмов (на примере Random Forest) при мультиклассовой идентификации культурных растений и сорняков в агроценозах северо-востока Казахстана на основе установления взаимосвязи между морфолого-анатомическими свойствами растительных тканей и эффективностью распознавания алгоритмом машинного обучения. В работе обоснованы основные причины биологической природы, проявляющиеся в оптическом перекрытии классов в пространстве гиперспектральных признаков. Сбор данных проведен гиперспектральным сенсором в диапазоне VIS-NIR. Анализ спектральных профилей проведен на основе оценки распределения дисперсии методом главных компонент и построения кривых отражения. Классификация извлеченных данных выполнена на основе алгоритма Random Forest. В качестве объектов исследования выбраны

посевы яровой пшеницы (*Triticum aestivum*) на разных фенофазах и пять доминирующих видов сорной растительности (*Euphorbia virgata*, *Descurainia sophia*, *Vicia hirsuta*, *Fallopia convolvulus*, *Echinochloa crus-galli*). Алгоритм показал эффективность сегрегации по типу «культура-сорняк» с общей точностью классификации на тестовых данных 72,94 %. Выявлена высокая точность идентификации пшеницы (MCC=0,994) за счет спектральной стабильности и выраженных признаков red edge (680–730 нм). Установлено, что снижение точности при попытке внутриклассового разделения сорняков связано с анатомическими особенностями отдельных представителей (плотные кутикулярные слои *Euphorbia virgata* приводят к проявлению высоких показателей отражения по сравнению с остальными видами; сложная архитектура *Vicia hirsuta* вызывает критический спектральный разброс; сходство строения мезофилла *Echinochloa crus-galli* с культурой, учитывая его принадлежность к злаковым, приводит к частичному оптическому слиянию с некоторыми признаками). Таким образом, разработанная модель позволяет решить главную задачу — обеспечение надежной дифференциации культурной растительности от сорного фона, что является важным условием для формирования технологического базиса точного земледелия и внедрения систем таргетного применения агрохимикатов.

Ключевые слова: гиперспектральное зондирование, яровая пшеница, сорная растительность, машинное обучение, видовая идентификация, видовая дифференциация, спектральные сигнатуры, точное земледелие.

Information about the authors

Ualiyeva Rimma Meyramovna — PhD, Professor, Department of Biology and Ecology, Toraighyrov University, Pavlodar, Kazakhstan; e-mail: ualiyeva.r@gmail.com; ORCID: <https://orcid.org/0000-0003-3551-5007>

Kaverina Mariya Mikhailovna (contact person) — Corresponding author, Junior Researcher, Department of Biology and Ecology, Toraighyrov University, Pavlodar, Kazakhstan; e-mail: k.ma96@mail.ru; ORCID: <https://orcid.org/0000-0002-3838-7656>

Osipova Anastasiya Vyacheslavovna — Master's student, Department of Biology and Ecology, Toraighyrov University, Pavlodar, Kazakhstan; e-mail: aanastasiyaaa@internet.ru; ORCID: <https://orcid.org/0000-0002-3947-5252>

Koppaev Altynbek Erkebulanovich — Master's student, Department of Biology and Ecology, Toraighyrov University, Pavlodar, Kazakhstan; e-mail: altynbek_koppaev@mail.ru; ORCID: <https://orcid.org/0009-0004-0476-9017>

Chaizabekova Meruyert Askarovna — Master of Natural Sciences, Department of Biology and Ecology, Toraighyrov University, Pavlodar, Kazakhstan; e-mail: meruyert130797@gmail.com; ORCID: <https://orcid.org/0009-0004-0468-956X>

Zhangazin Sayan Berikovich — PhD, Associate Professor, Acting Deputy Director for Research at the Institute of Natural Sciences, Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan; e-mail: sayanzhangazin@gmail.com; ORCID: <https://orcid.org/0000-0002-5813-8331>