Л.Н. Гумилев атындағы Еуразия ұлттық университетінің ХАБАРШЫСЫ.
BULLETIN of L.N. Gumilyov Eurasian National University.
ВЕСТНИК Евразийского национального университета имени Л.Н. Гумилева.
ISSN: 2616-7034. eISSN: 2663-130X

IRSTI
34.29.25:68.37.31:29.31.26
Research article

https://doi.org/10.32523/2616-7034-2026-155-2-38-56

Study on the application of hyperspectral imaging for identifying barley loose smut at different stages of development using machine learning methods

R.M. Ualiyeva[1], A.V. Osipova[2], M.M. Kaverina*[3], M.A. Zhaksybek[4], S.B. Zhangazin[5]

[1,2,3,4]Toraighyrov University, Pavlodar, Kazakhstan
[5]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

*E-mail: [1]ualiyeva.r@gmail.com, [2]aanastasiyaaa@internet.ru, *[3]k.ma96@mail.ru, [4]mzhaksybek@gmail.com, [5]sayanzhangazin@gmail.com*

Abstract. This paper presents a study of the potential of hyperspectral imaging for identifying barley infection by *Ustilago nuda* at various stages of disease development using machine learning methods. The study focused on barley samples (*Hordeum vulgare* L.) collected from agrocenoses in the northeastern part of Pavlodar region. Analysis of spectral characteristics revealed significant differences between healthy, infected, and desiccated plant areas, including reduced reflectance and the absence of a distinct red edge in infected tissues. The Maximum Entropy algorithm was employed for classification. The training set comprised 243 samples, while the testing set included 352 samples with pronounced class imbalance. The model demonstrated high classification accuracy of up to 95%. The results confirm the effectiveness of hyperspectral imaging combined with machine learning for disease detection and monitoring, including at early stages, which can be applied in precision agriculture systems to enhance phytosanitary control efficiency and promote sustainable grain production.

Keywords: hyperspectral imaging, loose smut, barley, monitoring of agrocenoses, disease stages.

Introduction

Barley loose smut, caused by the fungus *Ustilago nuda*, is among the economically significant diseases of cereal crops, leading to substantial yield losses and deterioration of grain quality. A distinctive feature of this disease is the latent nature of infection during the early stages of plant development, where infection occurs at the flowering stage, after which the pathogen persists

Received: 20.05.2026. Accepted: 18.03.2026. Available online: 30.06.2026

*Corresponding author

within tissues for an extended period without producing visual symptoms. Spores can overwinter on plant debris and in soil, infecting young plants in spring. Traditional diagnostic methods based on morphological traits prove largely ineffective for the timely detection of infection and prevention of its spread.

Driven by advances in precision agriculture technologies, remote diagnostic methods designed to detect physiological changes in plants before visible disease symptoms appear are gaining particular relevance. One of the most promising approaches is hyperspectral imaging, based on recording the reflectance properties of objects across a broad range of wavelengths. Hyperspectral imaging (HSI) surpasses conventional methods in speed, objectivity, and automation [1]. HSI is used to detect phytopathogens in cereal seeds [2], assess barley seed quality and hardness [3-5], and determine protein, starch, lipid, mycotoxin [6-8], and nutrient content in barley [9].

HSI also enables the detection of fungal infection in grains at the individual kernel level [10, 11], for instance, rice false smut [12], and facilitates mapping of infection foci across fields [13]. The accuracy of loose smut detection using HSI can reach 90% [12, 14], allowing automatic assessment of infection severity [15]. Classification models for detecting non-viable barley grains have also demonstrated accuracy exceeding 90% [16]. Classification models are developed using SVM, Random Forest, neural networks [17, 18], PLSR [19, 20], PLS-DA [21], and CNN [22], employing both pixel-based and object-based classification [23]. Hyperspectral imaging is actively applied to the detection of barley diseases, enabling the identification of infections at various developmental stages [24].

Despite significant progress in the field of hyperspectral plant analysis, questions concerning the identification of barley loose smut specifically remain insufficiently studied. The high sensitivity of hyperspectral methods and their potential for integration with machine learning algorithms open new prospects for developing accurate and automated diagnostic systems.

Accordingly, the aim of the present study is to investigate the application of hyperspectral imaging for identifying barley loose smut using machine learning methods.

Materials and research methods

The study focused on common barley (*Hordeum vulgare* L., 1753) infected with the loose smut pathogen *Ustilago nuda* ((Jensen) Rostr., 1889). Barley samples were collected from fields in the main grain-producing districts of the northeastern part of Pavlodar region, Republic of Kazakhstan. Healthy barley samples served as controls for comparing spectral patterns and determining reflectance characteristics. Infected and healthy samples were collected during the key barley developmental stages (BBCH 25-35, 39-59, 61-75). The climatic conditions of the study area are characterised by a sharply continental climate, uneven precipitation during the field season - from 15-22 mm to 65-70 mm, winds of 3-5 m/s from the NE and SW, mean air temperatures of +13 to +22 °C, and variation in productive soil moisture reserves from 30 to 90 mm throughout the growing season. Plants were collected from sampling plots using rectangular/triangular or checkerboard/diagonal transect patterns [25].

Analysis, hyperspectral imaging, and subsequent investigation were conducted at the Biological Research Laboratory of Toraighyrov University, Pavlodar, Kazakhstan. Imaging was performed using a FigSpec FS-13 camera operating in the 400-1000 nm wavelength range, encompassing both the visible and near-infrared regions. The imaging process employed a CMOS detector with a frame rate of 128 fps, spectral resolution of 2.5 nm, and spatial resolution of 1920 pixels, with a pixel size of 5.86 μm. The acquired images were imported into Breeze software for preprocessing and visualisation of spectral data. To obtain accurate, informative data, calibration

was performed using dark and white reference standards, and the model preprocessing stage included the application of a Savitzky-Golay filter for smoothing and noise reduction, SNV, and centering.

To reduce the dimensionality of the large hypercubes obtained, a PCA model was constructed, enabling visualisation of the most informative principal components (PC1 and PC2). Raw Spectrum plots were used to visualise the spectral characteristics of the studied samples, while Variance Scatter point clouds were employed to reveal the distribution of signatures across individual principal components and their variability with respect to features of disease lesions in the hyperspectral images. All these steps ensured noise minimisation, background pixel exclusion, and manual selection of regions of interest using a grid-based approach.

The imaged samples and extracted spectral features were subsequently used for machine learning to test the performance of the algorithmic model. The Maximum Entropy algorithm was selected as the basis for training, distinguished by its applicability to large datasets and suitability for multiclass tasks. The preceding segmentation stage (Figure 1) was conducted on the training set, which included 243 sample segments - 13.2% were healthy areas used as controls, and 86.8% were loose smut-infected areas. The application of cross-validation helped to improve training accuracy.

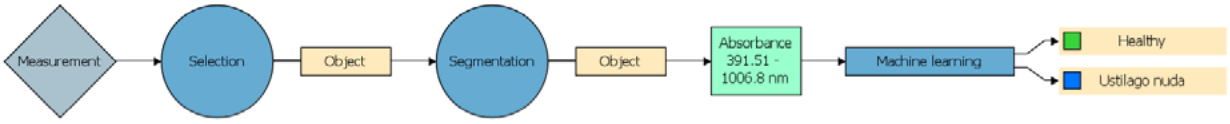

Figure 1. Workflow of the machine learning model from raw data acquisition to automatic detection of disease features

The performance of the machine learning algorithm was tested on samples infected with loose smut at different stages. Based on visual symptoms, and in accordance with State Standard 12044-93 “Methods for determining disease contamination” [26] and the pathogen developmental stages [27] (Figure 2), the infected samples were classified into three stages: early, middle, and late. The number of lesion segments selected for the study across the three developmental stages of loose smut in the testing set totaled 352.

The model was developed on spectral data subjected to analysis of variance (ANOVA) and processed using descriptive statistical methods at $p>0.05$. Model performance and effectiveness were evaluated using the Confusion Matrix and the key quality metrics Precision, Recall, and F1-score, while indicators such as Accuracy and Matthews Correlation Coefficient were used to confirm model performance.

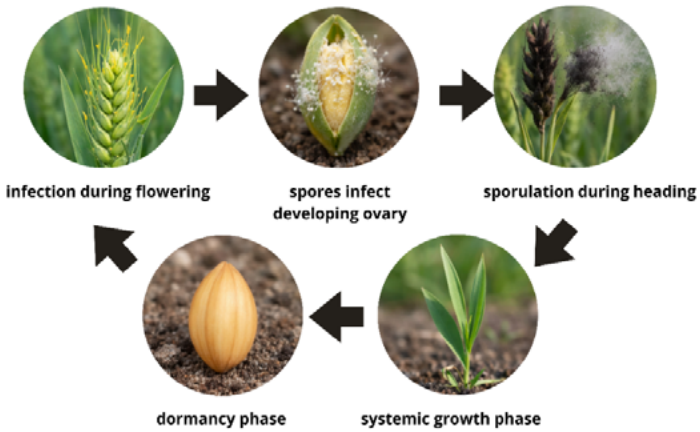

Figure 2. Life cycle of barley loose smut pathogen

Results

The figures present hyperspectral images of barley, Raw Spectrum plots with curves reflecting the spectral characteristics of plant areas, and Variance Scatter diagrams. The hyperspectral image is rendered in false colours corresponding to the intensity of light reflectance. Warm colours (red and yellow) indicate areas with high reflectance, while cool colours (cyan and blue) represent zones of low reflectance. The Raw Spectrum plot contains curves reflecting the spectral characteristics of a specific portion of the hyperspectral image and serves as a basic representation of spectral data, showing the dependence of an object's reflectance coefficient on wavelength (400-1000 nm). This indicator is a key source of information on the physiological state of plant tissues. Variance Scatter diagrams reflect the distribution of objects in principal component space, where t[1] explains the main share of variance and t[2] accounts for the presence and primary characteristics of additional variations. Such visualisation enables assessment of data structure, degree of homogeneity, and differences between disease stages.

Healthy sample

Figure 3 presents a hyperspectral image of a healthy barley sample. The Raw Spectrum plot shows that the highest reflectance intensity is exhibited by spike glumes, with a reflectance coefficient of 30% at the first peak and 27% at the second. The yellow, cyan, and blue curves also correspond to spike areas, with first-peak reflectance intensities of 19%, 18%, and 12%, and second-peak values of 28%, 20%, and 17%, respectively. The green spectrum represents the leaf and some awn areas, where the first-peak reflectance is 10% and the second-peak 17%. The lowest reflectance intensity is observed in the stem and awns, described by the red curve with a reflectance of 5% at the first peak and 16% at the second. The first peak range lies within 500-700 nm, and the second within 700-780 nm. Notably, a sharp decline in the curves with a pronounced red edge is observed at 700 nm.

The Variance Scatter diagram is characterised by a compact, dense cluster in the region of negative or near-zero t [1] values , high point concentration, and limited dispersion. The first principal component t[1] explains 56.7% of total variance, while t[2] explains only 2.16%, which is associated with low intra-class variability.

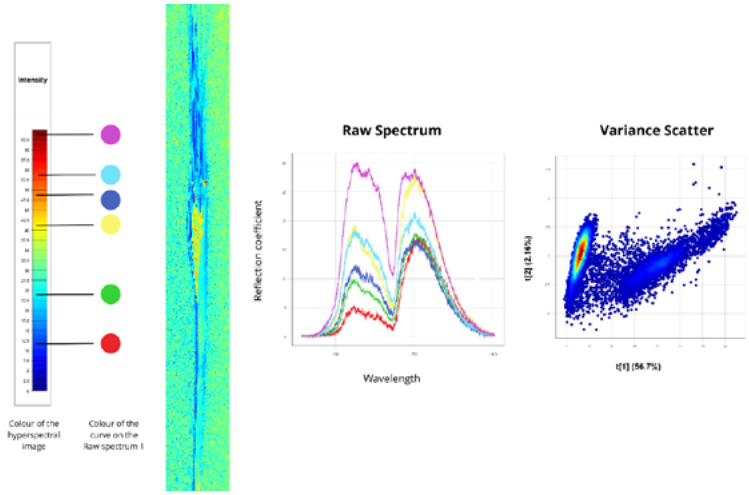

Figure 3. Spectral characteristics of healthy barley

Sample infected with loose smut

Figure 4 shows a barley spike infected with loose smut. In the Raw Spectrum plot, the red curve corresponds to the blue-coloured spike areas affected by the disease, characterised by low spectral intensity with a reflectance coefficient of 5-6%. The remaining spectra belong to intact plant tissue and exhibit higher reflectance. The green, yellow, and blue spectra describe green barley tissue and display two distinct peaks: first-peak reflectance values reach 22.5%, 30%, and 37%, with second-peak values of 20%, 30%, and 35%, respectively. Here, the first peak lies in the 550-700 nm range, while the second is recorded in the 700-780 nm range. The cyan and purple curves represent desiccated spike areas with reflectance of 37.5-46%, with the peak located in the 600-700 nm region.

In the Variance Scatter diagram, the first principal component t[1] explains 74.8% of total variance, while the second component t[2] accounts for approximately 6.17% of variance with less significant variations. Additionally, a broad point distribution, a pronounced arc-shaped structure, and strong elongation along t[1] are observed, indicating high heterogeneity of spectral features.

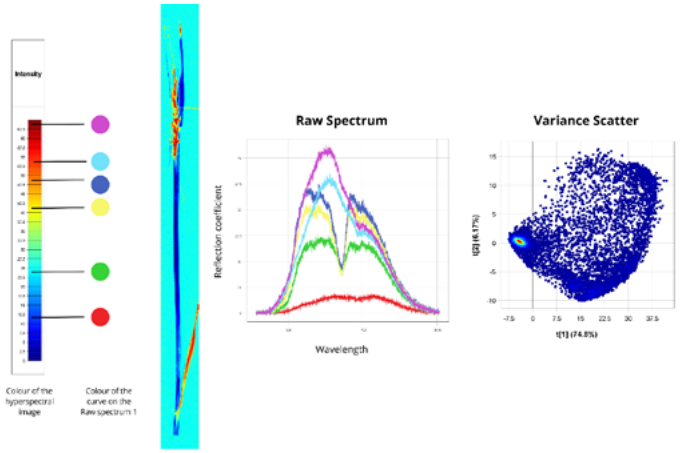

Figure 4. Spectral characteristics of barley infected with loose smut

Testing set

Early stage

Figure 5 shows a sample at the early stage of loose smut development.

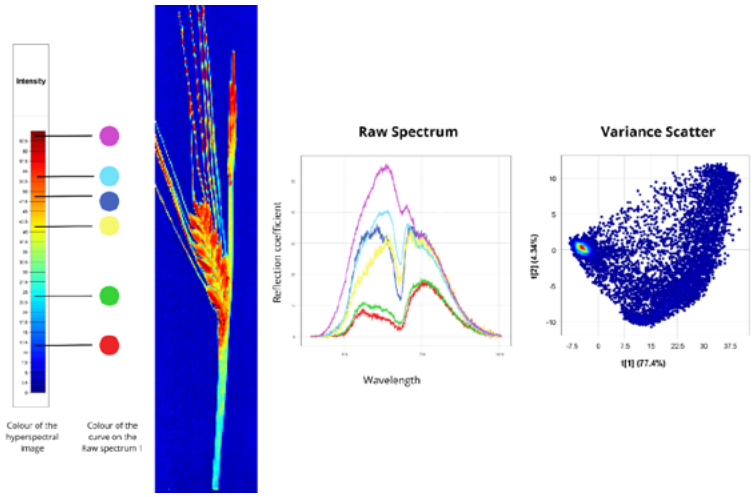

Figure 5. Spectral characteristics of the early stage of loose smut development

The red curve corresponds to infected areas with reflectance in the range of 7-17%, while the green spectrum describes awn areas with reflectance values from 10% to 19%. The yellow, blue, and cyan curves represent healthy stem and spike areas with reflectance of 30-40%, and the purple spectrum depicts healthy spike areas with reflectance reaching 55%. Overall, all curves except the purple one exhibit a pronounced red edge at 700 nm. In the Variance Scatter diagram, the first principal component t[1] explains 77.4% of total variance, while t [2] accounts for only 4.34% of the explained feature variation. The resulting point cloud forms a dense cluster with a distribution tail extending toward positive t [1] values and moderate dispersion. Here, spectral changes are just beginning to emerge, a transitional state is forming, and a significant portion of the sample retains healthy characteristics.

Middle stage

Figure 6 shows a sample at the middle stage of loose smut development, where the red curve corresponds to infected areas with reflectance in the 5-13% range. The green curve describes awn areas with reflectance varying from 10% to 17%. The purple spectrum is characterised by reflectance of 27% and represents unaffected spike areas. The yellow, blue, and cyan curves correspond to healthy stem and spike areas with reflectance ranging from 27% to 38%. All curves except the purple one exhibit a pronounced red edge at 700 nm. In the Variance Scatter diagram, the first principal component t [1] explains 62.4% of total variance, while t [2] accounts for only 4.52% of variance. A marked broadening of the distribution, a clear arc-shaped form, and a rightward shift of the distribution center (along t [1]) are observed, explained by intensifying physiological changes and increasing variability of spectral responses.

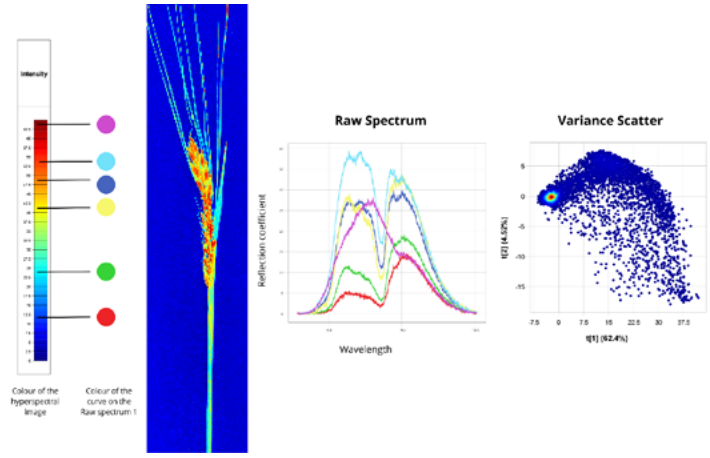

Figure 6. Spectral characteristics of the middle stage of loose smut development

Late stage

Figure 7 shows a sample at the late stage of loose smut development. In the Raw Spectrum plot, the red curve with a reflectance of 5-6% describes areas infected with loose smut. The purple and cyan spectra represent desiccated sample areas with a peak in the 550-700 nm region and reflectance of 35-37.5%. The remaining curves correspond to healthy areas and exhibit two peaks with a pronounced red edge. The first peak is characterised by reflectance values ranging from 16% to 30%, and the second peak from 15% to 28%. According to the Variance Scatter diagram, the first principal component t [1] explains approximately 58.6% of total variance, while the second principal component t [2] explains 8.26% of variance, indicating increased variability and heterogeneity of spectral signatures due to overt disease manifestation. In this case, the distribution exhibits maximum spread, a significant shift along t [1], and a high degree of dispersion and sparsity of points.

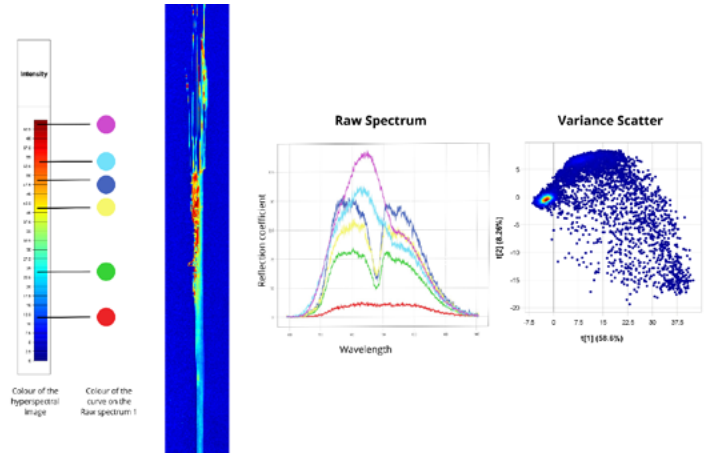

Figure 7. Spectral characteristics of the late stage of loose smut development

Statistical analysis

In order to identify patterns in the pathological process during smut infection, both healthy and infected samples were examined based on reflectance indicators and spectral parameters. For healthy barley, the maximum reflectance was recorded in the spike (mean 21%), the lowest

in the stem with awns (10.5%), and intermediate reflectance intensity was observed in the leaf (13.5%). Particularly noteworthy is the coefficient of variation, ranging from 6.59% for the leaf to 13.30% for the stem with awns, identified as within-group variability, where the greatest heterogeneity according to the indicators was found for stem structures. Additionally, healthy spikes may be characterised by a wider dynamic signal range, manifested primarily through maximum rate of change and delta reflectance values of 11.54 and 18, respectively (Table 1, Figure 8). In healthy plants, all structural elements exhibit reflectance within the 500-780 nm range, with a bandwidth of 280 nm, a contrast ratio of 1.56, a relative bandwidth of 0.219, and a contrast of 0.359 (Table 2, Figure 8). These parameters were thus used as controls for comparison with loose smut-infected samples.

When examining indicators for the second portion of the training set - namely, loose smut infections - all areas were divided into three functional zones: desiccated areas, visually healthy tissue, and tissue directly infected by the pathogen. Desiccated areas were characterised by high reflectance values (37.5 to 46%) combined with low variability (CV=2.59%) and moderate Rate of change and delta reflectance values (7.29 and 8.5, respectively) (Table 1, Figure 8). The spectral range for these areas is narrowed compared with healthy and diseased zones, spanning approximately 600-700 nm, with the contrast ratio decreasing to 1.17 and normalised contrast to 0.143 due to the narrowing of the main spectral response range, confirming homogeneous dehydration and tissue necrosis unaccompanied by active pigment decomposition (Table 2, Figure 8).

Table 1

Barley reflectance characteristics in the spectral domain

№	Species	Part	Reflectance (%)		Mean reflectance	Standard deviation	Coefficient of variation	Rate of change	Delta reflectance
			min	max					
Training set									
1	Healthy	Ear	12	30	21,0	2,29	10,89	11,54	18,0
		Leaf	10	17	13,5	0,89	6,59	4,49	7,0
		Stem and arista	5	16	10,5	1,40	13,30	7,05	11,0
2	*Ustilago nuda*	Dried area	37,5	46	41,8	1,08	2,59	7,29	8,5
		Healthy area	20	37	28,5	2,16	7,58	11,99	17,0
		Affected area	5	6	5,5	0,13	2,31	0,64	1,0
Testing set									
3	Early stage	Healthy area	10	55	32,5	5,72	17,58	28,85	45,0
		Affected area	7	17	12,0	1,27	10,58	6,41	10,0
4	Middle stage	Healthy area	10	38	24,0	3,56	14,82	17,95	28,0
		Affected area	5	13	9,0	1,02	11,29	5,13	8,0
5	Late stage	Dried area	35	37.5	36,3	0,32	0,88	1,96	2,5
		Healthy area	16	28	22,0	1,52	6,93	7,50	12,0
		Affected area	5	6	5,5	0,13	2,31	0,64	1,0

Table 2

Wavelength parameters of barley spectra

№	Species	Part	Wavelength (nm)		Spectral bandwidth	Contrast ratio	Relative bandwidth	Normalised contrast
			min	max				
Training set								
1	Healthy	Ear, leaf, stem and arista	500	780	280	1,56	0,219	0,359
2	*Ustilago nuda*	Dried area	600	700	100	1,17	0,077	0,143
		Healthy area	550	780	230	1,42	0,173	0,295
		Affected area	500	780	280	1,56	0,219	0,359
Testing set								
3	Early stage	Healthy, affected area	500	780	280	1,56	0,219	0,359
4	Middle stage	Healthy, affected area	500	780	280	1,56	0,219	0,359
5	Late stage	Dried area	550	700	150	1,27	0,120	0,214
		Healthy area	500	800	300	1,60	0,231	0,375
		Affected area	500	780	280	1,56	0,219	0,359

Healthy areas on the infected plant are distinguished by intermediate reflectance values, lower than those of desiccated zones, accompanied by elevated variation values of 7.58%. Overall, the spectral parameters of these areas are close to those of healthy tissue and fall within the reflectance wavelength range of the healthy sample, indicating partial preservation of the photosynthetic apparatus but with some shifts toward a stress state.

Loose smut-infected areas are marked by a sharp drop in reflectance to 5.5% with minimal variability of 2.31%. All other derived indicators, such as rate of change and delta, are also at minimal levels (0.64 and 1.0, respectively). Infected sample parts are characterised by low reflectance across a broad spectral range (500-780 nm), which is primarily associated with moist, decomposing tissues exhibiting high chlorophyll (absorption) and water content (Tables 1, 2, Figure 8).

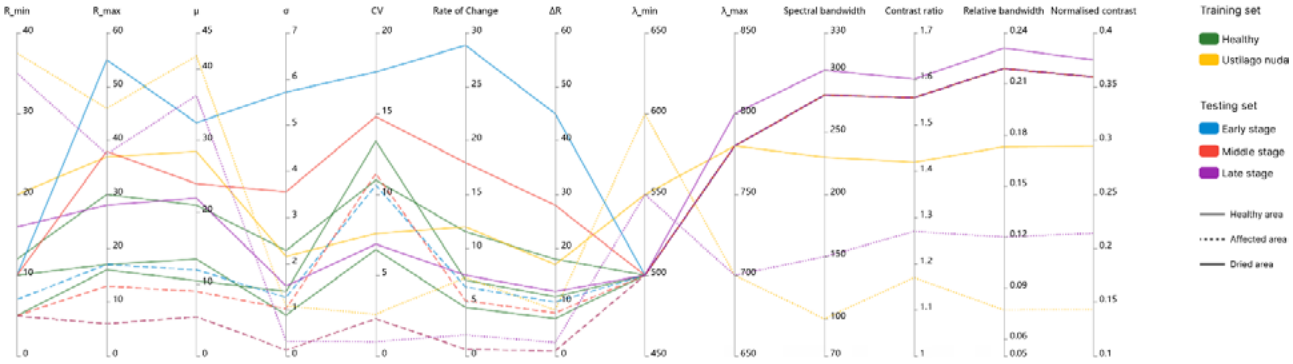

Figure 8. Visualisation of statistical data

A separate analysis was also conducted for the testing set, examining the dynamics of statistical indicators of disease progression. At the early stage, the healthy area is characterised by a mean reflectance of 32.5% and a relatively wide reflectance range (from 10% to 55%) - the highest among all groups - but is distinguished by high variability (CV=17.58%). A delta reflectance value of 45 indicates strong heterogeneity of tissue structures at the initial infection stage. In diseased areas at the early stage, mean reflectance is quite low (only 12%, with CV=10.58% and delta reflectance = 10). Reflectance here is higher than in infected areas of the training set, clearly indicating the initial degradation stage.

At the middle stage of loose smut infection, reflectance in the healthy region decreases (mean to 24%) compared with the early stage and variability remains high (CV=14.82%). Infected areas at this stage become darker, bringing statistical indicators closer to the training set values (reflectance 9%, CV=11.29%) (Table 1, Figure 8).

The late stage is characterised by clear spectral differentiation of zones: the desiccated region exhibits homogeneous and stable indicators with a low coefficient of variation of 0.88% and a narrowed spectral range from 550 to 700 nm, which, combined with a moderate contrast value (0.214), indicates dried and fully necrotised tissues. Healthy areas retain some variability (CV=6.93%) (Tables 1, 2, Figure 8), remaining in close agreement with the reference. However, the overall reflectance value is somewhat reduced, which may indicate partial tissue degradation. Minimal variability and near-complete correspondence with training set indicators are characteristic of infected areas, confirming a stable pathological tissue state. Accordingly, the late stage serves as a clear example of the stabilisation of necrotic and infected zones, accompanied by a simultaneous decline in the functional activity of tissues considered conditionally healthy.

Machine learning

Machine learning was performed using the Maximum Entropy algorithm (Figure 9). The training set comprised 243 samples: 32 (13.2%) healthy and 211 (86.8%) loose smut-infected. 93.8% of the healthy samples and 89.1% of the loose smut-infected samples were correctly classified. A total of 10.3% of samples were assigned to “No class”. The F1-score, which combines Precision and Recall into a single average metric and reflects the balance between the model’s ability to find class objects without including extraneous ones, was 96.8% for healthy samples and 94.2% for *Ustilago nuda*-infected samples. The testing set, which included previously trained samples as well, totaled 595 samples (352 of which were new and unknown to the model): 79 (13.3%) healthy and 516 (86.7%) loose smut-infected. As a result, 93.7% of healthy samples and 87% of diseased samples were correctly classified. A total of 12.1% of samples were assigned to “No

class." The F1-score was 96.7% for healthy samples and 93.1% for *Ustilago nuda* (Figures 9, 10).

In the training set, the proportion of correctly classified instances according to the Accuracy metric was 93.75% for healthy and 89.10% for infected samples. In the testing set, the results remained stable, with 93.67% for healthy and 87.02% for infected samples. Balanced Accuracy values reached 95.73% in the training set and 95.18% in the combined training and testing set. The high and stable values of this indicator demonstrate that the model effectively recognises both classes despite their imbalance. MCC values were 0.955 for the training set and 0.947 for the testing set. The virtually unchanged MCC values between the training and testing sets indicate good generalisation ability of the model and the absence of pronounced overfitting (Table 3).

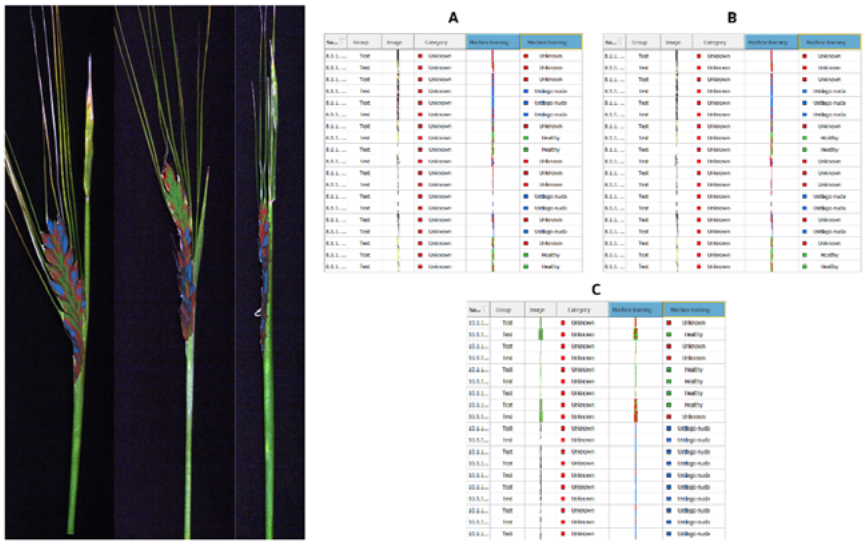

Figure 9. Results of machine learning using the Maximum Entropy algorithm (a - early stage, b - middle stage, c - late stage)

A

Actual classes	Total	Healthy	Ustilago nuda	No class
Healthy	79 (13.3%)	74 (93.7%)		5 (6.33%)
Ustilago nuda	516 (86.7%)		449 (87%)	67 (13%)
# Predicted	595 (100%)	74 (12.4%)	449 (75.5%)	72 (12.1%)
Correctly	523 (87.9%)			
Incorrectly	72 (12.1%)			
Precision		100%	100%	
Recall		93.7%	87%	
F-score		96.7%	93.1%	

B

Actual classes	Total	Healthy	Ustilago nuda	No class
Healthy	32 (13.2%)	30 (93.8%)		2 (6.25%)
Ustilago nuda	211 (86.8%)		188 (89.1%)	23 (10.9%)
# Predicted	243 (100%)	30 (12.3%)	188 (77.4%)	25 (10.3%)
Correctly	218 (89.7%)			
Incorrectly	25 (10.3%)			
Precision		100%	100%	
Recall		93.8%	89.1%	
F-score		96.8%	94.2%	

Figure 10. Confusion matrix (a - Training+Testing sets, b - Training set)

Table 3

Machine learning classification metrics: Maximum Entropy

Metrics	Healthy	*Ustilago nuda*	Average	Healthy	*Ustilago nuda*	Average
	Training set			Training+Testing sets		
Binary Accuracy	93,75	89,10	91,42	93,67	87,02	90,34
Balanced Accuracy	96,90	94,55	95,73	96,85	93,50	95,18
MCC	0,968	0,942	0,955	0,967	0,926	0,947

Discussion

Hyperspectral imaging enables the identification of pathogens based on their molecular structure. The dark colour of spores may be attributed to the presence of melanin-like pigments, as found in *Podospora anserina* [28], *Pestalotiopsis microspora* [29], and *Ustilago maydis* [30]. Melanin can protect fungal spores from ultraviolet radiation, enhance desiccation resistance, and reinforce the spore cell wall. Owing to their dark colouration, loose smut spores absorb more light than they reflect; consequently, the spectra of plant tissues infected by these fungi differ markedly, exhibiting reduced intensity. Healthy plant tissue exhibits high reflectance due to the green pigment chlorophyll and the reflectivity of the mesophyll, and the spectral curves of these areas differ clearly from those of tissues infected with loose smut. It was found that healthy plant curves exhibit a pronounced red edge at 700 nm, whereas the spectra of diseased areas lack distinct peaks and a red edge.

It was established that desiccated areas of infected plants also exhibit no pronounced red edge and appear as a spectral curve with a single high-reflectance peak. Data on the spectral differences among healthy, diseased, and desiccated areas are shown in Figure 11.

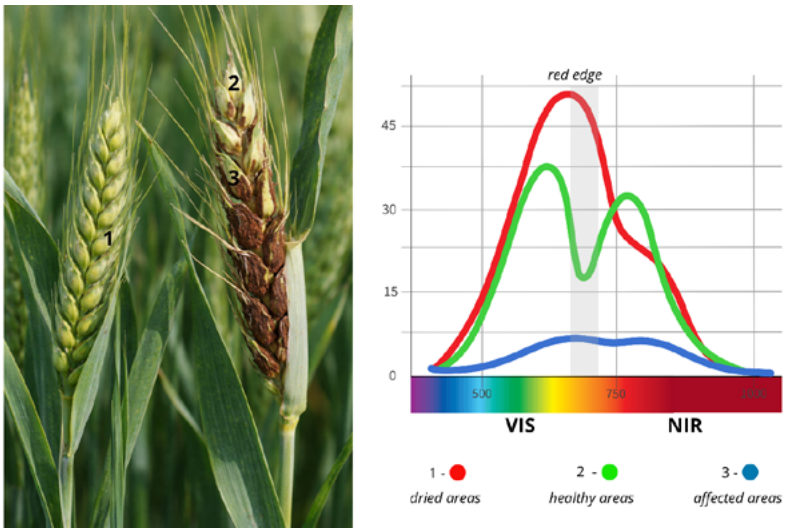

Figure 11. Spectral differences among healthy, infected, and desiccated barley tissues

The numerical spectral indicators obtained enable statistical substantiation of the differentiation among healthy, necrotised, and actively loose smut-infected tissues. These differences are governed by reflectance level, spectral bandwidth, and feature variability. Statistical analysis demonstrated that infected areas predominantly exhibit consistently low reflectance (5-6%, CV=3%), associated with high moisture content and pigment absorption under conditions of disrupted cellular structure. This signal is presumably largely stage-independent and may therefore serve as a reliable diagnostic marker. The reflectance level of desiccated tissues is characterised by high values (36-42%) and low variability (CV<3%), attributed to the dehydration process and the resulting strong scattering. The breakdown of individual cellular structures and pigments leads to spectral narrowing compared to the normal range, to approximately 600-700 nm. Healthy areas on infected plants exhibit signs of systemic stress and generally show intermediate reflectance values (22-32%). At early disease stages, a transient increase in reflectance may be observed, potentially associated with compensatory tissue activity.

Ustilago nuda develops gradually, and at early stages, changes in spectral characteristics may be subtle, complicating classification. High dimensionality and noise sensitivity lead to partial overlap of features from different classes. Meanwhile, the predominance of infected samples may influence decision boundaries; however, in this case, the effect is successfully compensated by the algorithm. The use of the Maximum Entropy algorithm, based on a probabilistic approach with regularization, enabled minimization of the imbalance effect, as confirmed by high Balanced Accuracy values. According to Binary and Balanced Accuracy metrics, the Maximum Entropy algorithm correctly accounts for class distribution and shows no pronounced bias toward the dominant class. High F1-score values indicate a balanced model with effective detection of infected samples. Notably, the slight decrease in this metric for infected plants is associated with the heterogeneity of *Ustilago nuda* disease manifestation, especially at early stages when spectral differences between classes are minimal. The increase in the proportion of uncertain classifications in the testing set may be due to overlapping spectral features between healthy and infected plants and the presence of intermediate physiological states. The higher accuracy for healthy samples, despite their underrepresentation, can be explained by the lower variability of their spectral characteristics. Healthy plants form a more homogeneous spectral profile, whereas infected samples exhibit a wide range of changes depending on the pathogen developmental stage and infection severity. The slight decrease in accuracy for infected samples in the testing set indicates the presence of challenging or borderline cases, which is typical for early plant disease diagnostics.

Since MCC is one of the most reliable quality metrics for binary classification with imbalanced data, the obtained values (~95%) indicate high agreement between model predictions and actual classes. It follows that infected zones are stable markers of active infection. Reflectance values for healthy tissues align with known data for cereals [31-33], and the typical contrast between low reflectance of infected tissues and high reflectance of necrotised tissues agrees with observations for other pathogens, though loose smut is characterised by particularly pronounced patterns [34]. Accordingly, the analysed statistical features can be used to develop methods for identifying barley smut using hyperspectral sensors.

Overall, the model demonstrated high effectiveness, with high classification accuracy and result reliability, robustness to class imbalance, and good generalisation ability. Despite a proportion of uncertain classifications, the model can be used for preliminary diagnosis of barley diseases. Further improvements in accuracy are possible through increasing the volume of balanced data and integrating ensemble and deep learning methods.

Conclusion

This study evaluated the effectiveness of hyperspectral imaging for identifying barley infection caused by loose smut (*Ustilago nuda*). It was established that hyperspectral data enable the detection of subtle changes in plant physiological state caused by infection development, even before pronounced morphological symptoms appear. Diseases accompanied by the formation of dark spores are characterised by reduced reflectance in the visible and near-infrared ranges and by spectral curves lacking a distinct red edge, which may serve as a diagnostic marker. The dark colouration of fungal plant diseases is driven by adaptations for spore dispersal and survival under agricultural field conditions. Based on spectral differences resulting from variations in pigment content between healthy and diseased tissues, a classification model was developed, achieving a detection accuracy of 95%. The application of the Maximum Entropy algorithm demonstrated high classification efficiency on both training and testing sets. The higher classification accuracy for healthy samples is associated with their lower spectral variability, whereas infected plants are characterised by heterogeneous features due to the different developmental stages of *Ustilago nuda*. The model exhibited high binary and balanced accuracy values, as well as a high F1-score, indicating its ability to correctly recognise both healthy and infected samples even under considerable class imbalance. The high Matthews Correlation Coefficient values confirm the reliability and robustness of the results.

Thus, the findings confirm the high diagnostic value of this approach for the early detection of phytopathogens. The study demonstrates that the application of hyperspectral imaging can enhance the efficiency of phytosanitary monitoring, enable early detection of barley loose smut, and reduce economic losses in agricultural production.

Author Contributions

U.R.M. - concept and supervision of the work; **O.A.V.** - discussion of the research results; **K.M.M.** - conducting the experiments; **Zh.M.A.** - writing the text; **Zh.S.B.** - editing the text of the article.

Funding

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23485162 “Development of an innovative method for monitoring and early diagnosis of grain crop diseases using hyperspectral sensing technology”).

Conflict of Interest

The authors declare no conflict of interest.

Compliance with ethical standards

This article does not contain a description of studies performed by the authors involving people or using animals as objects.

References

1. Moskovsky MN, Shogenov YH, Lavrov AV, Gulyaev AA, Efremenkov IY, Pyatchenkov DS, Belyakov MV. Spectral photoluminescent parameters of barley seeds (Hordeum vulgare) infected with Fusarium spp. Photochemistry and Photobiology. 2022;99(1). https://doi.org/10.1111/php.13645

2. Ferreira LDC, Carvalho ICB, Jorge LADC, Quezado-Duval AM, Rossato M. Hyperspectral imaging for the detection of plant pathogens in seeds: Recent developments and challenges. Frontiers in Plant Science. 2024;15:1387925. *https://doi.org/10.3389/fpls.2024.1387925*

3. Xiong C., She Y., Jiao X., Zhang T., Wang M., Wang M., Abd El-Aty A.M., Wang J., Xiao M. Rapid nondestructive hardness detection of black highland barley kernels via hyperspectral imaging. Journal of Food Composition and Analysis. 2024;127:105966. *https://doi.org/10.1016/j.jfca.2023.105966*

4. Lohani UC, Divyanth LG, Singla D, Senthilkumar T, Singh CB. Comparative study of digital and hyperspectral imaging for varietal discrimination of dill seeds. European Food Research and Technology. 2026;252:41. *https://doi.org/10.1007/s00217-025-05024-1*

5. Trujillo HA, Velasquez Vasconez PA, Alves RM, Iost Filho FH, Gomes-junior FG.. Hyperspectral imaging and machine learning for assessing germination of hydroprimed lettuce seeds. Journal of Tekirdağ Agricultural Faculty. 2026;23:56-100. https://doi.org/10.33462/jotaf.1586666

6. Sendin K, Williams PJ, Manley M. Near infrared hyperspectral imaging in quality and safety evaluation of cereals. Critical Reviews in Food Science and Nutrition. 2018;58:575-590. *https://doi.org/10.1080/10408398.2016.1205548*

7. An D, Zhang L, Liu Z, Liu J, Wei Y. Advances in infrared spectroscopy and hyperspectral imaging combined with artificial intelligence for detection of cereal quality. Critical Reviews in Food Science and Nutrition. 2023;63(29):9766-9796. *https://doi.org/10.1080/10408398.2022.2066062*

8. Liang Y, Li Z, Shi J, Zhang N, Qin Z, Du L, Zhai X, Shen T, Zhang R, Zou X, Huang X. Advances in hyperspectral imaging technology for grain quality and safety detection: A review. Foods. 2025;14:2977. *https://doi.org/10.3390/foods14172977*

9. Grieco M, Schmidt M, Warnemünde S, Backhaus A, Klück HC, Garibay A, Tandrón Moya YA, Jozefowicz AM, Mock HP, Seiffert U, Maurer A, Pillen K. Dynamics and genetic regulation of leaf nutrient concentration in barley based on hyperspectral imaging and machine learning. Plant Science. 2022;315:111123. *https://doi.org/10.1016/j.plantsci.2021.111123*

10. Femenias A, Gatius F, Ramos AJ, Teixido-Orries I, Marín S. Hyperspectral imaging for the classification of individual cereal kernels according to fungal and mycotoxins contamination: A review. Food Research International. 2022;155:111102, *https://doi.org/10.1016/j.foodres.2022.111102*

11. Sun H, Zhang L, Rao Z, Ji H. Determination of moisture content in barley seeds based on hyperspectral imaging technology. Spectroscopy Letters. 2020;53(10):751-762. *https://doi.org/10.1080/00387010.2020.1832531*

12. Wu N, Jiang H, Bao Y, Zhang C, Zhang J, Song W, Zhao Y, Mi C, He Y, Liu F. Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions. Sensors and Actuators B: Chemical. 2020;308:127696. *https://doi.org/10.1016/j.snb.2020.127696*.

13. An G, Xing M, He B, Kang H, Shang J, Liao C, Huang X, Zhang H. Extraction of areas of rice false smut infection using UAV hyperspectral data. Remote Sensing. 2021;13(16):3185. *https://doi.org/10.3390/rs13163185*

14. Bao D, Zhou J, Bhuiyan SA, Adhikari P, Tuxworth G, Ford R, Gao Y. Early detection of sugarcane smut and mosaic diseases via hyperspectral imaging and spectral-spatial attention deep neural networks. Journal of Agriculture and Food Research. 2024;18:101369. *https://doi.org/10.1016/j.jafr.2024.101369*

15. Thomas S, Behmann J, Steier A, Kraska T, Muller O, Rascher U, Mahlein AK. Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. Plant Methods. 2018;14:45. https://doi.org/10.1186/s13007-018-0313-8

16. McGoverin CM, Engelbrecht P, Geladi P, Manley M. Characterisation of non-viable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics. Analytical and Bioanalytical Chemistry. 2011;401:2283-2289. *https://doi.org/10.1007/s00216-011-5291-x*

17. Wan L, Li H, Li C, Wang A, Yang Y, Wang P. Hyperspectral sensing of plant diseases: Principle and methods. Agronomy. 2022;12:1451. *https://doi.org/10.3390/agronomy12061451*

18. Singh T, Sharma A, Garg NM, Iyengar SRS. Near-Infrared Hyperspectral Imaging in Tandem with Machine Learning Techniques to Identify the Near Geographical Origins of Barley Seeds. In: Tistarelli, M., Dubey, S.R., Singh, S.K., Jiang, X. (eds) Computer Vision and Machine Intelligence. Lecture Notes in Networks and Systems. 2023; 586. Springer, Singapore. *https://doi.org/10.1007/978-981-19-7867-8_53*

19. Sun T, Xu M, Jin Z. Grain geometry matters: Hyperspectral imaging challenges for moisture detection in individual kernels of barley. Journal of Cereal Science. 2026;127:104352. *https://doi.org/10.1016/j.jcs.2025.104352*

20. Caporaso N, Whitworth MB, Fisk ID. Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. Applied Spectroscopy Reviews. 2018;53:667-687. *https://doi.org/10.1080/05704928.2018.1425214*

21. Sun H, Zhang L, Li H, Rao Z, Ji H. Nondestructive identification of barley seed varieties using hyperspectral data from two sides of barley seeds. Journal of Food Process Engineering. 2021;44:e13769. *https://doi.org/10.1111/jfpe.13769*

22. Terentev A, Kuznetsova D, Fedotov A, Baranova O, Eremenko D. Cross-crop transferability of machine learning models for early stem rust detection in wheat and barley using hyperspectral imaging. Plants. 2025;14:3265. *https://doi.org/10.3390/plants14213265*

23. Chossegros M, Hubbard A, Burt M, Harrison RJ, Nellist CF, Grinberg NF. Hyperspectral image analysis for classification of multiple infections in wheat. Plant Methods. 2025;21:144. *https://doi.org/10.1186/s13007-025-01461-x*

24. Zhou RQ, Jin JJ, Li QM, Su ZZ, Yu XJ, Tang Y, Luo SM, He Y, Li XL. Early detection of Magnaporthe oryzae-infected barley leaves and lesion visualization based on hyperspectral imaging. Frontiers in Plant Science. 2019;9:1962. *https://doi.org/10.3389/fpls.2018.01962*

25. Polyakov IYa, Persov MP, Smirnov VA. Forecasting the development of agricultural crop pests and diseases (with practical exercises). Leningrad: Kolos; 1984. p. 318.

26. State Standard 12044-93. Methods for determining disease contamination. Moscow: Standartinform; 2011. p. 57.

27. Ökmen B, Schwammbach D, Bakkeren G, Neumann U, Doehlemann G. The *Ustilago hordei* - Barley Interaction is a Versatile System for Characterization of Fungal Effectors. Journal of Fungi, 2021;7(2):86. *https://doi.org/10.3390/jof7020086*

28. Gautier V, Levert E, Giraud T, Silar P. Important role of melanin for fertility in the fungus Podospora anserina. G3: Genes, Genomes, Genetics. 2021;11(8):jkab159. *https://doi.org/10.1093/g3journal/jkab159*

29. Yu X, Huo L, Liu H, Chen L, Wang Y, Zhu X. Melanin is required for the formation of multicellular conidia in the endophytic fungus Pestalotiopsis microspora. Microbiological Research. 2015;179:1-11. *https://doi.org/10.1016/j.micres.2015.06.004*

30. Reyes-Fernández EZ, Shi Y-M, Grün P, Bode HB, Bölker M. An unconventional melanin biosynthesis pathway in Ustilago maydis. Applied and Environmental Microbiology. 2021;87:e01510-20.. *https://doi.org/10.1128/AEM.01510-20*

31. Khan IH, Liu H, Li W, Cao A, Wang X, Liu H, Cheng T, Tian Y, Zhu Y, Cao W, Yao X. Early detection of powdery mildew disease and accurate quantification of its severity using hyperspectral images in wheat. Remote Sensing. 2021;13:3612. *https://doi.org/10.3390/rs13183612*

32. Yao Z, Lei Y, He D. Early visual detection of wheat stripe rust using visible/near-infrared hyperspectral imaging. Sensors. 2019;19:952. *https://doi.org/10.3390/s19040952*

33. Lim J, Kim G, Mo C, Oh K, Kim G, Ham H, Kim S, Kim MS. Application of near infrared reflectance spectroscopy for rapid and non-destructive discrimination of hulled barley, naked barley, and wheat contaminated with Fusarium. Sensors. 2018;18:113. *https://doi.org/10.3390/s18010113*

34. Ualiyeva RM, Kaverina MM, Osipova AV, Kairbayev YB, Zhangazin SB, Iksat NN, Mapitov NB. VNIR hyperspectral signatures for early detection and machine-learning classification of wheat diseases. Plants. 2025;14:3644. *https://doi.org/10.3390/plants14233644*

Машиналық оқыту әдістері негізінде арпаның тозаңды қаракүйесін дамудың әртүрлі кезеңдерінде анықтауда гиперспектральды визуализацияны қолдануды зерттеу

Р.М. Уалиева[1], А.В. Осипова[2], М.М. Каверина*[3], М.Ә. Жақсыбек[4], С.Б. Жангазин[5]

[1,2,3,4]Торайғыров университеті, Павлодар, Қазақстан
[5]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан

Аңдатпа. Мақалада машиналық оқыту әдістерін қолдана отырып, аурудың дамуының әртүрлі кезеңдерінде арпаның *Ustilago nuda* қоздырғышымен зақымдануын анықтау үшін гиперспектралды визуализацияның мүмкіндіктері зерттелген. Зерттеу нысаны ретінде Павлодар облысының солтүстік-шығысындағы агроценоздардан іріктеліп алынған арпа үлгілері (*Hordeum vulgare* L.) пайдаланылды. Спектрлік сипаттамаларды талдау өсімдіктердің сау, зақымданған және қураған бөліктері арасындағы елеулі айырмашылықтарды көрсетті, соның ішінде шағылысу қабілетінің төмендеуі және зақымданған тіндерде айқын red edge аймағының болмауы анықталды. Жіктеу үшін Maximum Entropy алгоритмі қолданылды. Оқыту жиынтығына 243 үлгі, ал тестілеу жиынтығына санаттар арасындағы айқын теңгерімсіздігі бар 352 үлгі енгізілді. Модель 95%-ға дейін жоғары жіктеу дәлдігін көрсетті. Алынған нәтижелер гиперспектралды визуализацияны машиналық оқыту әдістерімен бірге қолданудың ауруды диагностикалауда және оның дамуын, соның ішінде ерте кезеңдерде, мониторингілеуде тиімді екенін растады. Бұл тәсіл фитосанитарлық бақылаудың тиімділігін арттыру және астық өндірісінің тұрақты дамуын қамтамасыз ету мақсатында дәлме-дәл егіншілік жүйелерінде қолданылуы мүмкін.

Түйін сөздер: гиперспектралды визуализация, тозаңды қаракүйе, арпа, агроценозды бақылау, ауру кезеңдері.

Изучение применения гиперспектральной визуализации для идентификации пыльной головни ячменя на разных стадиях развития с использованием методов машинного обучения

Р.М. Уалиева[1], А.В. Осипова[2], М.М. Каверина*[3], М.Ә. Жақсыбек[4], С.Б. Жангазин[5]

[1,2,3,4]Торайгыров университет, Павлодар, Казахстан
[5]Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан

Аннотация. В статье представлено исследование возможностей гиперспектральной визуализации для идентификации поражения ячменя *Ustilago nuda* на различных стадиях развития заболевания с применением методов машинного обучения. Объектом исследования выступали образцы ячменя (*Hordeum vulgare* L.), отобранные в агроценозах северо-востока Павлодарской области. Анализ спектральных характеристик показал существенные различия между здоровыми, поражёнными и высохшими участками растений, включая снижение отражательной способности и отсутствие выраженного red edge у инфицированных тканей. Для классификации использован алгоритм Maximum Entropy. Обучающая выборка включала 243 образца, тестовая - 352 образца с выраженным дисбалансом классов. Модель показала высокую точность классификации до 95%. Полученные результаты подтверждают эффективность гиперспектральной визуализации в сочетании с машинным обучением для диагностики заболевания и мониторинга его развития в том числе на ранней стадии, что может быть использовано в системах точного земледелия для повышения эффективности фитосанитарного контроля и устойчивого развития зернопроизводства.

Ключевые слова: гиперспектральная визуализация, пыльная головня, ячмень, мониторинг агроценозов, стадии болезни.

Authors' information:

Ualiyeva Rimma Meyramovna – PhD, Professor of the Department of Biology and Ecology, Toraighyrov University, 64 Lomov St., 140008, Pavlodar, Kazakhstan.

Osipova Anastasiya Vyacheslavovna –Master's degree student, Department of Biology and Ecology, Toraighyrov University, 64 Lomov St., 140008, Pavlodar, Kazakhstan.

Kaverina Mariya Mikhailovna – Corresponding author, Junior Researcher at the Department of Biology and Ecology, Toraighyrov University, 64 Lomov St., 140008, Pavlodar, Kazakhstan.

Zhaksybek Meruyert Adilbekkyzy – PhD student of the Department of Biology and Ecology, group DB-12/2, Toraighyrov University, 64 Lomov St., 140008, Pavlodar, Kazakhstan.

Zhangazin Sayan Berikovich – PhD, Associate Professor, Deputy Dean for Research of the Faculty of Natural Sciences, L.N. Gumilyov, Eurasian National University, 2 Satpayev St., 010008, Astana, Kazakhstan.

Авторлар туралы мәлімет:

Уалиева Римма Мейрамовна – PhD, «Биология және экология» кафедрасының профессоры, «Торайғыров университеті» КеАҚ, Ломов көшесі, 64, 140008, Павлодар, Қазақстан.

Осипова Анастасия Вячеславовна – «Биология және экология» кафедрасының магистранты, «Торайғыров университеті» КеАҚ, Ломов көшесі, 64, 140008, Павлодар, Қазақстан.

Каверина Мария Михайловна – хат-хабар авторы, «Биология және экология» кафедрасының кіші ғылыми қызметкері, «Торайғыров университеті» КеАҚ, Ломов көшесі, 64, 140008, Павлодар, Қазақстан.

Жақсыбек Меруерт Әділбекқызы – ДБ-12/2 тобының докторанты, «Биология және экология» кафедрасы, «Торайғыров университеті» КеАҚ, Ломов көшесі, 64, 140008, Павлодар, Қазақстан.

Жангазин Саян Берикович – PhD, қауымдастырылған профессор, Жаратылыстану ғылымдары факультеті деканының ғылыми жұмыс жөніндегі орынбасары, «Л.Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ, Сәтбаев көшесі 2, 010008, Астана, Қазақстан.

Сведения об авторах:

Уалиева Римма Мейрамовна - PhD, профессор кафедры биологии и экологии, НАО «Торайгыров университет», ул. Ломова, 64, 140008, Павлодар, Казахстан.

Осипова Анастасия Вячеславовна - магистрант кафедры биологии и экологии, НАО «Торайгыров университет», ул. Ломова, 64, 140008, Павлодар, Казахстан.

Каверина Мария Михайловна - автор-корреспондент, младший научный сотрудник кафедры биологии и экологии, НАО «Торайгыров университет», ул. Ломова, 64, 140008, Павлодар, Казахстан.

Жақсыбек Меруерт Әділбекқызы - докторант группы ДБ-12/2 кафедры биологии и экологии, НАО «Торайгыров университет», ул. Ломова, 64, 140008, Павлодар, Казахстан.

Жангазин Саян Берикович - PhD, ассоциированный профессор, заместитель декана по научной работе факультета естественных наук, НАО «Евразийский национальный университет им. Л.Н. Гумилева», ул. Сатпаева, 2, 010008, Астана, Казахстан.