ISSN 1727-1320 (Print),
ISSN 2308-6459 (Online)

ВЕСТНИК ЗАЩИТЫ РАСТЕНИЙ

PLANT PROTECTION NEWS

2026 ТОМ VOLUME 109 ВЫПУСК ISSUE 1

Санкт-Петербург
St. Petersburg, Russia

OECD+WoS: 4.01+AM (Agronomy); 1.06+QU (Plant Sciences) https://doi.org/10.31993/2308-6459-2026-109-1-17459

Full-text article

EVALUATION OF A HYPERSPECTRAL IMAGING DATA PROCESSING PIPELINE FOR EARLY RUST DISEASE DIAGNOSIS IN GRAIN CROPS APPLIED TO WHEAT, RYE, AND BARLEY PHENOTYPING

D.V. Kuznetsova[1], O.A. Baranova[2], D.A. Emelyanov[2], D.Yu. Eremenko[1], A.A. Fedotov[1,2], A.B. Terentev[2]*

[1]Peter the Great Polytechnic University, St. Petersburg, Russia
[2]All-Russian Institute of Plant Protection, St. Petersburg, Russia

**corresponding author, e-mail: admin@vizr.spb.ru*

Hyperspectral sensing data processing pipeline, originally developed for the early diagnosis of rust diseases in grain crops, was assessed for its applicability for the task of phenotyping of healthy plants of wheat *Triticum aestivum*, barley *Hordeum vulgare*, and rye *Secale cereale*. Hyperspectral images of healthy plants, obtained under laboratory conditions using a Cubert Ultris 20 camera (450–874 nm range, 106 channels), were utilized. The effectiveness of various preprocessing schemes was compared: full (including normalization, smoothing, calculation of derivatives, and identification of extreme features), reduced, and minimal. Machine learning models were exploited for classification: logistic regression, support vector machine, and gradient boosting, trained on averaged spectra. It is shown that the use of a full pipeline optimized for phytopathological diagnostics leads to reduced classification accuracy in phenotyping tasks. The best results (F1 = 0.97 ± 0.025) were achieved using the original averaged spectral curves without additional transformations. It is concluded that for healthy wheat, barley, and rye phenotyping, absolute reflectance levels are informative, whereas for disease diagnostics, changes in the shape of the spectral curve are more important. The obtained results clarify the applicability limits of pipelines developed for phytosanitary purposes and can inform the development of remote monitoring and phenotyping systems for cereal crops.

Keywords: rye, barley, wheat, hyperspectral imaging, data processing pipeline, phenotyping, plant disease diagnostics

Submitted: 20.10.2025 *Accepted: 01.03.2026*

Introduction

Wheat *Triticum aestivum* L., barley *Hordeum vulgare* L., and rye *Secale cereale* L. are the key cereal crops in Russian agriculture. Collectively, they occupy the largest sown area within the national cropping structure and form the basis of the country's food security (Guzenko et al., 2024; Sysuev et al., 2025; Zimnyakov et al., 2020). Given the scale of cultivation of these crops and their high sensitivity to biotic and abiotic stresses, the development of technologies that ensure prompt monitoring of their condition is of particular importance (Lysov and Pavlyushin, 2022; Yakushev et al., 2022).

Plant phenotyping is the quantitative description of morphological, physiological, and biochemical properties reflecting plant responses to genetic factors and environmental conditions (Mishra et al., 2016; Pieruschka and Schurr, 2019). Phenotyping using high-throughput automated systems with various sensors plays an important role in plant protection. It enables the identification of pathogen resistance in varieties based on specific spectral traits. This supports breeding programs and facilitates the detection of early-stage diseases, allowing timely application of protective measures (Gavrilenko et al., 2021; Danilov et al., 2024).

Traditional visual and laboratory methods, although remaining the basis of phytopathological analysis, are characterized by high labor intensity and limited applicability to large areas (Mahlein, 2016). Therefore, modern plant production requires the implementation of new, automated methods for crop diagnostics and monitoring, enabling rapid assessment of plant physiological status and prediction of stress development (Farber et al., 2019; Zhang et al., 2019). In this regard, there is growing interest in remote and optical methods for assessing plant condition, ensuring objective and reproducible measurements with minimal time requirements (Bock et al., 2022; Terentev and Dolzhenko, 2023). In recent years, the development of phenotyping has significantly accelerated thanks to non-invasive optical technologies such as RGB imaging, multi- and hyperspectral imaging, thermography, and chlorophyll fluorescence measurements. These methods provide objective and reproducible data, complementing traditional imaging and molecular approaches (Li et al., 2014; Mahlein, 2016).

One of the most promising areas is the use of hyperspectral imaging methods (Terentev et al., 2022; Wan et al., 2022). These methods allow the identification of even small spectral differences in the objects under study, caused by both stress and other factors. They are particularly well suited for solving plant phenotyping tasks, including those within the same taxonomic group (Mishra et al., 2020; Sarić et al., 2022).

Hyperspectral imaging (HSI) records the reflectance of the plant canopy over tens or hundreds of narrow spectral ranges, enabling the detection of subtle differences in plant physiological states that are otherwise invisible to the naked eye (Cheshkova et al., 2022; Khan et al., 2022). The potential

of this method has been confirmed by numerous studies demonstrating the possibility of early detection of biotic and abiotic stresses in various agricultural crops (Lowe et al., 2017; Terentev et al., 2022). However, when applied to phenotyping tasks, the data often have high dimensionality. This can lead to feature redundancy, complicate analysis, and increase the risk of model overfitting (Sarić et al., 2022). Nevertheless, HSI has been effectively applied in laboratory, greenhouse, and field conditions to create spectral portraits of plants and perform their further classification (Mishra et al., 2020; Sarić et al., 2022).

Globally, the greatest attention has been paid to research on cereal crops, primarily wheat, for which a variety of approaches to spectral diagnostics of diseases and stress conditions have been developed and tested (Terentev et al., 2022). In Russian research, this area has been actively developing in recent years and is considered a promising monitoring tool (Yakushev et al., 2022; Zolotukhina et al., 2023; Terentev et al., 2023; Zolotukhina et al., 2024). However, the applicability of hyperspectral data processing algorithms developed for phytopathological diagnostics specifically in Russian conditions to a wide range of crops and phenotyping tasks remains insufficiently studied—constituting a methodological gap addressed in this study.

The authors previously developed an algorithmic scheme for preprocessing and analyzing hyperspectral data aimed at solving the problem of early diagnosis of rust diseases of grain crops, in particular stem rust of wheat and barley (caused by *Puccinia graminis* f. sp. *tritici*) and leaf rust of wheat (caused by *Puccinia triticina*) (Terentev et al., 2023; Terentev et al., 2025; Fedotov et al., 2025). The developed data processing pipeline was optimized for the early detection of physiological abnormalities associated with pathological tissue changes. When switching to tasks unrelated to the detection of disease symptoms, such as species phenotyping of healthy plants, the accuracy of the algorithm's prediction could decrease. In such tasks, spectral differences between crops are significantly less pronounced than between healthy and diseased plants, and the use of a full range of pre-treatments can lead to the loss of informative features associated with absolute reflectance levels.

This study aims to evaluate the portability and generality of a previously developed toolkit. Specifically, we aim to test the applicability of a data processing pipeline, including pixel-, curve-, and channel-level preprocessing strategies, as well as first derivatives and categorical derivatives, developed for disease diagnostics (Fedotov et al., 2025; Terentev et al., 2025), to the problem of classifying healthy plants of three crops—wheat, barley, and rye—based on hyperspectral data.

To achieve this goal, the following objectives were set:

1. To evaluate classification performance using the full preprocessing pipeline and its abbreviated versions, as well as minimal processing of the original hyperspectral data.
2. To compare the information content of different feature types—original spectral curves, first derivatives, categorical derivatives, and extreme features extracted from them.
3. To determine which processing stages have a positive or negative impact on the classification accuracy of healthy plants and to formulate recommendations for their application.

Thus, this work aims to evaluate the applicability of a hyperspectral data preprocessing pipeline, originally developed for automated diagnosis of rust diseases, to the task of automated phenotyping of healthy plants of three cereal crops.

Materials and Methods

The study was conducted on spring soft wheat (variety "Saratovskaya 74"), spring barley (variety "Lyuboyar"), and winter rye (variety "Volkhova"). In each experiment, wheat, rye, and barley were cultivated in six plastic containers (320 × 220 × 160 mm), with 75 seeds sown in each container. The sowing density was adjusted to correspond to the recommended field seeding rates for the respective varieties, recalculated according to the container area and official varietal descriptions. Each experiment included three independent biological replicates. The plants were grown under controlled conditions with regular irrigation and fertilization in a phytotron at a temperature of 23–25 °C, a 16-hour photoperiod, and a photosynthetic photon flux density of 202.5–270 μmol m^{-2} s^{-1} (15,000–20,000 lux), with relative humidity maintained at 60–70 % (Terentev et al., 2023).

Hyperspectral imaging took place in a dedicated light-isolated chamber. The camera was mounted horizontally on a tripod at a distance of 0.5 m above the plant canopy. The samples were illuminated by two 500 W halogen light sources positioned at an angle of 45° relative to the imaging plane. A dark background was utilized to minimize background reflectance and enhance segmentation of plant material. The frame covered an area of 20 × 20 cm. The configuration of the imaging system was developed in accordance with previously published experimental setups (Zhu et al., 2016; Wang et al., 2018; Gu et al., 2019).

The experiments employed an Ultris 20 hyperspectral snapshot camera (Cubert GmbH, Ulm, Germany), which acquired reflectance data within the 450–874 nm spectral range, distributed over 106 narrow bands with a spectral resolution of 4 nm. The image resolution amounted to 410 × 410 pixels. Camera calibration included white reference calibration using certified reflectance panels and dark current correction by covering the lens with a light-blocking cap. Geometric calibration was also performed. Data were then exported to a personal computer via Cubert-Pilot proprietary software (version 2.8.1, Cubert GmbH, Ulm, Germany) supplied by the manufacturer (Terentev et al., 2025).

Images were acquired daily at 12:00 PM from the 13th to the 18th day after sowing, corresponding to growth stages BBCH 12–13. Each daily data set (control + infected) comprised 144 images. The complete data set for each experiment consisted of 432 images. The data were stored in multichannel TIFF format (106 channels, 16-bit depth).

The following machine learning models were applied for spectral curve classification: support vector machine (Cortes and Vapnik, 1995), logistic regression (Hosmer et al., 2013), and gradient boosting (Friedman, 2002) implemented in LightGBM (Ke et al., 2017). A detailed description of the machine learning algorithms and the explored hyperparameter ranges can be found in (Terentev et al., 2025; Fedotov et al., 2025). All computations were performed in Python 3.10 using scikit-learn, Optuna, SciPy, Pandas, and Matplotlib/Seaborn.

The following metrics were used to evaluate the performance of machine learning models: recall, precision, and F1-score (Powers, 2011). Recall measures the proportion of correctly identified positive samples:

$$Recall = \frac{TP}{TP + FN},$$

where TP is the number of true positives and FN is the number of false negatives. Accuracy measures the proportion of predicted positive samples that are actually positive:

$$Precision = \frac{TP}{TP + FP},$$

where FP is the false positive rate. The F1 score, utilized as the primary evaluation metric in this study, is the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Prescision \times Recall}{Precision + Recall}.$$

Results

Data Processing

The general scheme of hyperspectral data processing is presented in Figure 1. In the first stage (Fig. 1a), pre-processing of hyperspectral images was performed, which included sequential smoothing of spectral curves using the Savitzky–Golay method (Savitzky and Golay 1964), pixel-by-pixel normalization (curve standardization) to eliminate the influence of illumination and shooting geometry, and pixel clustering using the k-means method (K-Means) (Hartigan and Wong, 1979) to highlight homogeneous areas of the plant and eliminate the background. After clustering, the mean spectrum value was calculated for each specimen. The final operations—curve-wise and channel-wise standardization—ensure data comparability between samples. Curve standardization was performed in the form of a z-transformation. This statistical normalization converts each data element into the number of standard deviations from the mean. As a result, the standardized reflectance in some spectra took negative values. The result of this stage was pre-processed average spectral curves used for feature construction; the full pre-processing pipeline is described in detail in (Fedotov et al., 2025).

In the second stage (Fig. 1b), feature spaces were formed. First derivatives characterizing local changes in slope and spectrum shape were calculated from the processed spectral curves. In addition, extreme features reflecting the position and amplitude of key reflectance maxima and minima were extracted.

Thus, three types of input data were formed: the original average curves (Xmean), the derived spectra, and the extreme features applied in solving the problem of classifying plant species.

Figure 2 shows the averaged spectral curves for the three studied crops—wheat, rye, and barley at different stages of preprocessing. In Fig. 2a the curves are presented before normalization. At this stage, different classes spectral profiles appear highly similar in shape, which indicates that overall reflectance patterns are close and may hinder reliable class separation. After applying variance normalization (Fig. 2b),

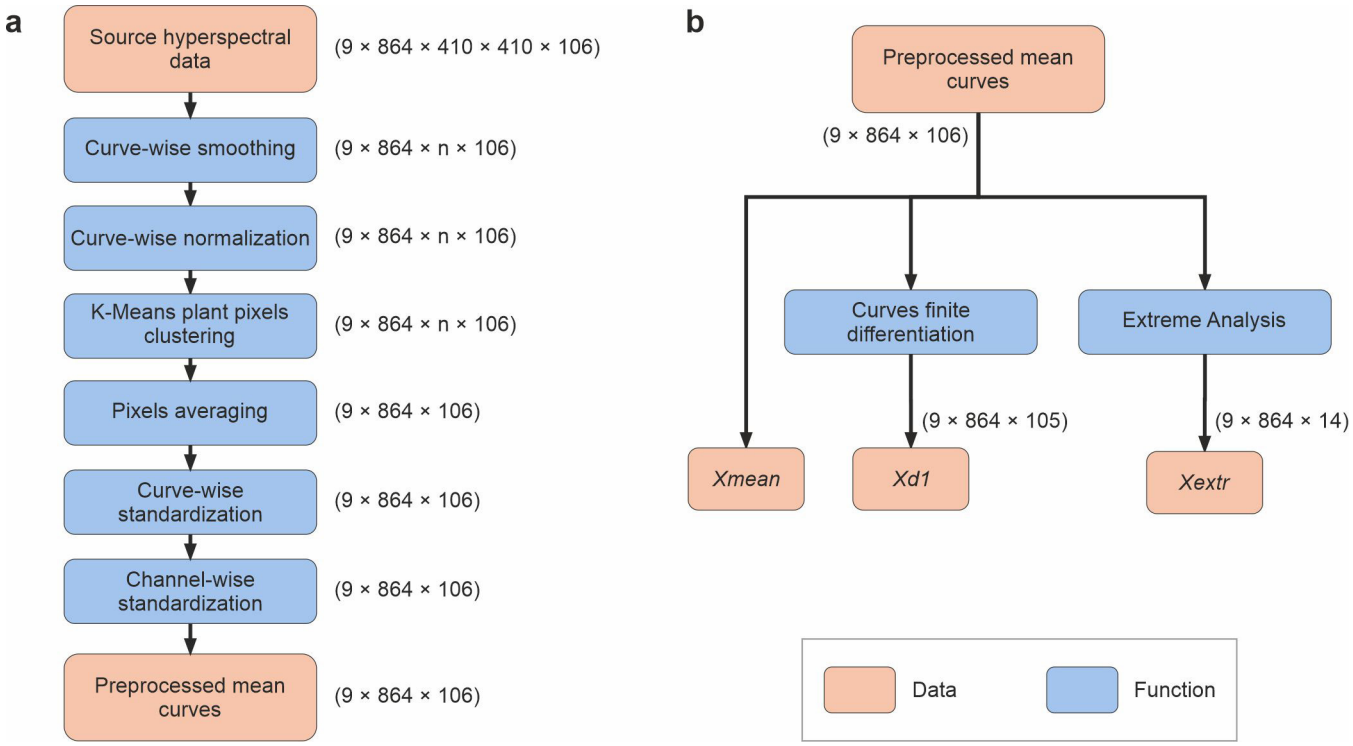

Figure 1. Diagram of the stages of processing and feature extraction from hyperspectral images: a – preliminary processing of the raw hyperspectral data; b – feature construction based on the obtained curves: calculation of derivatives, extrema, and formation of categorical features

Рисунок 1. Схема этапов обработки и формирования признаков из гиперспектральных изображений: a – этап предварительной обработки исходных гиперспектральных данных; b – этап построения признаков на основе полученных кривых: вычисление производных, экстремумов, формирование категориальных признаков

which reduces noise and scale differences influence, the curves become more comparable. As a result, subtle local differences between the classes start to emerge, particularly in specific wavelength regions.

After calculating the first derivatives (Fig. 2c), the spectral curves exhibit characteristic local extrema in the regions around 500 and 700 nm. At the same time, although the general shape of the curves remains similar, the derivative representation further enhances local variations between the classes, making class-specific features more distinguishable.

Figure 3 presents the results of an analysis of extreme features calculated from the derivatives of the spectral curves. These features reflect differences in the curves' minima and maxima, their number, and their locations along the wavelength range. Similar ranges of extreme points are observed for all three crops—around 500 nm and 750–800 nm. Minor shifts in extreme points between crops indicate minor differences in the optical properties of the leaves; however, no significant structural differences in the spectral shapes are observed.

Cross-validation

In this study, a Logistic regression model was employed as the baseline classifier for the three-class problem (wheat, barley, and rye). Logistic regression belongs to the class of linear probabilistic models and estimates posterior class probabilities based on a linear combination of input features followed by a normalization procedure. In the multiclass setting, the model simultaneously discriminates between all classes, providing a consistent probabilistic interpretation of predictions. Due to its linear structure, logistic regression offers a high degree of interpretability, allowing direct analysis of the contribution of individual spectral features to the decision function.

To assess the generalization ability of the models and select optimal hyperparameters, a nested cross-validation strategy was used. The dataset included nine independent subsamples grouped into three crop classes: three for wheat, three for barley, and three for rye. This data organization allowed the validation procedure to be implemented on completely independent data.

A cross-validation scheme was implemented. For each crop, pairs of training and validation sets were formed from the three available subsamples such that each set was ud as a validation set exactly once. A total of three independent partition combinations were obtained.

At each optimization step for a given set of hyperparameters, the model was trained on the combined data from one subsample of each class and validated on the corresponding paired subsample of the same classes. For each hyperparameter combination, the F1-score (averaged across classes) was

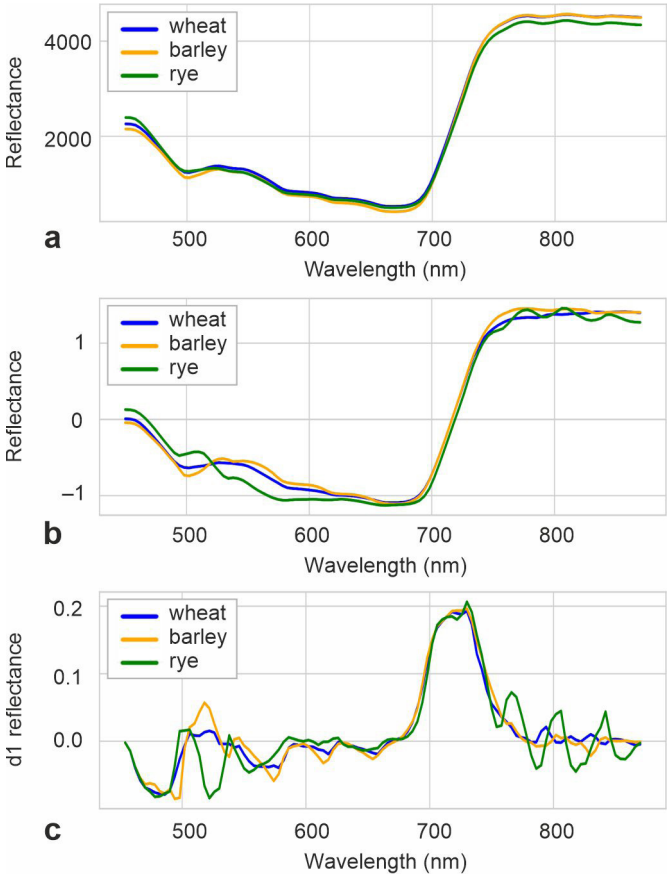

Figure 2. Average spectral reflectance curves of leaves of three crops: a – before preprocessing; b – after preprocessing, c – after first derivatives calculating

Рисунок 2. Средние спектральные кривые отражательной способности листьев трех культур. a – средние спектральные кривые до предварительной обработки, b – средние спектральные кривые после предварительной обработки, c – после вычисления первых производных

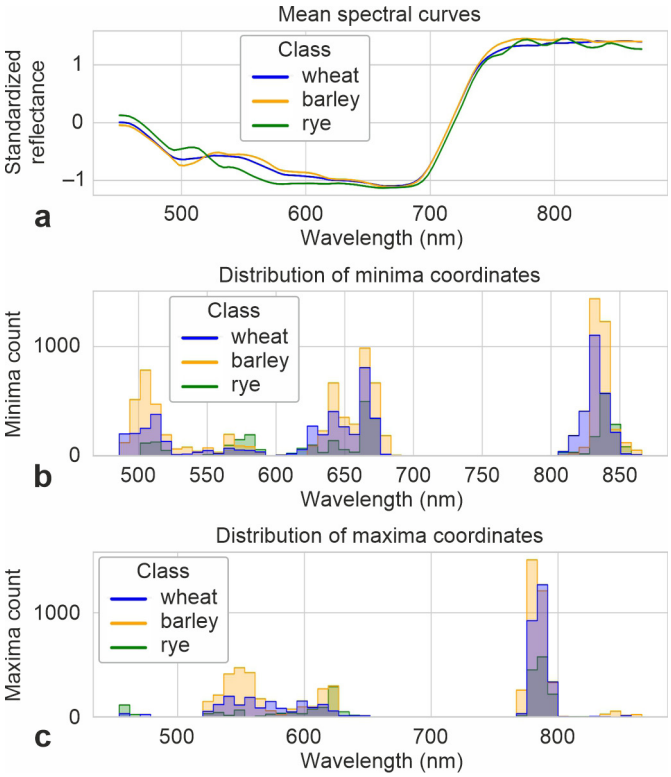

Figure 3. Distribution of spectral extremum coordinates for three crop species: a – average spectral curves after preprocessing; b and c – distribution of wavelengths corresponding to minima and maxima of the spectral curves for wheat, barley, and rye

Рисунок 3. Распределение координат спектральных экстремумов для трёх видов культур. a – средние спектральные кривые после предварительной обработки; b и c – панели показывают распределение длин волн, соответствующих минимумам и максимумам спектральных кривых для пшеницы, ячменя и ржи

calculated, and the final score was the average metric value across all three partition combinations. Optimization was performed using the Bayesian method with the objective function of maximizing the average F1-score.

After determining the optimal hyperparameters, a final evaluation of the model was conducted on independent test sets. For this, two data sets from each class were combined into a training set, with a third set retained for testing. This procedure was repeated so that each of the nine data sets was in the test set exactly once.

The model with optimal hyperparameters was trained on the combined training set and tested on the corresponding independent data set. The final model performance metrics (F1-score, precision, recall) were calculated as the average values across all three testing configurations.

Classification results

In a study on the classification of diseased and healthy plants, differences in the shape of spectral curves were a key factor in separating the classes (Terentev et al., 2025). Therefore, when constructing feature spaces, the primary focus was on analyzing the morphological features of the spectra, using first derivatives, local extrema, and other characteristics sensitive to shape changes. This approach allowed for the effective identification of differences associated with the physiological state of the plants.

As Figure 4 shows, the situation is fundamentally different in the current problem of classifying different crop species: wheat, rye, and barley. The average spectral curves of the studied classes exhibit high shape similarity, which reduces the information content of feature spaces based primarily on waveform analysis. When using such features, a significant portion of the spectral information associated with absolute reflectance and characteristic wavelength ranges may be lost, negatively impacting classification accuracy.

As a result, the best accuracy is achieved when using original or averaged spectral curves without extracting derived features, where full information on spectral amplitude is preserved. This confirms that absolute differences in spectral characteristics, rather than variations in curve shape, are decisive for plant classification.

Table 1 shows the values of the F1-score metric on the test data (test_f1) for three independent partitionss.

Table 1. Final classification results of three crop species according to F1-score, Precision, and Recall metrics on test data across three independent partitions

Таблица 1. Итоговые результаты классификации трёх видов культур по метрикам *F1-score*, *Precision* и *Recall* на тестовых данных для трёх независимых разбиений

Partition	F1-score (test)	Precision	Recall
1	0.964	0.967	0.9629
2	0.998	0.998	0.998
3	0.940	0.943	0.938

The average F1-score across all experiments was 0.967 ± 0.025. The small variation between partitions indicates model stability and the absence of overfitting. The best results were observed for the second partition (F1 = 0.998), which may be due to a more even distribution of samples across classes in the training and test sets.

Figure 5 shows the normalized confusion matrix, combining the results for all partitions. The diagonal elements of the matrix reflect the proportion of correctly classified samples for each class, while the off-diagonal elements indicate misclassifications between classes.

The classifier demonstrates a high degree of accuracy in recognizing all three plant classes. The diagonal elements of the matrix have values close to one: 0.98 for wheat, 1.00 for barley, and 0.96 for rye. Minor errors were observed in the classification of rye and wheat: approximately 2% of wheat

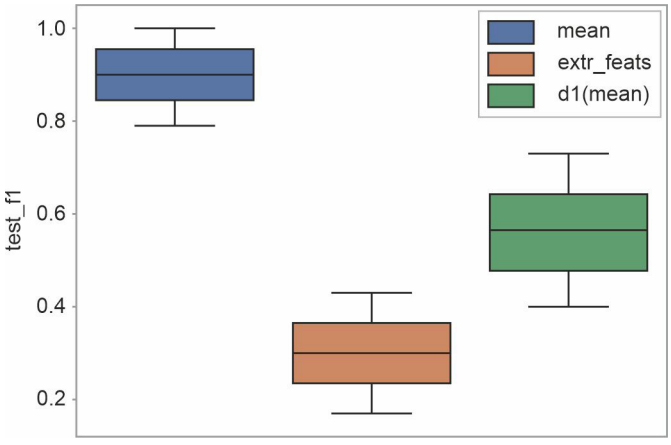

Figure 4. Classification results using different feature sets. F1-score values on the test data are presented for three types of features: *mean* – original averaged spectral curves, *d1(mean)* – first derivatives, and *extr_feats* – extreme features

Рисунок 4. Результаты классификации при использовании различных признаковых пространств. Представлены значения метрики F1-мер на тестовых данных для трёх типов признаков: *mean* – исходные усреднённые спектральные кривые, *d1(mean)* – первые производные, *extr_feats* – экстремальные признаки

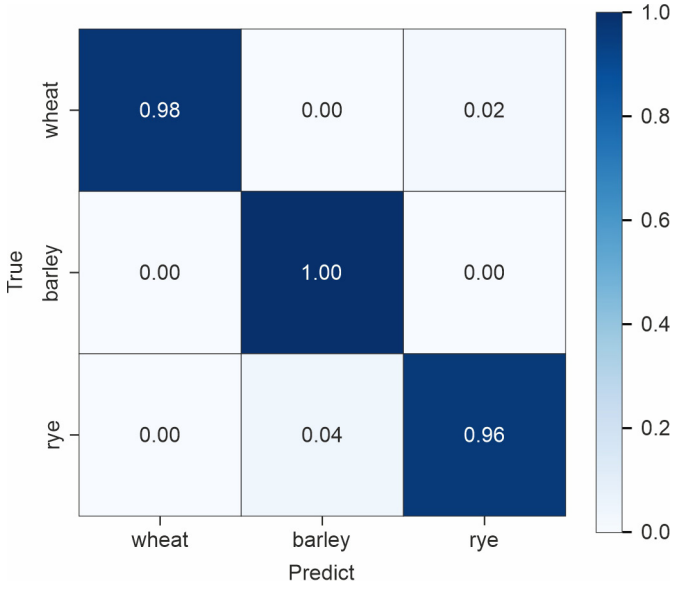

Figure 5. Normalized confusion matrix for classifying three cereal crop species. Diagonal values represent the proportion of correctly classified samples in each class

Рисунок 5. Нормализованная матрица ошибок для классификации трёх видов злаковых культур. Значения по диагонали отражают долю правильно классифицированных образцов каждого класса

samples were mistakenly classified as rye, and 4% of rye samples were mistakenly classified as barley.

Spectral Bands Contribution for Mean Curves and First-Derivative Curves

To interpret which spectral regions were primarily exploited by the classifier for decision-making, the best-performing logistic regression (LR) model was analyzed, which was selected according to the highest macro-F1 score. The analysis was restricted to the mean spectral representation, corresponding to the averaged spectral curve, as this feature space yielded the best classification performance among the considered representations.

For linear models, the informativeness of each wavelength was quantified using the absolute value of the model coefficient (|w|). In the multiclass setting, importance values were computed separately for each class based on the corresponding class-specific coefficient vector. The coefficients were subsequently normalized so that their sum across all wavelengths equaled 1, enabling interpretation of the values as proportional contributions of individual spectral regions.

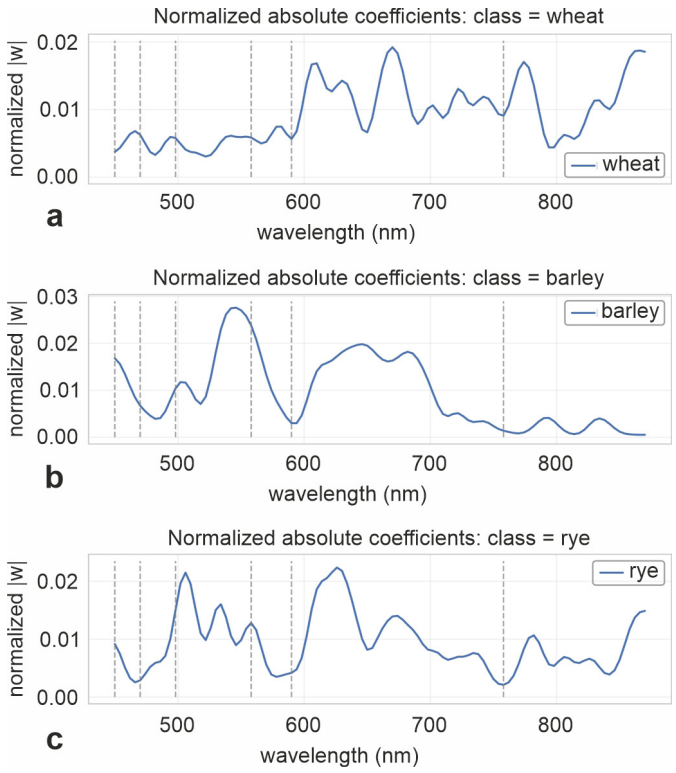

Figure 6. Profiles of the normalized feature importance $|w\lambda|$ as a function of wavelength for the logistic regression (LR) model using mean spectral curves: a – wheat; b – barley; c – rye. The vertical dashed lines indicate the boundaries of the predefined spectral ranges: blue (440–470 nm), blue-green (470–500 nm), green (500–560 nm), orange (560–590 nm), red (590–760 nm), and infrared (760–874 nm)

Рисунок 6. Профили нормализованной важности признаков $|w\lambda|$ в зависимости от длины волны для модели логистической регрессии с использованием средних спектральных кривых. a – пшеница; b – ячмень; c – рожь. Вертикальными пунктирными линиями отмечены границы спектральных диапазонов: синий (440–470 нм), сине-зеленый (470–500 нм), зеленый (500–560 нм), оранжевый (560–590 нм), красный (590–760 нм), инфракрасный (760–874 нм)

The boundaries of predefined spectral bands are indicated by vertical dashed lines in the importance profiles (Fig. 6). Aggregated contributions per spectral band are summarized as heatmaps (Fig. 7).

Figure 7 presents the normalized absolute coefficients of the logistic regression model for each class, reflecting the relative importance of individual spectral channels. The results demonstrate that the classifier relies on different spectral regions for distinguishing between the considered crops.

For the wheat class, the importance is distributed across a wide range of wavelengths, with pronounced contributions in the red and near-infrared regions (approximately 600–750 nm and above 800 nm). In contrast, the barley class exhibits a more localized importance profile, with a dominant peak in the green spectral region (approximately 520–560 nm) and moderate contributions in the red edge range (around 600–700 nm). The importance rapidly decreases in the near-infrared region, suggesting that barley is primarily distinguished by reflectance characteristics in the visible spectrum. The rye class demonstrates a mixed importance pattern, with notable contributions in both the green region (around 500–540 nm) and the red edge (approximately 600–650 nm), as well as moderate importance in the near-infrared range. This distribution suggests that rye shares spectral characteristics with both wheat and barley, leading to a less distinct importance profile.

A quantitative summary of band-wise contributions is shown in Fig. 7. For the logistic regression model using the mean spectral representation, the red band dominates across all classes (0.458–0.496), followed by the infrared region, whose contribution varies substantially between classes (from 0.054

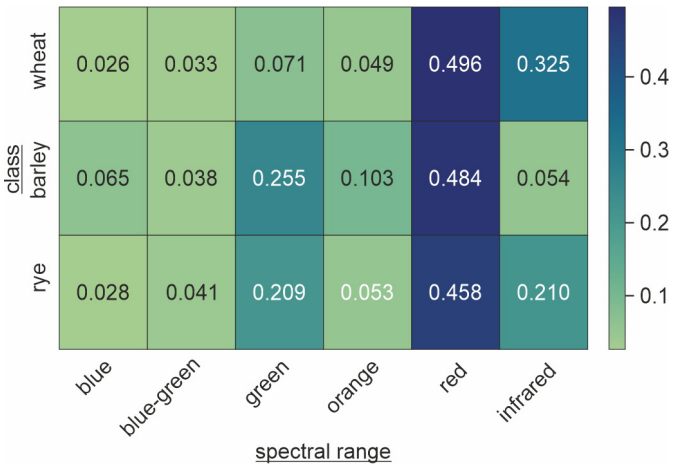

Figure 7. Distribution of the aggregated normalized feature importance $|w\lambda|$ across predefined spectral ranges for the logistic regression (LR) model based on mean spectral curves for each class (wheat, barley, rye). Spectral ranges: blue (440–470 nm), blue–green (470–500 nm), green (500–560 nm), orange (560–590 nm), red (590–760 nm), and infrared (760–874 nm)

Рисунок 7. Распределение суммарной нормированной важности признаков $|w\lambda|$ по спектральным диапазонам для модели логистической регрессии (LR), построенной на основе средних спектральных кривых, для каждого класса (пшеница, ячмень, рожь). Спектральные диапазоны: синий (440–470 нм), сине-зелёный (470–500 нм), зелёный (500–560 нм), оранжевый (560–590 нм), красный (590–760 нм) и ближний-инфракрасный (760–874 нм)

for barley to 0.325 for wheat). The green band also provides a notable contribution, particularly for barley (0.255) and rye (0.209), while its impact is smaller for wheat (0.071). The remaining spectral regions, including blue, blue–green, and orange, contribute relatively little to the overall importance (all below approximately 0.10).

Overall, the results demonstrate a consistent pattern across classes. The red region (590–760 nm) represents the primary source of discriminative information for all crops, while the contribution of the near-infrared range (760–874 nm) varies depending on the class. In particular, wheat exhibits a strong reliance on the infrared region, whereas barley is characterized by a much weaker infrared contribution and a more pronounced dependence on the green spectral range.

The analysis showed that, despite observable differences in spectral importance profiles, the spectral curves of the three crops remain similar in shape and exhibit comparable characteristics in both the visible and near-infrared ranges. As a result, class discrimination is achieved through subtle differences in the relative contributions of spectral regions rather than through large structural differences in the spectral signatures. The best classification performance was obtained using the original averaged spectral curves, indicating that absolute reflectance values contain sufficient discriminative information without the need for additional transformations.

Discussion

The obtained results revealed significant differences between hyperspectral data preprocessing strategies traditionally utilized in phytopathological diagnostics. For phenotyping healthy plants, minimal preprocessing proved superior. Various classification methods based on hyperspectral imaging have firmly established themselves as promising tools for the early detection of diseases, as they enable the identification of biochemical and structural changes in leaves long before visible symptoms appear (Mahlein et al., 2018; Nalepa, 2021). Such effects are associated with chlorophyll degradation, tissue damage, and changes in water balance, which manifest as localized alterations in spectral curve shapes (Jacquemoud, Ustin, 2019; Cheshkova et al., 2022; Tanner et al., 2022; Szechyńska-Hebda et al., 2025).

Disease detection studies often employ a standard preprocessing pipeline which includes smoothing (Savitzky–Golay filters), scatter/scale correction (MSC, SNV), continuum removal, and first- or second-derivative calculations—to reduce the influence of illumination and imaging geometry artifacts and highlight spectral features related to pigment composition and tissue structural changes during infection. These methods have been widely described in reviews and empirical studies and have demonstrated practical utility in plant pathology diagnostics (Zhang et al., 2020; Witteveen et al., 2022; Xu et al., 2022). The exploitation of such methods is particularly justified in stress analyses, where the informative signal is expressed primarily in the form of the spectral curve.

At the same time, studies focused on phenotyping and interspecies/varietal classification indicate that optimal preprocessing depends on the nature of the discriminatory information. In problems where key information is embedded in the spectral amplitude (reflectance level), aggressive normalization and differentiation can diminish the informative content. Comparative studies demonstrate that no preprocessing scheme is universally applicable, and in some cases, minimal processing or the use of raw spectra yields superior results for interspecies classification (Grzybowski et al., 2021; Wu et al., 2022; Cushnahan et al., 2024).

The impact of modern approaches on preprocessing requirements is particularly noteworthy. While classical machine learning methods benefit from carefully chosen preprocessing, modern deep neural network architectures can automatically learn robust representations and partially offset the need for manual preprocessing. Several studies have shown that deep learning methods can reduce the importance of traditional preprocessing steps; however, testing and comparing strategies for each specific application remains necessary (Paoletti et al., 2019; Bock et al., 2020; Helin et al., 2022).

In our experiments with wheat, rye, and barley, the full preprocessing scheme, previously optimized for early rust detection, offered no advantage when phenotyping healthy plants. Minimal preprocessing preserved absolute reflectance, which proved decisive for classification. These observations align with generalizations in the literature, where the impact of preprocessing on classification quality depends on the analysis objective, wavelength range, and sensor characteristics (Li et al., 2021; Witteveen et al., 2022; Cozzolino et al., 2023; Ram et al., 2024).

Our results indicate that spectral differences among the studied crops are less pronounced than those between healthy and diseased plants. Close phylogenetic relationships and similar leaf structures contribute to overlapping spectral curves, limiting interspecies discriminability (Rocchini et al., 2022; Thornley et al., 2022). In our data, spectra of diseased plants exhibit more significant shifts compared to interspecific differences within Triticeae (Terentev et al., 2025). Phytopathological processes alter not only absolute reflectance but also the wavelength distribution of reflectance, consistent with the concept of a disease-specific "spectral fingerprint," which reflects changes in absorption, scattering, and fluorescence (Mahlein, 2016). Reviews emphasize that the diagnostic value of hyperspectral data arises from spectral patterns rather than individual amplitude values (Sarić et al., 2022). Studies on feature selection and interspecies classification further confirm that informative signals are associated with local variations, gradients, and the position of characteristic inflections in the spectral curve. However, for closely related crops, absolute reflectance dominates over spectral shape differences for classification. Thus, the role of spectral curve shape aligns with the general principle of phenotyping: pathological processes alter the spectrum as a holistic structure, providing stronger class separability than interspecific differences. Consequently, in the system studied, the influence of the crop factor is subordinate to the influence of the phytopathological state, which should be considered when developing spectral diagnostic methods.

A more detailed interpretation of band-wise importance further clarifies the physiological basis of interspecific discrimination. From a physiological standpoint, the dominance of the red region (590–760 nm) for the mean spectral curves can be attributed to differences in pigment composition among

crop species. Reflectance in the red range is strongly governed by chlorophyll absorption, and even moderate interspecific differences in chlorophyll concentration or pigment ratios (chlorophyll a/b, carotenoids) may lead to measurable variations in reflectance (Jacquemoud and Baret, 1990; Gitelson et al., 2003). Such pigment-related differences likely explain why the red band provides the highest discriminative contribution when using the original mean spectral curves. In contrast, the increased importance of the near-infrared (760–874 nm) region for the first-derivative representation reflects structural differences between leaves of the studied crops. NIR reflectance is primarily controlled by internal leaf structure, including mesophyll organization, cell size, and intercellular air spaces (Slaton et al., 2001; Ustin et al., 2009). Wheat, barley, and rye exhibit species-specific anatomical traits that influence scattering properties in this spectral domain. The first derivative enhances subtle changes in spectral slope, making structural differences more pronounced and shifting the center of informativeness toward the infrared domain. Thus, the transition from red-dominated importance in mean spectra to infrared-dominated importance in first-derivative spectra is consistent with the complementary roles of pigment-related and structural features in interspecific crop discrimination. These mechanisms fundamentally differ from stress-induced spectral alterations. Pathogen-induced stress primarily affects biochemical and physiological processes, including chlorophyll degradation, chloroplast disruption, changes in water status, and modifications of cellular reflectance properties. Such processes generate localized variations in spectral shape and shifts in absorption features within the 550–750 nm and 950–1300 nm ranges (Carter, Knapp, 2001; Mahlein et al., 2018; Zhang et al., 2020). In contrast, interspecific morphological differences are largely determined by epidermal structure, leaf thickness, and internal air-space architecture, which mainly influence reflectance amplitude in the near-infrared region (Gitelson et al., 2002; Jacquemoud, Ustin, 2019). In phenotyping tasks, where discriminative information resides primarily in reflectance amplitude rather than localized spectral deformations, derivative-based transformations and intensive normalization may attenuate the informative signal and amplify noise. This explains the decrease in classification accuracy observed when applying the full phytopathology-oriented preprocessing pipeline to healthy crop discrimination.

Therefore, designing hyperspectral data processing pipelines should involve comparing multiple preprocessing options and selecting the optimal strategy based on empirical evaluation (e.g., cross-validation and holdout test sets). No single preprocessing scheme guarantees optimal results; empirical testing remains essential. In practice, researchers routinely evaluate preprocessing combinations and choose the best option based on validation metrics, making the approach empirically justified. Accordingly, when developing a hyperspectral pipeline, it is recommended to: (1) design several alternative preprocessing schemes, (2) assess their impact using cross-validation and a holdout test set, and (3) select and fix the optimal strategy for the specific problem and sensor (Li et al., 2021; Witteveen et al., 2022; Xu et al., 2022; Ram et al., 2024; Terentev et al., 2025).

The scientific novelty of this work lies in demonstrating, for the first time using experimental data from three major grain crops, that preprocessing efficiency strongly depends on the nature of the classification task. The results highlight the need for systematic domestic research to develop adaptive hyperspectral processing pipelines based on empirical observations and datasets from national research programs and grant projects. Such developments will ensure the independence and reproducibility of domestic solutions in phytosanitary monitoring and crop phenotyping. Although hyperspectral preprocessing and disease detection have been extensively studied internationally, most research focuses on early stress detection or single-species phenotyping. This study addresses the gap by systematically comparing preprocessing strategies across multiple major grain crops, highlighting task-dependent efficiency and providing empirical guidance for domestic crop phenotyping.

Despite the high accuracy achieved in classifying wheat, barley, and rye using raw spectral curves, this study has several limitations. First, the experiments were conducted under controlled laboratory conditions, which may not fully represent field variability. Second, only three crop species were considered, and further testing is needed to evaluate the generalizability of the preprocessing pipeline to additional cereals or varieties. Future work will focus on extending the approach to field conditions, exploring additional crop species, and integrating deep learning models to potentially reduce manual preprocessing requirements while maintaining high classification accuracy.

Conclusions

The study showed that the effectiveness of hyperspectral data processing algorithms is determined not by fixed preprocessing rules, but by the specific characteristics of the task at hand. For classifying healthy barley, rye, and wheat plants, using raw or minimally processed spectra is optimal, as the absolute reflectance values contain key discriminatory information. In contrast, for the early diagnosis of diseases, where changes are physiological and biochemical in nature, methods based on spectral curve shape analysis and derived features prove to be most effective. Therefore, when designing applied hyperspectral data processing pipelines, it is important to consider the study objective and sensor properties, to test multiple preprocessing strategies and select the one that performs best according to empirical metrics.

Acknowledgements

The study was conducted using the equipment of the Shared Research Facility “Innovative Plant Protection Technologies” at the All-Russian Institute of Plant Protection (VIZR). This research was funded by the Russian Science Foundation (RSF), grant number 25-26-00188.

References

Гузенко АЮ, Солонкин АВ, Донцова АА (2024) Сравнительный анализ фотосинтетического потенциала новых сортов ярового ячменя в зоне засушливого климата Нижнего Поволжья. *Зерновое хозяйство России* 16(2):88–97. http://doi.org/10.31367/2079-8725-2024-91-2-88-97

Лысов АК, Павлюшин ВА (2022) Фитосанитарное проектирование агроэкосистем и дистанционное зондирование. *СПДЗЗ из космоса* 19(5):101–112. http://doi.org/10.21046/2070-7401-2022-19-5-101-109

Сысуев ВА, Уткина ЕИ, Шешегова ТК (2025) Озимая рожь – основа здоровья населения и продовольственной независимости России. *Вестник РАН* (1):16–27. http://doi.org/10.7868/S3034520025010026

Якушев ВП, Якушев ВВ, Блохина СЮ, Блохин ЮИ et al (2022) Перспективы выявления идентификационных показателей состояния посевов по аэрокосмическим снимкам и специализированным полевым исследованиям. *СПДЗЗ из космоса* 19(4):113–127. http://doi.org/10.21046/2070-7401-2022-19-4-113-127

Bock CH, Chiang KS, Del Ponte EM (2022) Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods to maximize accuracy. *Trop Plant Pathol* 47(1):25–42. http://doi.org/10.1007/s40858-021-00439-z

Bock CH, Barbedo JGA, del Ponte EM, Bohnenkamp D et al (2020) From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. *Phytopathol Res* 2(1):9. http://doi.org/10.1186/s42483-020-00049-8

Carter GA, Knapp AK (2001) Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am J Bot* 88(4):677–684. http://doi.org/10.2307/2657068

Cheshkova AF (2022) A review of hyperspectral image analysis techniques for plant disease detection and identification. *Vavilov J Genet Breed* 26(2):202. http://doi.org/10.18699/vjgb-22-25

Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. http://doi.org/10.1023/A:1022627411411

Cozzolino D, Williams PJ, Hoffman LC (2023) An overview of pre-processing methods available for hyperspectral imaging applications. *Microchem J* 193:109129. http://doi.org/10.1016/j.microc.2023.109129

Cushnahan TA, Grafton MCE, Pearson D, Ramilan T (2024) Hyperspectral data can differentiate species and cultivars of C3 and C4 turf despite measurable diurnal variation. *Remote Sens* 16(17):3142. http://doi.org/10.3390/rs16173142

Danilov R, Kremneva O, Sereda I, Gasiyan K et al (2024) Study of the spectral characteristics of crops of winter wheat varieties infected with pathogens of leaf diseases. *Plants* 13(14):1892. http://doi.org/10.3390/plants13141892

Farber C, Kurouski D, Leiva V, Sanchez L et al (2019) Advanced spectroscopic techniques for plant disease diagnostics: a review. *TrAC Trends Anal Chem* 118:43–49. http://doi.org/10.1016/j.trac.2019.05.022

Fedotov A, Eremenko D, Kuznetsova D, Baranova O et al (2025) Advanced hyperspectral image processing and machine learning approaches for early detection of wheat stem rust. *Front Plant Sci* 16:1725017. http://doi.org/10.3389/fpls.2025.1725017

Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38(4):367–378. http://doi.org/10.1016/S0167-9473(01)00065-2

Gavrilenko TA, Khiutti AV, Klimenko NS, Antonova OY et al (2021) Phenotypic and DNA marker-assisted characterization of Russian potato cultivars for resistance to potato cyst nematodes. *Agronomy* 11(12):2400. http://doi.org/10.3390/agronomy11122400

Gitelson AA, Zur Y, Chivkunova OB, Merzlyak MN (2002) Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochem Photobiol* 75(3):272–281. http://doi.org/10.1562/0031-8655(2002)0750272ACCIPL2.0.CO2

Gitelson AA, Viña A, Ciganda V, Rundquist DC et al (2003) Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys Res Let* 30(5):1248. http://doi.org/10.1029/2002GL016450

Gu Q, Sheng L, Zhang T, Lu Y et al (2019) Early detection of tomato spotted wilt virus infection in tobacco using the hyperspectral imaging technique and machine learning algorithms. *Comput Electron Agric* 167:105066. http://doi.org/10.1016/j.compag.2019.105066

Grzybowski M, Wijewardane NK, Atefi A, Ge Y et al (2021) Hyperspectral reflectance-based phenotyping for quantitative genetics in crops: progress and challenges. *Plant Commun* 2(4):100209. http://doi.org/10.1016/j.xplc.2021.100209

Hartigan JA, Wong MA (1979) Algorithm AS 136: A K-Means clustering algorithm. *J R Stat Soc Ser C Appl Stat* 28:100–108. http://doi.org/10.2307/2346830

Helin R, Indahl UG, Tomic O, Liland KH (2022) On the possible benefits of deep learning for spectral preprocessing. *J Chemometr* 36(2):e3374. http://doi.org/10.1002/cem.3374

Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression. Hoboken: John Wiley and Sons. http://doi.org/10.1002/9781118548387

Jacquemoud S, Baret F (1990) PROSPECT: A model of leaf optical properties spectra. *Remote Sens Environ* 34(2):75–91. http://doi.org/10.1016/0034-4257(90)90100-Z

Jacquemoud S, Ustin SL (2019) Leaf optical properties. In: Hyperspectral Remote Sensing of Vegetation. Cambridge: Cambridge University Press. 83–108. http://doi.org/10.1017/9781108686457

Ke G, Meng Q, Finley T, Wang T et al (2017) LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 30

Khan A, Vibhute AD, Mali S, Patil CH et al (2022) A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. *Ecol Inform* 69:101678. http://doi.org/10.1016/j.ecoinf.2022.101678

Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. *Sensors* 14(11):20078–20111. http://doi.org/10.3390/s141120078

Li YH, Tan X, Zhang W, Jiao QB et al (2021) Research and application of several key techniques in hyperspectral image preprocessing. *Front Plant Sci* 12:627865. http://doi.org/10.3389/fpls.2021.627865

Lichtenthaler HK, Gitelson A, Lang M (1996) Non-destructive determination of chlorophyll content of leaves of a green and an aurea mutant of tobacco by reflectance measurements.

J Plant Physiol 148(3–4):483–493. http://doi.org/10.1016/S0176-1617(96)80283-5

Lowe A, Harrison N, French AP (2017) Hyperspectral image analysis techniques for detection and classification of early onset of plant disease and stress. *Plant Methods* 13(1):80. http://doi.org/10.1186/s13007-017-0233-z

Mahlein AK (2016) Plant disease detection by imaging sensors: parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis* 100(2):241–251. http://doi.org/10.1094/PDIS-03-15-0340-FE

Mahlein AK, Kuska MT, Behmann J, Polder G et al (2018) Hyperspectral sensors and imaging technologies in phytopathology: state of the art. *Annu Rev Phytopathol* 56(1):535–558. http://doi.org/10.1146/annurev-phyto-080417-050100

Mishra KB, Mishra A, Klem K, Govindjee (2016) Plant phenotyping: a perspective. *Indian J Plant Physiol* 21(4):514–527. http://doi.org/10.1007/s40502-016-0271-y

Mishra P, Lohumi S, Khan HA, Nordon A (2020) Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. *Comput Electron Agric* 178:105780. http://doi.org/10.1016/j.compag.2020.105780

Moghimi A, Yang C, Marchetto PM (2018) Ensemble feature selection for plant phenotyping: a journey from hyperspectral to multispectral imaging. *IEEE Access* 6:56870–56884. http://doi.org/10.1109/ACCESS.2018.2872801

Nalepa J (2021) Recent advances in multi- and hyperspectral image analysis. *Sensors* 21(18):6002. http://doi.org/10.3390/s21186002

Paoletti ME, Haut JM, Plaza J, Plaza A (2019) Deep learning classifiers for hyperspectral imaging: a review. *ISPRS J Photogramm Remote Sens* 158:279–317. http://doi.org/10.1016/j.isprsjprs.2019.09.006

Pieruschka R, Schurr U (2019) Plant phenotyping: past, present, and future. *Plant Phenomics* 2019:7507131. http://doi.org/10.34133/2019/7507131

Powers DMW (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol* 2(1):37–63. http://doi.org/10.48550/arXiv.2010.16061

Ram BG, Oduor P, Igathinathane C, Howatt K et al (2024) A systematic review of hyperspectral imaging in precision agriculture: analysis of its current state and future prospects. *Comput Electron Agric* 222:109037. http://doi.org/10.1016/j.compag.2024.109037

Rocchini D, Santos MJ, Ustin SL, Féret JB et al (2022) The spectral species concept in living color. *J Geophys Res Biogeosci* 127(9):e2022JG007026. http://doi.org/10.1029/2022JG007026

Sarić R, Nguyen VD, Burge T, Berkowitz O et al (2022) Applications of hyperspectral imaging in plant phenotyping. *Trends Plant Sci* 27(3):301–315. http://doi.org/10.1016/j.tplants.2021.12.003

Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627–1639. http://doi.org/10.1021/ac60214a047

Slaton MR, Hunt ER Jr, Smith WK (2001) Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am J Bot* 88(2):278–284. http://doi.org/10.2307/2657019

Szechyńska-Hebda M, Hołownicki R, Doruchowski G, Sas K et al (2025) Application of hyperspectral imaging for early detection of pathogen-induced stress in cabbage as case study. *Agronomy* 15(7):1516. http://doi.org/10.3390/agronomy15071516

Tanner F, Tonn S, de Wit J, Van den Ackerveken G et al (2022) Sensor-based phenotyping of above-ground plant-pathogen interactions. *Plant Methods* 18(1):35. http://doi.org/10.1186/s13007-022-00853-7

Terentev A, Dolzhenko V, Fedotov A, Eremenko D (2022) Current state of hyperspectral remote sensing for early plant disease detection: a review. *Sensors* 22(3):757. http://doi.org/10.3390/s22030757

Terentev A, Dolzhenko V (2023) Can metabolomic approaches become a tool for improving early plant disease detection and diagnosis with modern remote sensing methods? A review. *Sensors* 23(12):5366. http://doi.org/10.3390/s23125366

Terentev A, Badenko V, Shaydayuk E, Emelyanov D et al (2023) Hyperspectral remote sensing for early detection of wheat leaf rust caused by Puccinia triticina. *Agriculture* 13(6):1186. http://doi.org/10.3390/agriculture13061186

Terentev A, Kuznetsova D, Fedotov A, Baranova O et al (2025) Cross-Crop Transferability of Machine Learning Models for Early Stem Rust Detection in Wheat and Barley Using Hyperspectral Imaging. *Plants* 14(21):3265. http://doi.org/10.3390/plants14213265

Thornley RH, Verhoef A, Gerard FF, White K (2022) The feasibility of leaf reflectance-based taxonomic inventories and diversity assessments of species-rich grasslands: a cross-seasonal evaluation using waveband selection. *Remote Sens* 14(10):2310. http://doi.org/10.3390/rs14102310

Ustin SL, Gitelson AA, Jacquemoud S, Schaepman M et al (2009) Retrieval of foliar information about plant pigment systems from high-resolution spectroscopy. *Remote Sens Environ* 113(S1):S67–S77. http://doi.org/10.1016/j.rse.2008.10.019

Wan L, Li H, Li C, Wang A et al (2022) Hyperspectral sensing of plant diseases: principle and methods. *Agronomy* 12(6):1451. http://doi.org/10.3390/agronomy12061451

Wang D, Vinson R, Holmes M, Seibel G et al (2018) Early tomato spotted wilt virus detection using hyperspectral imaging technique and outlier removal auxiliary classifier generative adversarial nets (OR-AC-GAN). *Proc ASABE Annu Int Meeting*. http://doi.org/10.13031/aim.201800660

Witteveen M, Sterenborg HJ, van Leeuwen TG, Aalders MC et al (2022) Comparison of preprocessing techniques to reduce nontissue-related variations in hyperspectral reflectance imaging. *J Biomed Opt* 27(10):106003–106003. http://doi.org/10.1117/1.JBO.27.10.106003

Wu H, Song Z, Niu X, Liu J et al (2022) Classification of Toona sinensis young leaves using machine learning and UAV-borne hyperspectral imagery. *Front Plant Sci* 13:940327. http://doi.org/10.3389/fpls.2022.940327

Xu Y, Wu W, Chen Y, Zhang T et al (2022) Hyperspectral imaging with machine learning for non-destructive classification of Astragalus membranaceus var. mongholicus, Astragalus membranaceus, and similar seeds. *Front Plant Sci* 13:1031849. http://doi.org/10.3389/fpls.2022.1031849

Yakushev VP, Kanash EV, Rusakov DV, Yakushev VV et al (2022) Correlation dependences between crop reflection indices, grain yield and optical characteristics of wheat leaves at different nitrogen level and seeding density. *Sel'skokhozyaistvennaya Biologiya* 57(1):98–112. http://doi.org/10.15389/agrobiology.2022.1.98eng

Zhang J, Huang Y, Pu R, Gonzalez-Moreno P et al (2019) Monitoring plant diseases and pests through remote sensing technology: a review. *Comput Electron Agric* 165:104943. http://doi.org/10.1016/j.compag.2019.104943

Zhang N, Yang G, Pan Y, Yang X et al (2020) A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sens* 12(19):3188. http://doi.org/10.3390/rs12193188

Zhu H, Cen H, Zhang C, He Y (2016) Early detection and classification of tobacco leaves inoculated with tobacco mosaic virus based on hyperspectral imaging technique. *Proc ASABE Annu Int Meeting*. http://doi.org/10.13031/aim.20162460422

Zimnyakov VM, Kurochkin AA, Bogomazov SV, Varlamova EN (2020) Wheat production in Russia. *Volga Region Farmland* 1(5):12–16. http://doi.org/10.36461/np.2020.54.1.003

Zolotukhina A, Machikhin A, Guryleva A, Gresis V et al (2023) Extraction of chlorophyll concentration maps from AOTF hyperspectral imagery. *Front Environ Sci* 11:1152450. http://doi.org/10.3389/fenvs.2023.1152450

Zolotukhina A, Machikhin A, Guryleva A, Gresis V et al (2024) Evaluation of leaf chlorophyll content from acousto-optic hyperspectral data: a multi-crop study. *Remote Sens* 16(6):1073. http://doi.org/10.3390/rs16061073

Translation of Russian References

Guzenko AYu, Solonkin AV, Dontsova AA (2024) Comparative analysis of the photosynthetic potential of new spring barley varieties in the arid climate zone of the Lower Volga. *Zernovoe khozyaystvo Rossii* 16(2):88–97. http://doi.org/10.31367/2079-8725-2024-91-2-88-97

Lysov AK, Pavlyushin VA (2022) Phytosanitary design of agroecosystems and remote sensing. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa* 19(5):101–112. http://doi.org/10.21046/2070-7401-2022-19-5-101-109

Sysuev VA, Utkina EI, Sheshegova TK (2025) Winter rye – the basis of population health and food independence of Russia. *Vestnik Rossiyskoy akademii nauk* (1):16–27. http://doi.org/10.7868/S3034520025010026

Yakushev VP, Yakushev VV, Blokhina SY, Blokhin YI et al (2022) Prospects for identifying indicators of crop status using satellite imagery and specialized field studies. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa* 19(4):113–127. http://doi.org/10.21046/2070-7401-2022-19-4-113-127

Вестник защиты растений, 2026, 109(1), с. 78–88

OECD+WoS: 4.01+AM (Agronomy); 1.06+QU (Plant Sciences) https://doi.org/10.31993/2308-6459-2026-109-1-17459

Полнотекстовая статья

ОЦЕНКА ПРИМЕНИМОСТИ КОНВЕЙЕРА ОБРАБОТКИ ГИПЕРСПЕКТРАЛЬНЫХ ДАННЫХ, РАЗРАБОТАННОГО ДЛЯ РАННЕЙ ДИАГНОСТИКИ РЖАВЧИННЫХ ЗАБОЛЕВАНИЙ ЗЕРНОВЫХ, ДЛЯ ФЕНОТИПИРОВАНИЯ ПШЕНИЦЫ, РЖИ И ЯЧМЕНЯ

Д.В. Кузнецова[1], О.А. Баранова[2], Д.А. Емельянов[2], Д.Ю. Еременко[1], А.А. Федотов[1,2], А.Б. Терентьев[2]*

[1]*Политехнический университет им. Петра Великого, Санкт-Петербург*

[2]*Всероссийский научно-исследовательский институт защиты растений, Санкт-Петербург*

**ответственный за переписку, e-mail: admin@vizr.spb.ru*

Представлены результаты оценки применимости конвейера обработки данных гиперспектрального зондирования, изначально разработанного для ранней диагностики ржавчинных заболеваний зерновых культур, к задаче фенотипирования здоровых растений пшеницы *Triticum aestivum*, ячменя *Hordeum vulgare* и ржи *Secale cereale*. Использованы гиперспектральные снимки здоровых растений, полученные в лабораторных условиях с применением камеры Cubert Ultris 20 (диапазон 450–874 нм, 106 каналов). Проведено сравнение эффективности различных схем предобработки — полной (включающей нормализацию, сглаживание, расчёт производных и выделение экстремальных признаков), сокращённой и минимальной. Для классификации применялись модели машинного обучения — логистическая регрессия, метод опорных векторов и градиентный бустинг, обученные на усреднённых спектрах. Показано, что применение полного конвейера, оптимизированного для задач фитопатологической диагностики, приводит к снижению точности классификации в задачах фенотипирования. Наилучшие результаты (F1 = 0.97 ± 0.025) достигнуты при использовании исходных усреднённых спектральных кривых без дополнительных преобразований. Сделан вывод, что для задач фенотипирования здоровых растений пшеницы, ячменя и ржи информативны абсолютные уровни отражения, тогда как для диагностики заболеваний — изменения формы спектральной кривой. Полученные результаты уточняют границы применимости конвейеров, разработанных для фитосанитарных целей, и могут быть использованы при создании систем дистанционного мониторинга и фенотипирования злаковых культур.

Ключевые слова: рожь, ячмень, пшеница, гиперспектральная съёмка, конвейер обработки данных, фенотипирование, диагностика заболеваний растений

Поступила в редакцию: 20.10.2025 *Принята к печати: 01.03.2026*