

Received 7 May 2025, accepted 22 May 2025, date of publication 2 June 2025, date of current version 10 June 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3575699

RESEARCH ARTICLE

Deep Spectral-Spatial Transformer for Robust Hyperspectral Image Segmentation in Varying Field Conditions

ANDREY MAKAROV^{1,2}, ISLOMBEK MIRPULATOV³, NIKITA FIRSOV¹,
VALERIY LOBANOV^{1,4}, SVETLANA ILLARIONOVA^{1,3,5}, VLADIMIR PODLIPNOV¹,
YULIYA VYBORNOVA¹, DMITRII SHADRIN^{1,6}, ANDREY RASTORGUEV¹,
ROMAN SKIDANOV^{1,2}, EVGENY BURNAEV^{1,3,7}, AND ARTEM NIKONOROV¹

¹Samara National Research University, 443086 Samara, Russia

²Institute of Image Processing Systems, NRC “Kurchatov Institute,” 123098 Samara, Russia

³Skolkovo Institute of Science and Technology (Skoltech), 121205 Moscow, Russia

⁴Adyghe State University, 385000 Maykop, Russia

⁵Biomedically Informed Artificial Intelligence Laboratory (BIMAI-Lab), University of Sharjah, Sharjah, United Arab Emirates

⁶Institute of Information Technology and Data Science, Irkutsk National Research Technical University, 664074 Irkutsk, Russia

⁷Autonomous Non-Profit Organization Artificial Intelligence Research Institute (AIRI), 105064 Moscow, Russia

Corresponding author: Svetlana Illarionova (s.illarionova@skoltech.ru)

This work was supported by the Ministry of Economic Development of Russian Federation under Agreement 000000C313925P3U0002 and Agreement No 139-15-2025-003 Dated 16.04.2025.

ABSTRACT Hyperspectral imaging serves as a powerful tool for environmental studies, enabling the capture of significant properties in the objects being analyzed. These hidden properties may not be discernible through traditional RGB analysis, thus enhancing the distinguishability of target classes. Machine learning (ML) algorithms are widely utilized for processing hyperspectral data in tasks such as image classification or segmentation. Despite the rich hyperspectral feature space, ML-based approaches often struggle with the challenges posed by varying outdoor lighting conditions, leading to instability in the developed solutions. Furthermore, traditional segmentation methods designed for RGB images need to be adapted to ensure robust performance when applied to hyperspectral images (HSI). This work proposes a robust method for the classification of HSI in the outdoor environment under varying lighting conditions. A calibration procedure for hyperspectral sensor sensitivity has been proposed, ensuring improved quality of input data and, consequently, enhanced performance of models in solving hyperspectral image classification tasks. The paper also proposes a novel architecture, the Deep Spectral Spatial Transformer (DSST), which is employed as the classifier. This architecture leverages a deeper feature extractor to ensure robustness in various lightning conditions. Among these, standardization within individual channels yielded the best results. The applicability of the proposed approach is validated in the weed detection task in an agricultural field. The data collected were captured using a pushbroom hyperspectral sensor. A series of comparative experiments were conducted on a range of classification algorithms, encompassing both traditional ML algorithms and contemporary neural network architectures developed in recent years. The proposed DSST architecture demonstrated superior performance metrics, with values of 0.911 and 0.907 on F1-Score and Accuracy, respectively.

INDEX TERMS Hyperspectral imaging, computer vision, precision agriculture, transformers, pushbroom, machine learning.

I. INTRODUCTION

The associate editor coordinating the review of this manuscript and approving it for publication was Wei He.

Hyperspectral imaging enables the capture of spectral information across different wavelengths, facilitating a

comprehensive analysis of the reflectance properties of various surfaces. This technology relies on the quantification of light acquired by photodiodes and its conversion into hundreds or thousands of bands [1]. Two primary types of sensors are utilized for HSI acquisition: charge-coupled devices (CCD) and complementary metal-oxide-semiconductor (CMOS), which differ in their processing of incoming energy [2]. Recent advancements in sensing techniques and device development have sparked increased interest in the utilization of hyperspectral imaging across various scientific and commercial domains, owing to their enhanced availability, lightweight nature, and accuracy. The hyperspectral imaging techniques surpass traditional RGB sensing by enabling spectrum analysis beyond the visible spectrum, thereby facilitating improved differentiation between objects and materials of both natural and artificial origin.

The hyperspectral imaging techniques are employed in various applications where visual similarity between considered classes occurs. The distanciation of artificial materials based on spectrum signatures is implemented for the separation of polyolefin waste plastics, which is not always feasible in the visible spectral range [3]. Currently, the hyperspectral imaging devices are widely integrated on board uncrewed aerial vehicles (UAVs) for spatial data mining and land cover analysis. For example, they can be utilized in geological applications and mineral exploration studies [4]. In forestry research, hyperspectral imaging is used to estimate various inventory properties such as dominant species, age, height, and growing stock volume, which are challenging to estimate based solely on RGB scanning [5]. Additionally, the hyperspectral imaging plays a significant role in precision agriculture by classifying plant types and assessing growing conditions for informed decision-making by farmers [6].

Similar to general RGB images, HSI can be processed automatically to facilitate quantitative and qualitative assessment of studied objects and phenomena. The typical pipeline for conducting analysis based on the hyperspectral imaging involves the following steps: data collection, imagery processing, algorithm development, and its subsequent application. The first stage usually involves spectral calibration and correction tasks. Imagery pre-processing addresses problems such as noise reduction, data standardization, and extraction of non-mixed spectra [7].

For the automatic extraction of desired information, classical computer vision approaches or ML techniques can be employed [8]. Classical approaches often rely on hyperspectral vegetation indices that use physically-informed formulations to measure structural, biochemical, and physiological parameters of surfaces [9]. These indices are computed using combinations of multiple bands and compared against empirically predefined thresholds to classify surface types. ML algorithms, when trained with labeled data, offer another approach for information extraction, with varying levels of complexity and data handling specificity.

The task of classifying HSI involves the per-pixel mapping of hyperpixels in a HSI to class labels. To solve this task, classical ML methods such as SVM [10], [11], Random Forest [12], [13], KNN [14], [15], and others can be applied. Despite the rich spectral information provided by the high dimensionality of the hyperspectral imaging data, specialized algorithms are necessary for processing this data type. Conventional ML and deep learning (DL) algorithms developed for general computer vision tasks often struggle to handle the hyperspectral imaging data accurately. The use of advanced DL approaches has significantly improved classification performance, including the use of convolutional networks employing 1D, 2D, 3D convolutions, as well as their combinations [16], [17], [18], [19], [20]. Meta-learning [21], GANs [22] and KANs [19] based approaches are also used, showing high classification quality. Recently, approaches based on transformer architectures have been gaining popularity [23], [24], [25].

Another challenge arises from variations in lighting conditions, particularly in outdoor (field) environments, which introduce additional processing instabilities in captured spectrum measurements. This challenge becomes more pronounced with small datasets where obtaining a sufficient number of samples is difficult. While image augmentations in general computer vision are designed to address this problem [26], applying color augmentation straightforwardly to multispectral or the hyperspectral imaging data is complex due to intricate spectrum connections [27]. Consequently, several studies focus on minimizing the impact of external conditions on acquired the hyperspectral imaging data through pixel- and channel-wise data standardization [28].

In this study, our aim is to address several key challenges encountered in the hyperspectral imaging data analysis. The first challenge involves developing a customized DL architecture to solve the image segmentation task for the hyperspectral imaging with a limited number of training samples. Another challenge arises from the diverse outdoor (field) conditions influenced by weather and lighting. Lastly, our objective is to establish a pipeline for HSI data processing and segmentation. For the experiments we used data obtained by a pushbroom hyperspectrometer built according to the Offner scheme (hereinafter we will use the term - "hyperspectrometer"). This type of hyperspectrometer is more affordable and allows for the acquisition of a significantly larger number of spectral channels compared to full-frame hyperspectrometers, making it advantageous for precision agriculture applications [29], [30]. For spectral calibration of the HSI, we used the ColorChecker Video as a cost-effective alternative to the Spectralon diffuse reflectance standard. To achieve invariance to lighting conditions, we developed and proposed a transformer architecture with a spatial-spectral feature extractor (Figure 1). The experiments included comparative analysis of several existing approaches based on ML and DL algorithms. The main contributions of the study are the following:

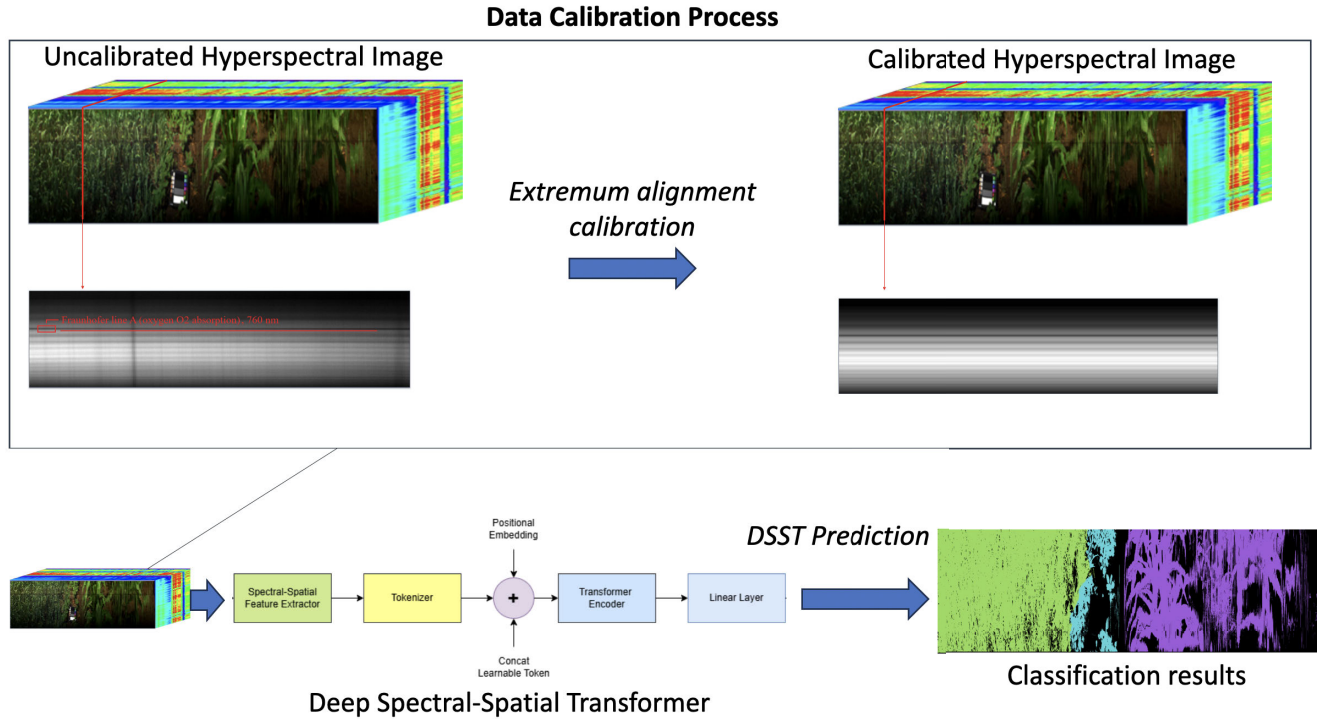


FIGURE 1. Proposed study workflow for hyperspectral image segmentation using deep spectral-spatial transformer.

- We proposed an effective pipeline for HSI analysis using ML-based algorithms for lighting invariance improvement;
- We explored different preprocessing techniques for HSI analysis;
- We proposed a hyperspectral sensor sensitivity calibration algorithm that improves the quality of the acquired data and, consequently, enhances the performance of classification models;
- We conducted comprehensive study with different ML and DL algorithms;
- We proposed an novel deep spatial-spectral architecture for HSI data analysis based on transformer block (DSST), which is designed to enhance the robustness of the system.

II. MATERIALS AND METHODS

A. DATASET COLLECTION AND PREPROCESSING

1) STUDY AREA

The study area for the survey is located in the Samara region, Privolzhsky district, Privolzhye, in the fields of Agriculture Holding Tsurulev E.P., near the coordinates latitude 52.81, longitude 48.61. The survey dates were from July 10 to 20, 2021. The selected survey areas included cultivated and irrigated plots, predominantly with a homogeneous distribution of a single agricultural crop across the survey area, as well as areas where several agricultural crops bordered each other.

2) HYPERSPECTRAL IMAGING

For the experiments we used data obtained by a pushbroom hyperspectrometer built according to the Offner scheme (hereinafter we will use the term - hyperspectrometer). A hyperspectral sensor mounted on a special tripod equipped with an angular rotation drive capable of setting the rotation speed in the range of 0.2 to 3 revolutions per minute. The hyperspectrometer was positioned during the survey so that the slit aperture was perpendicular to the spatial scanning vector. This hyperspectrometer is described in detail and was used in [31]. Each object has a unique spectral reflection characteristic, allowing for the unambiguous identification of the materials that make up its structure. In addition to the two standard coordinates, a spectral coordinate is added to each pixel of the hyperspectral image, providing a three-dimensional data space. In other words, each point of the image corresponds to a spectrum obtained at that point of the imaged object.

The spectrum $S_i(\mathbf{x})$ (a hyperpixel at the point $\mathbf{x}$) registered in the HSI [32], [33] at a given point is formed as follows:

$$S_i(\mathbf{x}) = \int_{\lambda_i}^{\lambda_i + \Delta\lambda_i} I(\lambda)R(\mathbf{x}, \lambda)T_S(\lambda)d\lambda \quad (1)$$

where

- λ is the wavelength;
- λ_i is the wavelength i -th spectral channel of the HSI;
- $\Delta\lambda_i$ is the width of the i -th spectral registration interval;
- T is the spectral transmittance of the sensor of the registering the hyperspectral imaging camera;



FIGURE 2. Color synthesis of HSI 1.



FIGURE 3. Color synthesis of HSI 2.

- R is the object's reflection spectrum;
- I is the spectrum of light source. The range of wavelengths is determined by the spectral sensitivity range of the registering camera.

Assuming equal spectral registration intervals of the HSI, in Equation 1 we obtain:

$$S_i(\mathbf{x}) = \int_{\lambda_i}^{\lambda_i + \Delta\lambda} I(\lambda)R(\mathbf{x}, \lambda)T_S(\lambda)d\lambda \quad (2)$$

Two HSI of agricultural crops obtained by the pushbroom hyperspectrometer are presented in Figures 2 and 3. These images have resolution of 960×3000 pixels and 250 spectral channels, with a spectral resolution of approximately 2.5 nm in the range of 420-980 nm. The study location contains three vegetation classes: oats, corn, and the weed species *Amaranthus* growing between them.

The tripod is also equipped with a mechanical device that allows setting various angles of inclination of the camera relative to the subject being photographed. The imaging was conducted under partially cloudy weather conditions with moderate wind speeds of 2 to 4 m/s. The camera's height above ground level was 1.9 meters. For capturing

objects, a lens with a fixed focal length MIR-1V 2.8/37 (Russia) was chosen, with the aperture set at f/3.2. The selection of this lens was based on its sufficient field of view from such a short distance to the objects being photographed. The equivalent focal length for the sensor with a crop factor of 2.7 is approximately 85 mm, corresponding to an angle of view of about 25 degrees. The Offner optical scheme utilizes a reflective diffraction grating to ensure adequate illumination on the matrix sensor, providing sufficient brightness [29], [30].

Image capturing is carried out using a CMV4000 sensor with a resolution of 2040×2048 , which, with 2×2 binning, yields a frame with a final resolution of 1020×1024 . Binning enhances light sensitivity and compensates for sensor noise. Field imaging was conducted with an exposure time of 65 ms. The exact exposure time was selected based on the 'spectral sweep' signal during shooting, with the frame being captured the hyperspectrometer pointed at the white area of the calibrated XriteColorChecker Video colour scale [34].

A dataset derived from initial HSI, which had undergone various spectral calibrations, was used. For this dataset, the annotation masks were obtained according to [34], with

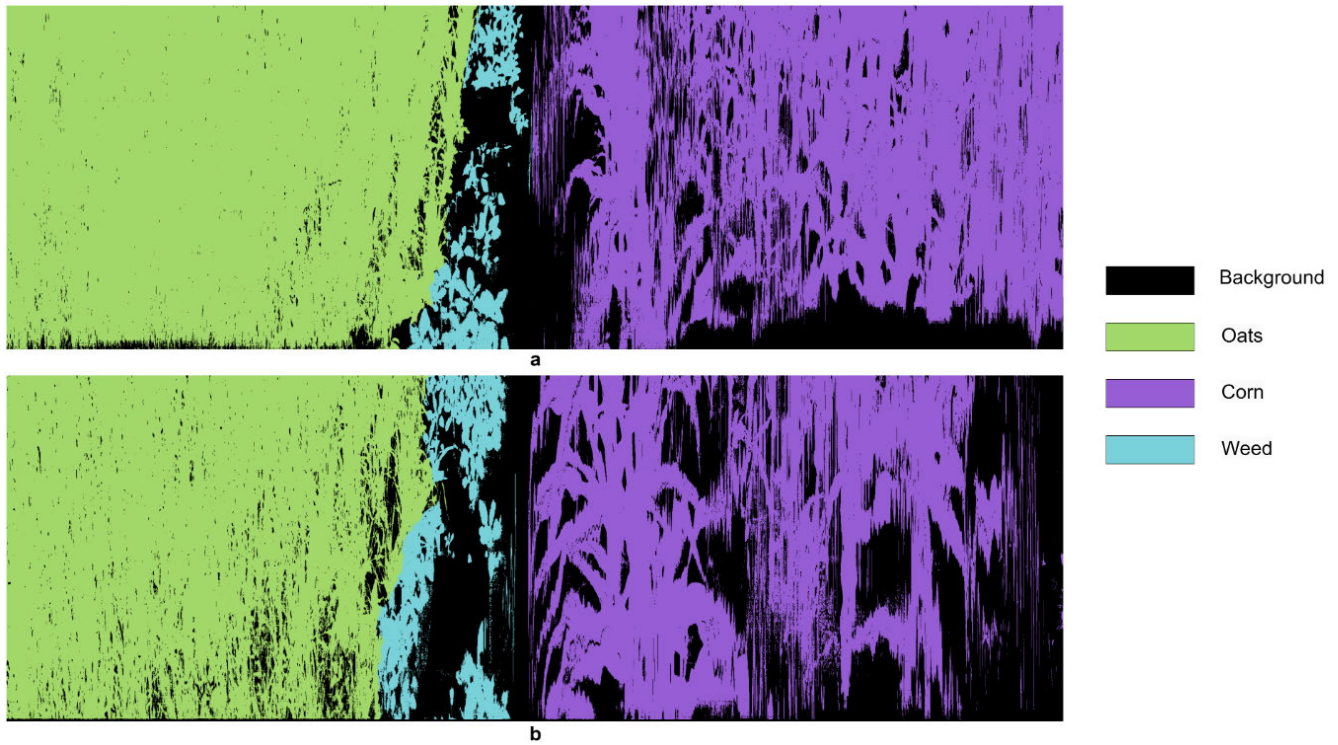


FIGURE 4. Manually annotated ground truth mask, (a) HSI 1, (b) HSI 2.

the exception of isolating the “soil” (background) class into a separate category. To distinguish vegetation, the Normalized Difference Vegetation Index (NDVI) was applied followed by threshold-based binarization [34]. Subsequently, each vegetation class was manually annotated. The visual representation of the masks is shown in Figure 4.

3) CALIBRATION OF HYPERSPECTROMETER SENSITIVITY INHOMOGENEITY

The HSI are a sequence of video frames or a sequence of image sets that can be represented as a third-order tensor $S(i, j, k) : 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K$, where j, k are the pixel coordinates for each frame i . The i -th HSI frame is the spectrally decomposed image of the slit that was recorded by the photodetector matrix. For the formed HSI, it will be referred to as the i -th “slice”. It is assumed that k is the row number (spectral channel) of the HSI, and j is the column number (in the direction of the slit diaphragm).

It is acknowledged that under conditions of uniform illumination of the entrance pupil of the hyperspectrometer, the ideal outcome would be the acquisition of an HSI exhibiting uniform brightness in each k -th channel along the j direction.

A series of experiments were conducted to validate the developed HSI system and ascertain its dependency on the central wavelength in relation to the pixel number within the row of the photodetector matrix. For this purpose,

hyperspectral video frames of a homogeneous white test object were acquired under uniform illumination:

- by a scattered laser light beam with a wavelength in the red spectrum region;
- by sunlight.

Additionally, ‘dark current’ images of the hyperspectrometer were obtained. Under the given shooting parameters, the dark currents were absent, so dark current correction was not required. Dark currents did not affect the recorded signal of the white object.

The following was found:

- 1) An investigation of the test object illuminated by a laser beam (wavelength 640 nm) revealed no geometric distortions in the video frames; however, the transmission intensity along the slit was nonuniform (Figures 5 and 6).
- 2) An analysis of a video frame obtained during the recording of a uniformly illuminated white test object under sunlight revealed that the Fraunhofer A line (oxygen O_2 absorption), located at a wavelength of 760 nm, does not exhibit geometric distortions (Figures-and 8).

The HSI exhibit a shift in the local maxima and minima (extrema) in the rows of the white test object’s cross section, depending on the pixel number j (j denoting the row number of the cross-section, and the pixel number along the slit of the hyperspectrometer). Additionally, variations in light transmission intensity are observed along the entire length

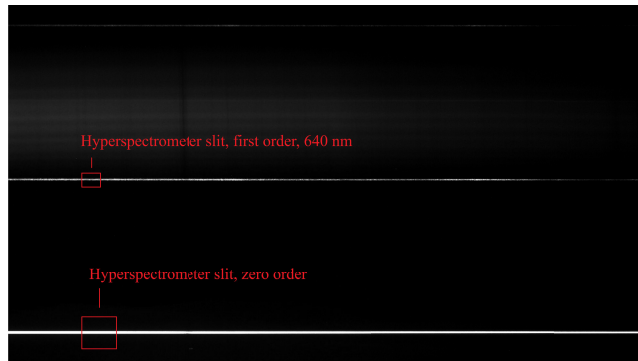


FIGURE 5. A frame of hyperspectral video data representing the image of the slit diaphragm of the hyperspectrometer in the zero and first diffraction orders.

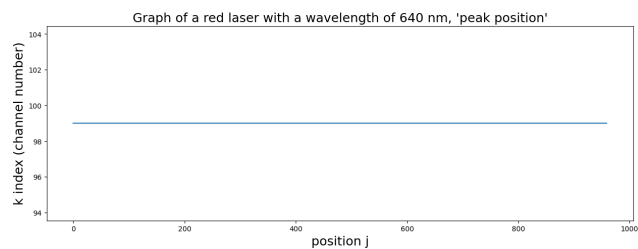


FIGURE 6. Image of the positions of maxima for red light (640 nm).



FIGURE 7. Spectral unfolding of the image of a white object illuminated by sunlight.

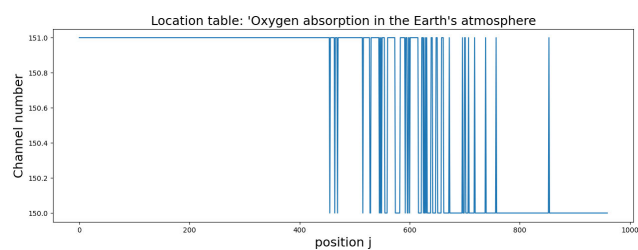


FIGURE 8. Image of the positions of minima at the oxygen absorption line.

of the slit (Figures 9–12). These shifts in local extrema are caused by the dependence of the hyperspectrometer's sensitivity not only on the row number but also on the column number of the frame.

As detailed in [35], wavelength values for each pixel along the slit j and for each hyperspectral imaging channel k are provided, an outcome resulting from geometric distortions referred to as “smile distortions”. Therefore, a relationship

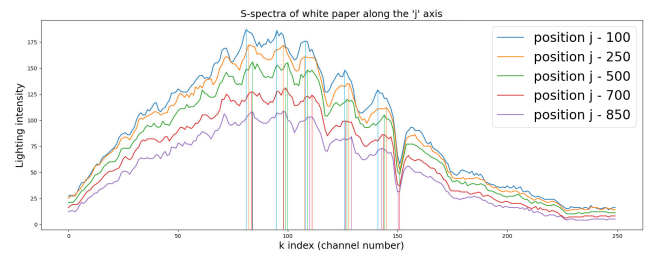


FIGURE 9. Spectrum of the white test object depends on column number in the frame, with marked local extrema.

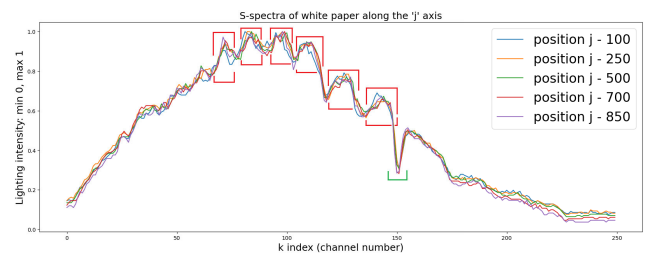


FIGURE 10. Normalized spectrum (ranging from 0 to 1) of the white test object across different frame columns. (1) Red highlights indicate regions where the shift of local extrema is clearly observed. (2) The green highlight marks the region of the Fraunhofer A line, where the local minimum remains unchanged across all frame columns.

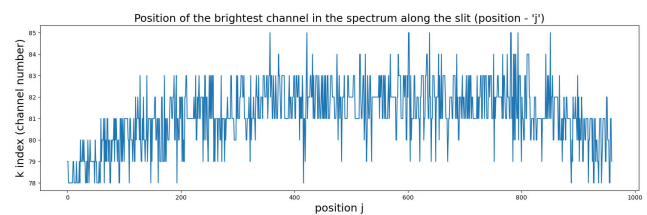


FIGURE 11. Positions of the local spectral extrema depends on the pixel number in the frame column (demonstrating distortions).

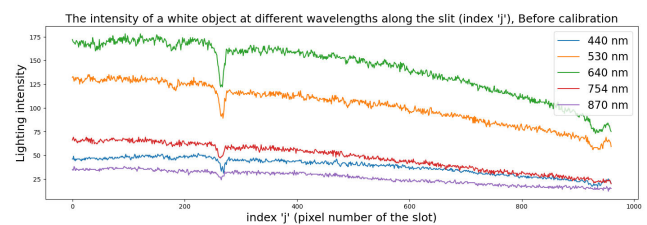


FIGURE 12. Signal from the white test object in different rows of HSI splits (spectral channels).

exists between the channel number k and the row number j , which is attributable to geometric distortions.

While it may appear that the hyperspectrometer utilized in this study displays comparable distortions, this is not the case. The performance evaluation of the hyperspectrometer reveals that it does not demonstrate a dependence of the laser line position at a wavelength of 640 nm (Figure 5) and the Fraunhofer line at 760 nm (Figure 7) on the row number. In other words, there is no evidence of any distortion in these instances.

Thus, a correction coefficient matrix is needed to eliminate signal non-uniformity of the hyperspectrometer along the slit diaphragm at a fixed wavelength range. This calibration will correct the non-uniformity of the signal throughout the hyperspectral imaging system, following the procedure described in Section 4.2.1 of [36]. The calibration data can be described as follows. Consider the i -th slice of the HSI; for simplicity, we denote it as $S_i(j, k)$. The signal in the spectral channel $S_k(j)$ before calibration is defined as:

$$S_k(j) = S_{0,k}(j) \cdot t_k(j) \cdot v(j) \quad (3)$$

where $S_{0,k}(j)$ is the calibrated or “reference” signal at a given wavelength k , $t_k(j)$ are the hyperspectrometer distortion coefficients (these coefficients depend on the column number k and row number j , caused both by the optical system’s non-uniformity and the hyperspectrometer sensor’s sensitivity variation). For the studied hyperspectrometer, the shifts of local extrema are determined, as illustrated in Figures 9-11. The term $v(j)$ represents the hyperspectrometer distortion coefficients (these coefficients depend on the column number j and are caused by changes in the slit diaphragm width along the slit aperture). The distortion coefficients $v(j)$ and $t_k(j)$ are demonstrated in Figure 12.

At the calibration stage of the hyperspectral imaging system, the *lifting function* $l_k(j)$ is selected to compensate for intensity distortions introduced by $t_k(j)$ and $v(j)$, and to obtain a signal close to the source HSI.

To correct the non-uniform transmission of the hyperspectrometer, the HSI of a uniform white object was obtained, denoted as $W(i, j, k)$. During the shooting process, the homogeneous white object was positioned perpendicular to the optical axis of the hyperspectrometer and illuminated exactly uniformly. The test object was placed at a distance that allowed full coverage of the hyperspectrometer’s field of view. The reflected light from the ‘white reference’ created an uniform illumination at the input aperture of the hyperspectrometer.

In the first stage, $W(i, j, k)$ is averaged over I of the white object’s HSI to minimize and compensate for additive noise from both the sensor and the electronic circuitry:

$$W_{\text{mean}}(j, k) = \frac{1}{I} \sum_{i=1}^I W(i, j, k) \quad (4)$$

In the second stage, the signal of the white object that will serve as the reference is determined. The brightest signal of the white object is found by selecting the column index j (i.e., the column along the slit) within an HSI slice, since distortions lead to intensity reduction along the slit diaphragm at index j . The column index j_{max} is determined as the one where the sum of all spectral channels K in the array $W_{\text{mean}}(j, k)$ is maximal:

$$j_{\text{max}} = \arg \max_j \sum_{k=1}^K W_{\text{mean}}(j, k) \quad (5)$$

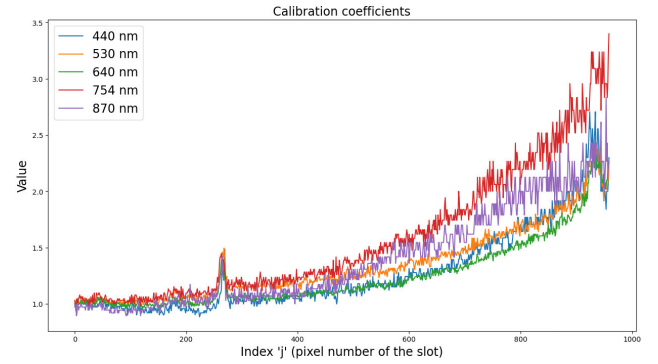


FIGURE 13. Graph of correction coefficients $l_k(j)$.

In the third stage, the brightest column $W_{\text{max}}(k)$ is designated as the reference signal (or reference spectrum). Then, for each wavelength k , a vector of values based on $W_{\text{max}}(k)$ is created:

$$W_k^*(j) = W_{\text{max},k} \quad (6)$$

At the fourth stage, a correction coefficient vector is generated for each wavelength k to compensate for the influence of $t_k(j)$ and $v(j)$ on the corrected signal $S_{0,k}(j)$:

$$l_k(j) = \frac{W_k^*(j)}{W_{\text{mean},k}(j)} \quad (7)$$

To perform the HSI slice calibration procedure for each wavelength k , the following transformation is applied:

$$S_{0,k}(j) = S_k(j) \cdot l_k(j) \quad (8)$$

To perform the calibration procedure for the entire HSI, the operation described above must be repeated for each wavelength of all slices included in the HSI. The experimentally obtained calibration coefficients at specific wavelengths are presented in Figure 13.

Setting a parametric approximation for the vectors $l_k(j)$ allows trends in Figure 13 to be evaluated. In this case, the approximating function for each wavelength k can be found under the condition:

$$\sum_{j=1}^J [m_k(j, a_k) - l_k(j)]^2 \rightarrow \min \quad (9)$$

In this study, the approximation $m_k(j, a_k)$ is performed using a quadratic parabola, which allows for the assessment of factors influencing signal non-uniformity in the HSI slice and the creation of a low-parametric distortion correction model that is less sensitive to noise:

$$m_k(j, a_k) = a_{k,1} + a_{k,2}j + a_{k,3}j^2 \quad (10)$$

Using this approximation (see Figure 14), correction coefficients for compensating $t_k(j)$ were calculated. Figure 15 shows the signal from the white test object in various rows of the HSI slice after applying analytical approximation calibration $m_k(j)$. With this approximation, it is possible

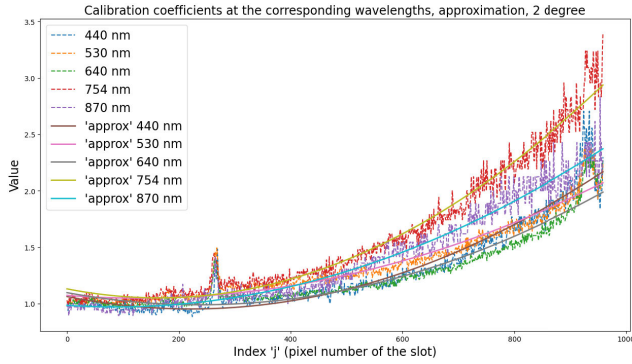


FIGURE 14. Type of approximating polynomials $m_k(j)$ (solid lines).

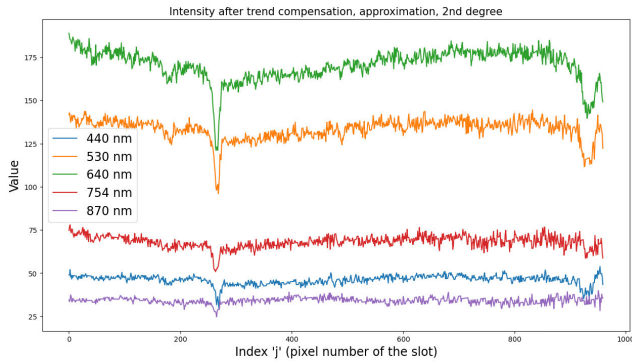


FIGURE 15. Signal from the white test object in different rows of the HSI slice (spectral channels) after applying analytical approximation calibration $m_k(j)$.

to determine coefficients for compensating $v(j)$, assuming the calibration coefficients for $h(j)$ (see Figure 16). After compensating $t_k(j)$ and $v(j)$, the best-case signal $S_{0,k}(j)$ nearly coincides with the reference signal at all coordinates j (see Figure 22). Thus, the corrected function using the approximation $l_k^*(j)$ is defined as:

$$l_k^*(j) = m_k(j, a_k) \cdot h(j) \quad (11)$$

Further studies are expected to explore the possibility of using the obtained approximation coefficients to compensate for non-uniformity in video frames obtained at the actual spatial coordinates of the video frame. In the context of real-world recording, the hyperspectrometer is susceptible to various factors (e.g., temperature, vibrations) that can lead to a shift in the image recorded by the system relative to the j, k coordinates of the photodetector matrix.

Figure 17 presents the spectral scan of the image of a homogeneous white test object after the calibration procedure without approximation. The positions of the local extrema of the spectra depending on the pixel number in the column of the HSI cross-section are shown in Figure 18. The signal from the white test object in different columns of the HSI cross-section after calibration is presented in Figure 19. The signal from the white test object in different rows of the HSI cross-section (spectral channels) after the calibration

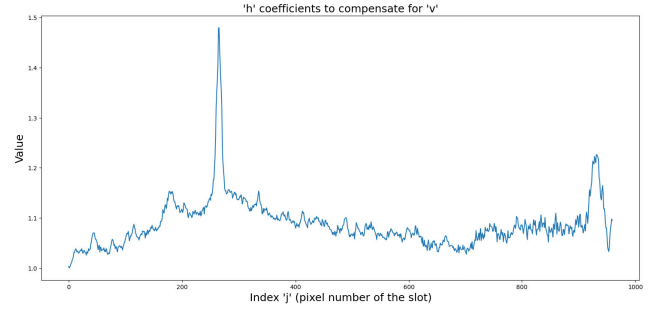


FIGURE 16. Type of correction coefficients $l_k(j)$ used for compensation.



FIGURE 17. Spectral scan of the image of a homogeneous white test object after the calibration procedure with coefficients $l_k(j)$ (without approximation) (before calibration see Figure 7).

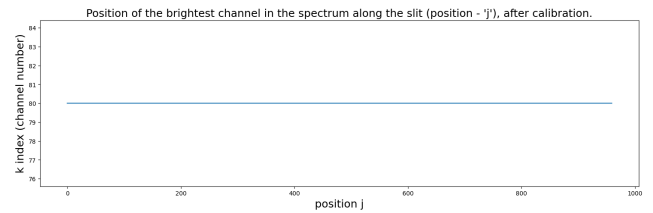


FIGURE 18. Positions of local extrema of spectra depending on the pixel number in the column of the HSI cross-section (before calibration see Figure 11).

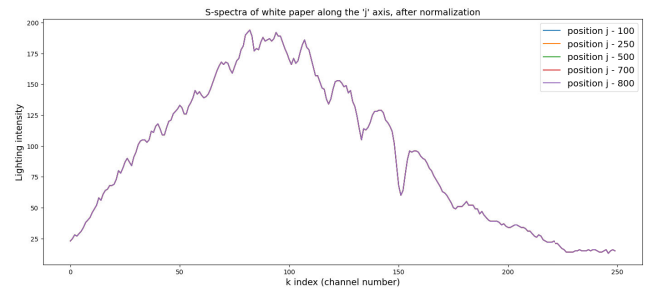


FIGURE 19. Signal from the white test object in different columns j of the HSI cross-section after calibration.

procedure is shown in Figure 20. Visualization of the calibration coefficients for HSI calibration is presented in Figure 21.

4) SPECTRAL CALIBRATION

Converting the discrete representation of Equation 2, the signal registered by the hyperspectrometer at point x for wavelength λ_i can be expressed as:

$$S(x, \lambda_i) = R(x, \lambda_i) I(\lambda_i) T(\lambda_i) \quad (12)$$

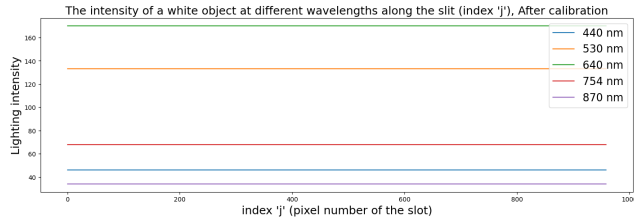


FIGURE 20. Signal from the white test object in different rows of the HSI cross-section (spectral channels) after the calibration procedure.



FIGURE 21. Visualization of the correction coefficients $I_k(j)$ for HSI calibration.

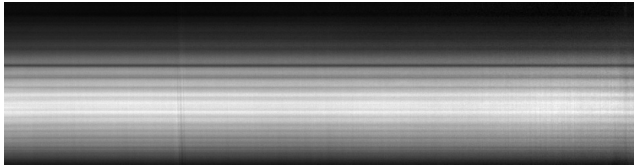


FIGURE 22. Spectral scan of the image of a homogeneous white test object after the calibration procedure with coefficients $I_k(j)$ (with approximation) (before calibration see Figure 7).

where:

- S is the signal registered by the hyperspectrometer;
- R is the reflection spectrum of the object;
- I is the illumination source spectrum;
- T is the spectral sensitivity of the hyperspectrometer;
- λ_i is the i -wavelength.

To eliminate the dependency of the hyperspectral data on the illumination conditions, imaging conditions can be used such that each recorded scene includes an object with a reference reflection spectrum:

$$S_{ref}(x, \lambda_i) = R_{ref}(x, \lambda_i) I(\lambda_i) T(\lambda_i) \quad (13)$$

Experimentally, this can be obtained by measuring the light source with the same device, where the registered signal will not contain a reflection spectrum:

$$S_{ls}(\lambda_i) = I(\lambda_i) T(\lambda_i) \quad (14)$$

Then, by dividing (Equation 13) by (Equation 14), the desired reflection spectrum of the reference object can be obtained:

$$R_{ref}(x, \lambda_i) = \frac{S_{ref}(x, \lambda_i)}{S_{ls}(x, \lambda_i)} \quad (15)$$

Next, it is necessary to divide the registered signal in each pixel of the HSI (Equation 12) by the signal for the reference object (Equation 13) to obtain the relative reflection spectrum

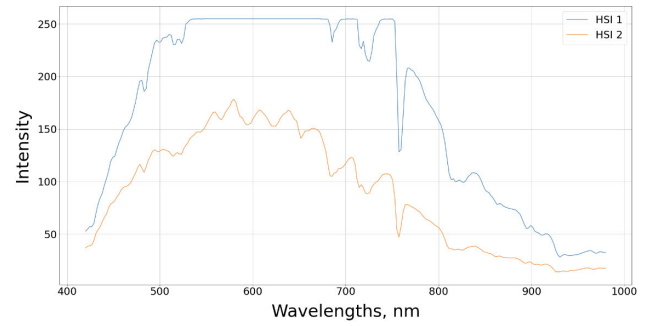


FIGURE 23. Spectral intensity for the white section ColorChecker (100%) of the uncalibrated HSI.

of the object:

$$RR(x, \lambda_i) = \frac{S(x, \lambda_i)}{S_{ref}(x, \lambda_i)} \quad (16)$$

Therefore, the spectrum of the illumination source and the sensor's sensitivity are eliminated from the equation. To obtain the actual reflection spectrum of the objects (Equation 15), it is required to multiply by (Equation 16), yielding:

$$R(x, \lambda_i) = \frac{S(x, \lambda_i)}{S_{ref}(x, \lambda_i)} R_{ref}(x, \lambda_i) \quad (17)$$

Special calibration objects with Lambertian scattering surfaces, such as Spectralon, are commonly used as reference objects [37], [38]. Spectralon is hydrophobic and thermally stable up to 350°C. This durable material ensures high accuracy and reproducibility of data. Unlike other standards made from polytetrafluoroethylene (PTFE), Spectralon can be repeatedly cleaned and polished in case of contamination or scratches.

In this study, spectral calibration was performed using a XRiteColorChecker Video (ColorChecker) like inexpensive replacement for Spectralon. For further application of the calibrated hyperspectral images in classification tasks, both the RR values from (Equation 16) and the R values from (Equation 17) can be used.

Spectral calibration was conducted using the light gray area of the ColorChecker, as overexposure was observed on the white area (Figures 4 and 23). The reflection spectra of vegetation before and after calibration are presented in Figures 24 and 25.

5) DATA SCALING

The values of HSI data can significantly vary in scale (within a single class) both channel-wise and pixel-wise [39], which may lead to a deterioration in the quality of data analysis. This problem is particularly relevant for data acquired under different external conditions. To bring the data to a uniform scale and compensate for such variations, we employed standardization method.

Standardization transforms the data so that its mean value is 0 and its standard deviation is 1, based on the following

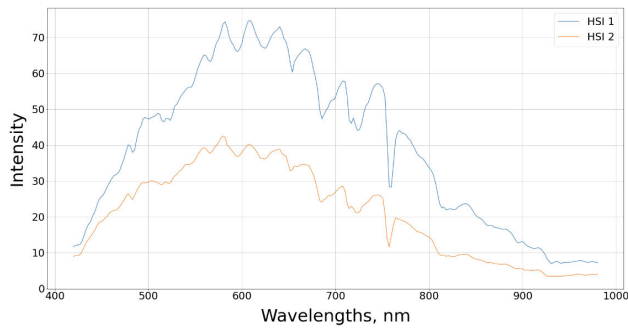


FIGURE 24. Spectral intensity for the grey section ColorChecker (50%) of the uncalibrated HSI.

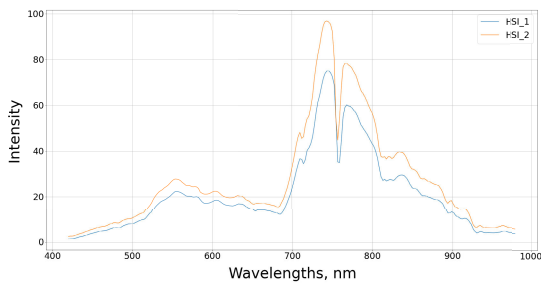


FIGURE 25. Spectral intensity for vegetation of the uncalibrated HSI.

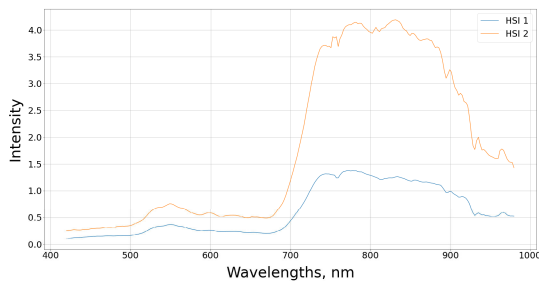


FIGURE 26. Spectral intensity for vegetation of the calibrated HSI.

equation:

$$X_{\text{stand}} = \frac{X - \bar{X}}{\sigma_X} \quad (18)$$

where:

X is the value to be standardized;

$\bar{X}$ is the mean value;

σ_X is the standard deviation.

This data scaling method can be applied both pixel-wise and channel-wise.

6) TABULAR DATA

In order to establish a baseline classification algorithm, a series of classical ML models were chosen. For ease of use, HSIs were transformed into a tabular format ‘CSV’, in accordance with approach described in [40]. For this

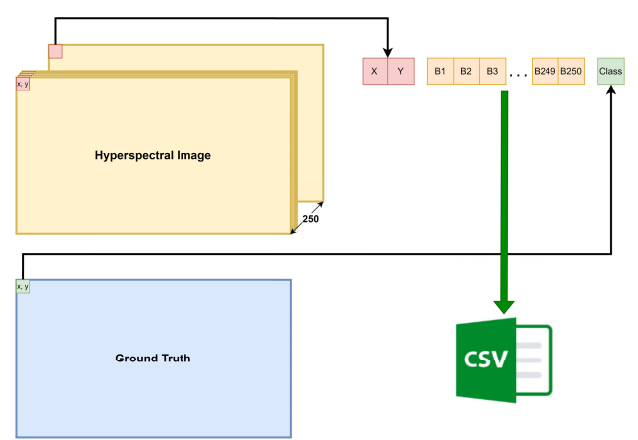


FIGURE 27. Full pipeline of conversion from image to table.

purpose, each pixel of the image was presented as a row with the pixel coordinates, a list of channel values, and the target variable. As a result, for each image, a table was obtained with dimensions of 2,928,000 by 253. This table consisted of the following columns x , y , $band1$, $\dots$, $band250$, and target. The data was saved in CSV format (Figure 27).

B. MACHINE LEARNING ALGORITHMS

1) MODELS

In this study, the baseline models used were the Random Forest (RF) algorithm [41] implemented in the sklearn library and the CatBoost algorithm [42] represented in the CatBoost package. Both methods are based on decision trees, which are employed for classification or regression tasks by partitioning the data according to specific rules.

In the RF method, the original data is sampled, and each tree is trained on its own subset of this data. This approach helps to reduce overfitting and improve the model’s generalization ability by introducing diversity into the training samples.

In contrast, in CatBoost, each new tree is created to minimize the error made by the previous trees. This iterative training process allows for the sequential improvement of the model’s accuracy by correcting the errors of preceding steps, thereby achieving higher predictive capability.

2) EXPERIMENTAL SETUP

A grid search with 5-fold cross-validation (Grid Search CV from the Sklearn package) was employed for hyperparameter tuning. Each model’s unique set of hyperparameters was selected to optimally adjust each model according to its specific characteristics.

For the Random Forest model, the following hyperparameters were tuned:

- `max_features`: auto, sqrt, log2. This parameter determines the number of features to consider when looking for the best split at each node. The value “auto” selects all features, “sqrt” selects the square root of the total number of features,

and “log2” selects the base-2 logarithm of the number of features.

- max_depth: 4, 6, 8, 10, 12, 14, 16, 18. This parameter sets the maximum depth of the tree. It controls how deep each tree can grow. Increasing this value can lead to more complex models, potentially improving their accuracy, but it can also result in overfitting.

- criterion: gini, entropy. Both criteria are used to measure the quality of a split in decision trees.

For the **CatBoost** model, the following parameters were used:

- depth: 3, 5, 7, 10, 12. This parameter specifies the depth of the tree. Deeper trees can model more complex relationships but are also more prone to overfitting.

- iterations: 100 to 1000 in steps of 100. This parameter indicates the number of training iterations (number of trees). A larger number of iterations can improve model accuracy but also increases training time and the risk of overfitting.

- learning_rate: 0.01, 0.02, 0.03, 0.04. This parameter determines the step size at each iteration while updating the model. Lower values ensure slower and more stable learning, while higher values lead to faster but potentially less stable learning.

C. DEEP LEARNING

1) MODELS

Classical architectures used for RGB segmentation may be less applicable for HSI due to the larger size in the spectral domain compared to RGB, lower spatial resolution, and the need for creating extensive annotated datasets. Therefore, for the task of HSI classification using neural networks, architectures with a small input window size in spatial coordinates are typically employed, serving as an additional source of information to spectral features. The effective extraction of features considering the spatial-spectral components of the object directly influences the performance of the neural network classifier. Currently, transformer-based approaches lead in the field of hyperspectral image classification [23], [24], [25], [43].

We propose a new architecture Deep Spectral-Spatial Transformer (DSST), based on the transformer network described in the following work [23]. This network consists of a spatial-spectral feature extractor, a tokenizer, a transformer encoder, and a classification linear layer (Figure 28). The original architecture of the Spectral-Spatial Feature Tokenization Transformer (SSFTT) employs two sequential layers of 3D and 2D convolutions as the spatial-spectral feature extraction block for further processing by the tokenizer and transformer block. This architecture demonstrated high classification metric values and reduced training and inference times. A modification of the SSFTT is proposed, involving the replacement of the feature extractor block with a deeper one, which will enable more effective extraction of spatial-spectral features.

As the spatial-spectral feature extractor, the Normalized-Modified 3D Convolutional Neural Network (NM3DCNN)

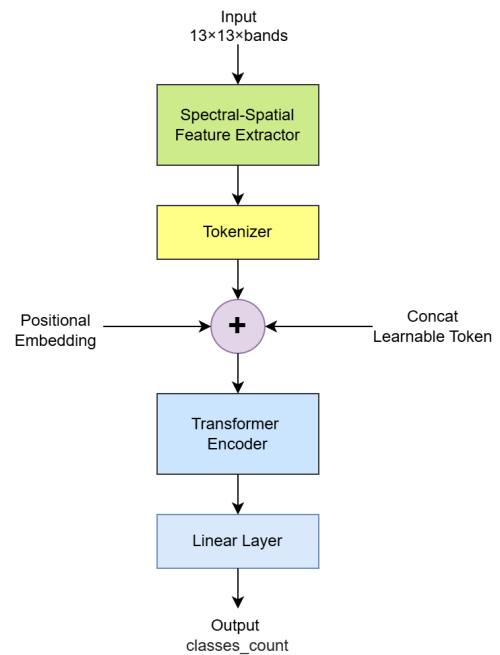


FIGURE 28. The block diagram of the modified neural network classifier DSST.

network from the studies [34], [44] is proposed (Figure 29), without the last linear layer and with the addition of a 2D convolution at the network’s output. This approach allows for the extraction of both spatial-spectral features and various spectral features through the use of blocks with parallel 3D convolutions and batch-norm layers (BNConvBlock), each having its own kernel size (Figure 30). This architecture has demonstrated high classification metric values in the referenced works compared to other spectral-spatial architectures [34].

For the purpose of comparing the effectiveness of the developed architecture, the one-dimensional convolutional network M1DCNN, which operates solely with spectral information based on [18], was used. This architecture was improved by adding an additional 1D convolutional layer and a batch normalization layer, which allowed it to achieve higher metric values. Furthermore, the original architectures of the spatial-spectral network NM3DCNN [34], [44] and the transformer network SSFTT [23] were also compared.

We also compare the GSC-ViT (Groupwise Separable Convolutional Vision Transformer) Network that is a lightweight and efficient model designed for HSI classification, integrating convolutional operations with Vision Transformers [45]. The key innovation lies in the groupwise separable convolution module, which combines grouped pointwise convolution and group convolution to reduce the number of parameters while enhancing the model’s ability to capture local spectral-spatial features. Additionally, the model utilizes a groupwise separable multihead self-attention module, which replaces the traditional multihead

self-attention in Transformers, enabling efficient extraction of both local and global spatial information.

In the context of our work, we also explored the use of the CLOLN (Channel-Layer-Oriented Lightweight Spectral-Spatial Network) architecture [46]. Deep neural networks with a large number of parameters and layers are typically employed for hyperspectral images. CLOLN is designed to reduce the complexity of the network while maintaining high classification quality. Within this architecture, the COS2m module is utilized. The COS2m module is a lightweight structure for extracting spectral-spatial features, employing pointwise convolution to reduce spectral channels, followed by 3D convolution with expansion and compression for efficient feature extraction with minimal computational overhead. It combines spectral and spatial features using depthwise convolution and 3D convolution, ensuring high efficiency and reduced computational cost. This network also incorporates a novel activation function that considers neighboring pixel information (NPAF), which enhances information transfer and improves the network's approximation capability through a single activation layer.

The last two architectures were proposed in 2024 and have demonstrated excellent performance on standard datasets for hyperspectral image classification tasks.

2) EXPERIMENTAL SETUP

Due to the different types of architectures used for the experiment, the processing and dimensionality of the input data were different. For example, for the M1DCNN architecture, the dimensionality of the input data is $1 \times 1 \times B$, where B is the number of HSI channels. For transformer-based architectures, the dimensionality of the data was $N \times N \times B$, where N is the spatial size of the image, more precisely the patch size. The training dataset consisted of 5% patches from one of the HSIs, and 1% was used for validation. The entire second HSI was used as the test sample on which the classification metrics were computed. The experiment was also performed in the reverse case. Adam and SGD were used as the optimizer. The training step was chosen using a scheduler that decreased the training step every 5-10 epochs, depending on the complexity of the model. Training continued until the metrics and loss function reached a plateau. The presented results show the best metrics of each model after hyperparameter search (Table 1).

D. PERFORMANCE METRICS

In our study, we utilized the metrics Accuracy (Equation 19) and F1-score (Equation 20). Accuracy is a fairly straightforward and interpretable metric, however, its weakness is inability to handle class imbalance. Therefore, we additionally evaluated the F1-score, which is the harmonic mean of Precision (Equation 21) and Recall (Equation 22) and works better with class imbalance. It is also important to note that we used the F1-Score weighted modification, which calculates

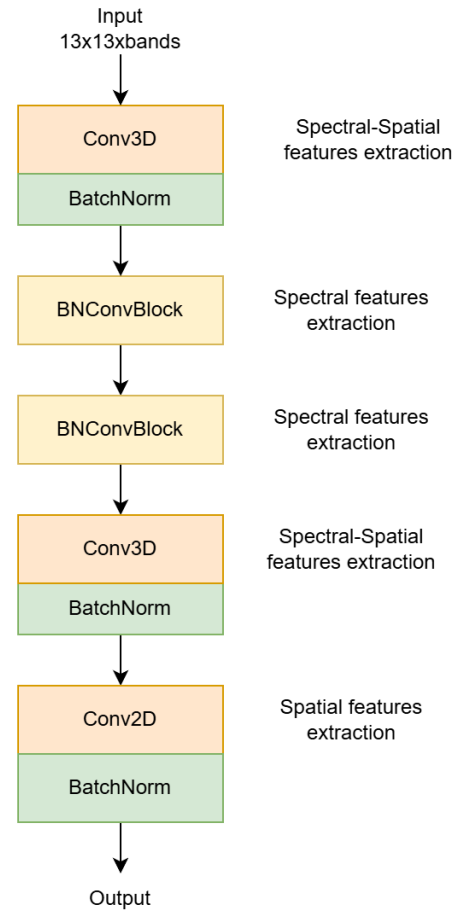


FIGURE 29. The block diagram of the feature extractor based on NM3DCNN.

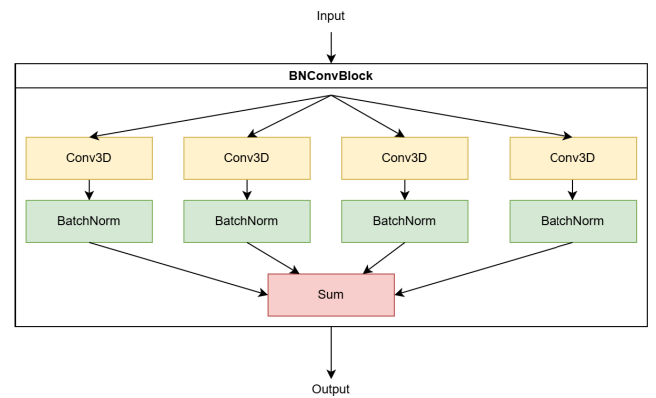


FIGURE 30. The diagram of the BNConvBlock.

the class-weighted F1-Score (Equation 23).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)$$

TABLE 1. Parameters to obtain the best performance for DL models.

	CLOLN	GSCViT	MIDCNN	NM3DCNN	SSFTT	DSST
Patch size	13	8	1	7	13	13
Optimizer	Adam	Adam	SGD	SGD	Adam	Adam
Learning rate	0.00005	0.00005	0.01	0.005	0.00005	0.0001
Step size	5	5	10	5	5	5
Gamma	0.5	0.5	0.5	0.5	0.5	0.5
Epoch	50	50	30	30	40	35

TABLE 2. Random forest predictions on scaled data, (1) train on light HSI 1, test on dark HSI 2, (2) train on dark HSI 2, test on light HSI 1.

	F1			Acc		
	1	2	AVG	1	2	AVG
Unscaled	0.676	0.698	0.687	0.655	0.704	0.679
Pixel-wise	0.584	0.606	0.595	0.632	0.631	0.631
Channel-wise	0.844	0.661	0.752	0.868	0.679	0.773

TABLE 3. Catboost predictions on scaled data, (1) train on light HSI 1, test on dark HSI 2, (2) train on dark HSI 2, test on light HSI 1.

	F1			Acc		
	1	2	AVG	1	2	AVG
Unscaled	0.660	0.667	0.663	0.650	0.658	0.654
Pixel-wise	0.437	0.465	0.451	0.526	0.548	0.537
Channel-wise	0.866	0.671	0.768	0.885	0.686	0.785

$$\text{Precision} = \frac{TP}{TP + FP} \quad (21)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (22)$$

where:

- TP is the number of true positive results;
- TN is the number of true negative results;
- FP is the number of false positive results; - FN is the number of false negative results.

$$\text{F1-Score}_{\text{weighted}} = \sum_{i=1}^n \frac{N_i}{N} \cdot \text{F1-Score}_i \quad (23)$$

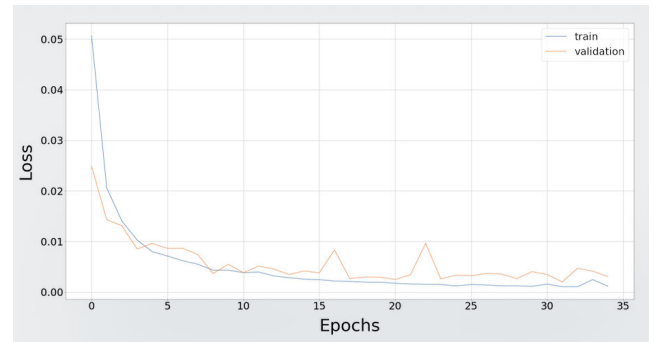
where:

- n is the number of classes;
- N_i is the number of true instances for class i ;
- N is the total number of instances across all classes;
- F1-Score_i is the F1-Score for class i .

For simplicity, we will denote the weighted F1-score as F1 and Accuracy as Acc.

III. RESULTS AND DISCUSSION

In the initial stage of the research, experiments were conducted to standardize the features of hyperspectral images, including per-pixel and per-channel standardizations (Tables 2 and 3). These experiments were performed using classical ML models (RF, CatBoost). Based on the obtained data, per-channel standardization demonstrated the best results for HSI classification. Consequently, this preprocessing method was employed in all subsequent experiments. It is also important to note that without standardization, the algorithms perform significantly worse compared to their performance with standardization.

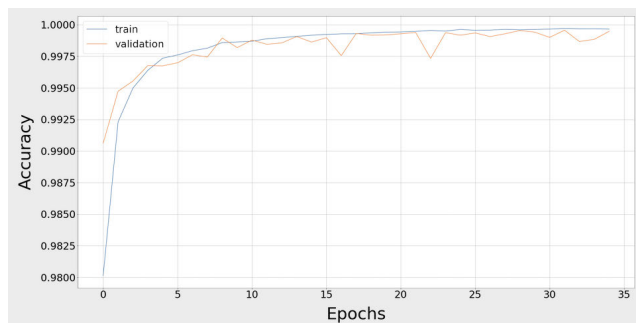
**FIGURE 31.** Loss function plots for the training set and the validation data sets for DSST.

The training of the best DSST model until the loss and accuracy functions plateaued (Figures 31 and 32) required 35 epochs and took approximately 180 minutes on an Nvidia RTX 3090 GPU with 24 GB of video memory. For classical ML methods, optimal parameters were selected using grid search. For Random Forest, the best parameters were: $\text{max_depth} = 14$, $\text{max_features} = \log_2$, $\text{criterion} = \text{gini}$. For CatBoost, the optimal parameters were: $\text{depth} = 12$, $\text{learning_rate} = 0.04$, $\text{iterations} = 1000$.

Table 4 presents the training results of all the classifiers discussed above. As evident from the obtained data, using HSI with greater scene illumination yields higher classification metric values compared to using darker HSI. Conversely, training on darker HSI results in a significant loss of classification accuracy. This phenomenon can be attributed to the fact that under low illumination, pixels of the target classes lose information at the ends of the spectra and exhibit more noise in the central part of the spectrum, leading to a potential loss of key information. Adequate illumination,

TABLE 4. Comparison of models' performance metrics for each calibration and ablation on DSST components NM3DCNN and SSFTT.

	Catboost		Random Forest		CLOLN [46]		GSCViT [45]		M1DCNN [18]		NM3DCNN [34]		SSFTT [23]		DSST	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Without calibration																
1	0,872	0,864	0,841	0,824	0,921	0,927	0,930	0,934	0,900	0,882	0,918	0,920	0,899	0,903	0,934	0,940
2	0,673	0,690	0,656	0,645	0,727	0,723	0,724	0,727	0,810	0,805	0,810	0,813	0,786	0,790	0,832	0,839
AVG	0,773	0,777	0,749	0,735	0,824	0,825	0,827	0,831	0,855	0,844	0,864	0,867	0,843	0,847	0,883	0,890
Spectral gray calibration																
1	0,872	0,863	0,842	0,826	0,954	0,960	0,943	0,946	0,908	0,898	0,929	0,928	0,912	0,917	0,939	0,944
2	0,674	0,689	0,657	0,646	0,784	0,787	0,738	0,740	0,833	0,832	0,814	0,817	0,796	0,801	0,857	0,863
AVG	0,773	0,776	0,750	0,736	0,869	0,874	0,841	0,843	0,871	0,865	0,872	0,873	0,854	0,859	0,898	0,904
Extremum alignment and gray calibration																
1	0,893	0,884	0,860	0,842	0,941	0,945	0,941	0,945	0,913	0,903	0,937	0,938	0,918	0,923	0,948	0,954
2	0,728	0,737	0,728	0,719	0,815	0,823	0,760	0,760	0,836	0,837	0,830	0,833	0,817	0,823	0,857	0,865
AVG	0,811	0,811	0,794	0,781	0,878	0,884	0,851	0,853	0,875	0,870	0,884	0,886	0,868	0,873	0,903	0,910
Extremum alignment (approximation) and gray calibration																
1	0,893	0,884	0,862	0,844	0,941	0,949	0,929	0,936	0,919	0,910	0,931	0,932	0,936	0,941	0,943	0,947
2	0,732	0,737	0,737	0,728	0,842	0,850	0,812	0,816	0,831	0,830	0,837	0,840	0,815	0,823	0,870	0,875
AVG	0,813	0,811	0,800	0,786	0,892	0,900	0,871	0,876	0,875	0,870	0,884	0,886	0,876	0,882	0,907	0,911

**FIGURE 32.** Accuracy metric plots for the training and validation data sets for DSST.

on the other hand, preserves this crucial information, thereby enhancing the classification performance.

From Table 4, it is evident that all calibrations positively impact Accuracy and F1-score metrics. All models show the best results when using calibration with calibration coefficients (with approximation) and gray illumination alignment for the two HSI (“Extremum alignment (approximation) and gray calibration”). Prediction using the original HSI showed worse results compared to prediction using the calibrated HSI. Gray calibration eliminated the variability of the recorded signal depending on the incident illumination intensity. Extremum alignment eliminated the heterogeneity of the hyperspectrometer signal along the slit diaphragm in a fixed wavelength range. The combined use of these two calibrations allowed us to achieve the best results. Therefore, the proposed pipeline for HSI analysis using ML-based algorithms for lighting invariance improvement showed effective way for performance improvement.

Among classification methods using only spectral information (RF, CatBoost, and M1DCNN), M1DCNN demonstrated the best classification quality metrics. However, all algorithms experienced a significant performance decrease when trained on the dark HSI dataset and tested on the light HSI. Along with M1DCNN, methods incorporating spatial-spectral information (CLOLN, GSCViT, NM3DCNN, SSFTT, and DSST) showed superior performance. Specifically, NM3DCNN and SSFTT showed comparable results but slightly underperformed in terms of average Accuracy and F1-score compared to CLOLN and GSCViT algorithms. The ablation part of the study consists of comparing the results of NM3DCNN and SSFTT, on the basis of which the DSST was created with the DSST itself. However, their combination in the DSST model yielded the highest performance among all models with an Accuracy of 0.907 and F1-score of 0.911. The improved performance can be attributed to deeper extraction of spatial-spectral features and the integration of a transformer block, which effectively captures both local and global contextual information. This improvement was observed even when models were trained on poorly illuminated HSI, highlighting the robustness and efficiency of the DSST model in handling various lighting conditions. Although the proposed architecture of the DSST in combination with calibration methods shows a significant increase in classification accuracy, on poor quality data due to the lack of information in poorly illuminated areas, lower classification metrics are observed. This problem requires further research and other approaches.

Based on visual results, it can be concluded that even the best baseline classical ML model, CatBoost, almost fails to handle the weed detection task. In contrast, the DSST model

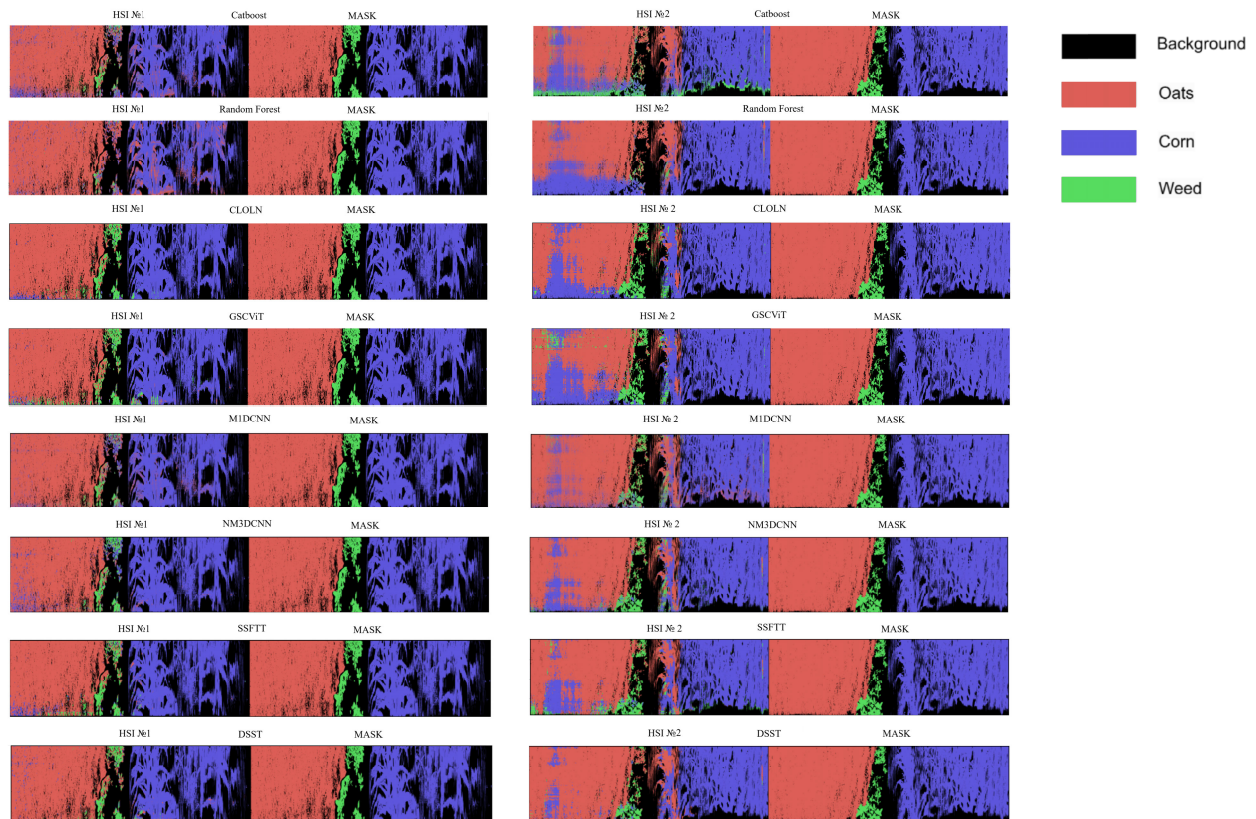


FIGURE 33. The best model inference. (a) Hyperspectral Image 1, (b) Hyperspectral image 2.

demonstrates good performance, although it slightly confuses the Oats and Corn classes (Figure 33).

Moreover, the use of advanced techniques such as transformer blocks facilitates the capture of complex dependencies within the data, further improving the classification accuracy. This study underscores the importance of integrating spatial and spectral information in hyperspectral image analysis and demonstrates that sophisticated models like DSST can significantly outperform traditional methods, especially in challenging scenarios with low illumination.

IV. CONCLUSION

The development of robust algorithms for hyperspectral images is crucial for various environmental applications. Variations in outdoor lighting conditions may lead to inaccuracies in target object recognition and classification. Therefore, in this study, we focus on tackling this problem using advanced deep learning custom architectures that can process HSI with hundreds spectral bands. We conducted a comprehensive study with several ML and DL architectures to compare various approaches in the task of weed segmentation in agricultural fields. Within the framework of the research, an algorithm for calibrating the sensitivity of a hyperspectral sensor has been developed and proposed, ensuring improvement in the quality of acquired measurement data. The implementation of the proposed

calibration methodology has allowed achieving a significant increase in the efficiency of models when solving the segmentation task. Several preprocessing techniques were considered several preprocessing techniques to obtain a more uniform spectral and spatial representation of the studied images. The proposed DL transformer architecture, called DSST, leads to the best segmentation results in terms of F1-score and accuracy compared to other state-of-the-art and classical ML baseline algorithms. The high performance is associated with a deep feature extractor developed specifically for high-dimensional feature space. Despite the efforts made, the classification of bad illuminated HSI remains a challenging problem. This underscores the necessity of further research and the development of advanced calibration and preprocessing methods to enhance the accuracy and reliability of classification algorithms. Further development of the research involves adding dimensionality reduction methods of HSI to the workflow pipeline, optimizing the classifier architecture (as only the feature extractor replacement was explored in this work), increasing the dataset through capturing images at different times of day and weather conditions.

ACKNOWLEDGMENT

(Andrey Makarov, Islombek Mirpulatov, and Nikita Firsov contributed equally to this work.)

REFERENCES

- [1] D. Wu and D.-W. Sun, "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals," *Innov. Food Sci. Emerg. Technol.*, vol. 19, pp. 1–14, Jul. 2013.
- [2] T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, and J. Sousa, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sens.*, vol. 9, no. 11, p. 1110, Oct. 2017.
- [3] A. Pimenov, S. Nesteruk, N. Mikhailova, A. Stepanov, N. Sliusar, V. Korotaev, A. Baldycheva, and A. Somov, "Discrete spectral sensing system for separation of polyolefin waste plastics," in *Proc. IEEE 28th Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2023, pp. 1–4.
- [4] S. Peyghambari and Y. Zhang, "Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review," *J. Appl. Remote Sens.*, vol. 15, no. 3, Jul. 2021, Art. no. 031501.
- [5] S. Illarionova, D. Shadrin, P. Tregubova, V. Ignatiev, A. Efimov, I. Oseledets, and E. Burnaev, "A survey of computer vision techniques for forest characterization and carbon monitoring tasks," *Remote Sens.*, vol. 14, no. 22, p. 5861, Nov. 2022.
- [6] B. Lu, P. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, p. 2659, Aug. 2020.
- [7] B. Jia, W. Wang, X. Ni, K. C. Lawrence, H. Zhuang, S.-C. Yoon, and Z. Gao, "Essential processing methods of hyperspectral images of agricultural and food products," *Chemometric Intell. Lab. Syst.*, vol. 198, Mar. 2020, Art. no. 103936.
- [8] A. Wang, Y. Wang, and Y. Chen, "Hyperspectral image classification based on convolutional neural network and random forest," *Remote Sens. Lett.*, vol. 10, no. 11, pp. 1086–1094, Nov. 2019.
- [9] D. A. Roberts, K. L. Roth, E. B. Wetherley, S. K. Meerdink, and R. L. Perroy, "Hyperspectral vegetation indices," in *Hyperspectral Indices and Image Classifications for Agriculture and Vegetation*. Boca Raton, FL, USA: CRC Press, 2018, pp. 3–26.
- [10] O. Okwuashi and C. E. Ndehedehe, "Deep support vector machine for hyperspectral image classification," *Pattern Recognit.*, vol. 103, Jul. 2020, Art. no. 107298.
- [11] J. Ren, R. Wang, G. Liu, Y. Wang, and W. Wu, "An SVM-based nested sliding window approach for spectral-spatial classification of hyperspectral images," *Remote Sens.*, vol. 13, no. 1, p. 114, Dec. 2020.
- [12] F. Tong and Y. Zhang, "Exploiting spectral-spatial information using deep random forest for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [13] Z. Wang, Z. Zhao, and C. Yin, "Fine crop classification based on UAV hyperspectral images and random forest," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 4, p. 252, Apr. 2022.
- [14] D. Akila, A. Bhaumik, S. Doss, and A. Ameen, "Hyperspectral image classification by means of supapixel representation with knn," in *Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2019*. Cham, Switzerland: Springer, 2021, pp. 369–378.
- [15] S. Jin, F. Zhang, Y. Zheng, L. Zhou, X. Zuo, Z. Zhang, W. Zhao, W. Zhang, and X. Pan, "CSKNN: Cost-sensitive K-nearest neighbor using hyperspectral imaging for identification of wheat varieties," *Comput. Electr. Eng.*, vol. 111, Oct. 2023, Art. no. 108896.
- [16] J. Abdulridha, Y. Ampatzidis, P. Roberts, and S. C. Kakarla, "Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence," *Biosyst. Eng.*, vol. 197, pp. 135–148, Sep. 2020.
- [17] M. Q. Alkhatib, M. Al-Saad, N. Aburaed, S. Almansoori, J. Zabalza, S. Marshall, and H. Al-Ahmad, "Tri-CNN: A three branch model for hyperspectral image classification," *Remote Sens.*, vol. 15, no. 2, p. 316, Jan. 2023.
- [18] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Aug. 2015.
- [19] N. Firsov, E. Myasnikov, V. Lobanov, R. Khabibullin, N. Kazanskiy, S. Khonina, A. A. Butt, and A. Nikonov, "HyperKAN: Kolmogorov–Arnold networks make hyperspectral image classifiers smarter," *Sensors*, vol. 24, no. 23, p. 7683, Nov. 2024. [Online]. Available: <https://www.mdpi.com/1424-8220/24/23/7683>
- [20] H. Firat, M. E. Asker, and D. Hanbay, "Classification of hyperspectral remote sensing images using different dimension reduction methods with 3D/2D CNN," *Remote Sens. Appl., Soc. Environ.*, vol. 25, Jan. 2022, Art. no. 100694.
- [21] L. Hu, W. He, L. Zhang, and H. Zhang, "Cross-domain meta-learning under dual-adjustment mode for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5526416.
- [22] J. Zhao, L. Hu, L. Huang, C. Wang, and D. Liang, "MSRA-G: Combination of multi-scale residual attention network and generative adversarial networks for hyperspectral image classification," *Eng. Appl. Artif. Intell.*, vol. 121, May 2023, Art. no. 106017.
- [23] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522214.
- [24] D. Hong, B. Zhang, X. Li, Y. Li, C. Li, J. Yao, N. Yokoya, H. Li, P. Ghamisi, X. Jia, A. Plaza, P. Gamba, J. Atli Benediktsson, and J. Chanussot, "SpectralGPT: Spectral remote sensing foundation model," 2023, *arXiv:2311.07113*.
- [25] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.
- [26] S. Nesteruk, S. Illarionova, and A. Somov, "Image dataset augmentation a survey and taxonomy," *Meas. Instrum. Mach. Vis.*, vol. 2024, pp. 110–136, Feb. 2024.
- [27] S. Illarionova, S. Nesteruk, D. Shadrin, V. Ignatiev, M. Pukalchik, and I. Oseledets, "MixChannel: Advanced augmentation for multispectral satellite images," *Remote Sens.*, vol. 13, no. 11, p. 2181, Jun. 2021.
- [28] B. Kumar, O. Dikshit, A. Gupta, and M. K. Singh, "Feature extraction for hyperspectral image classification: A review," *Int. J. Remote Sens.*, vol. 41, no. 16, pp. 6248–6287, Jun. 2020.
- [29] V. Podlipnov and R. Skidanov, "Calibration of an imaging hyperspectrometer," *Comput. Opt.*, vol. 41, no. 6, pp. 869–874, 2017.
- [30] S. V. Karpeev, S. N. Khonina, A. R. Murgadulov, and M. V. Petrov, "Alignment and study of prototypes of the Offner hyperspectrometer," *VESTNIK Samara University. Aerosp. Mech. Eng.*, vol. 15, no. 1, pp. 197–206, Apr. 2016.
- [31] N. Kazanskiy, N. Ivliev, V. Podlipnov, and R. Skidanov, "An airborne Offner imaging hyperspectrometer with radially-fastened primary elements," *Sensors*, vol. 20, no. 12, p. 3411, Jun. 2020.
- [32] S. Hu, R. Hou, L. Ming, S. Meifang, and P. Chen, "A hyperspectral image reconstruction algorithm based on RGB image using multi-scale atrous residual convolution network," *Frontiers Mar. Sci.*, vol. 9, Jan. 2023, Art. no. 1006452.
- [33] M. E. Klein, B. J. Aalderink, R. Padoan, G. De Bruin, and T. A. G. Steemers, "Quantitative hyperspectral reflectance imaging," *Sensors*, vol. 8, no. 9, pp. 5576–5618, Sep. 2008.
- [34] N. Firsov, V. Podlipnov, N. Ivliev, P. Nikolaev, S. Mashkov, P. Ishkin, R. Skidanov, and A. Nikonov, "Neural network-aided classification of hyperspectral vegetation images with a training sample generated using an adaptive vegetation index," *Comput. Opt.*, vol. 45, no. 6, pp. 887–896, Dec. 2021.
- [35] J. Jablonski, C. Durell, E. T. Slonecker, K. K. A. Wong, B. Simon, A. Eichelberger, and J. Osterberg, "Best practices in passive remote sensing VNIR hyperspectral system hardware calibrations," *Proc. SPIE*, vol. 9860, pp. 13–42, May 2016.
- [36] H. Aasen, E. Honkavaara, A. Lucieer, and P. J. Zarco-Tejada, "Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows," *Remote Sens.*, vol. 10, no. 7, p. 1091, Jul. 2018.
- [37] J. J. Sousa, P. Toscano, A. Matese, S. F. Di Gennaro, A. Berton, M. Gatti, S. Poni, L. Pádua, J. Hruška, R. Morais, and E. Peres, "UAV-based hyperspectral monitoring using push-broom and snapshot sensors: A multisite assessment for precision viticulture applications," *Sensors*, vol. 22, no. 17, p. 6574, Aug. 2022.
- [38] M. S. Shaikh, K. Jaferzadeh, B. Thörnberg, and J. Casselgren, "Calibration of a hyper-spectral imaging system using a low-cost reference," *Sensors*, vol. 21, no. 11, p. 3738, May 2021.
- [39] F. Cao, Z. Yang, J. Ren, M. Jiang, and W.-K. Ling, "Does normalization methods play a role for hyperspectral image classification?" 2017, *arXiv:1710.02939*.
- [40] I. Mirpulatov, S. Illarionova, D. Shadrin, and E. Burnaev, "Pseudo-labeling approach for land cover classification through remote sensing observations with noisy labels," *IEEE Access*, vol. 11, pp. 82570–82583, 2023.

- [41] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Jun. 2001.
- [42] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, Dec. 2018, pp. 6639–6649.
- [43] Y. Ma, Y. Lan, Y. Xie, L. Yu, C. Chen, Y. Wu, and X. Dai, "A spatial-spectral transformer for hyperspectral image classification based on global dependencies of multi-scale features," *Remote Sens.*, vol. 16, no. 2, p. 404, Jan. 2024.
- [44] N. A. Firsov, V. V. Podlipnov, N. A. Ivliev, D. D. Ryskova, A. V. Pirogov, A. A. Muzyka, A. R. Makarov, V. E. Lobanov, V. I. Platonov, A. N. Babichev, V. A. Monastyrskiy, V. I. Olgarenko, D. P. Nikolaev, R. V. Skidanov, A. V. Nikonov, N. L. Kazanskiy, and V. A. Soifer, "Ensembles of spectral-spatial convolutional neural network models for classifying soil types in hyperspectral images," *Comput. Opt.*, vol. 47, no. 5, pp. 795–805, 2023.
- [45] Z. Zhao, X. Xu, S. Li, and A. Plaza, "Hyperspectral image classification using groupwise separable convolutional vision transformer network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5511817.
- [46] C. Li, B. Rasti, X. Tang, P. Duan, J. Li, and Y. Peng, "Channel-layer-oriented lightweight spectral-spatial network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5504214.



ANDREY MAKAROV was born in 1999. He is currently pursuing the Ph.D. degree with the Department of Supercomputing and General Informatics, Samara University. He is an Engineer with the Laboratory of Intelligent Video Data Analysis, Institute of Image Processing Systems, NRC "Kurchatov Institute."



ISLOMBEK MIRPULATOV was born in 1996. He received the bachelor's and master's degrees in computer science from Lomonosov Moscow State University, Moscow, Russia, in 2019 and 2023, respectively. He is currently pursuing the Ph.D. degree with Skoltech. He is a Research Engineer in Artificial Intelligence Center at Skoltech. His research interests include ML, deep neural networks, and computer vision.



NIKITA FIRSOV was born in 1998. He received the degree from Samara National Research University, in 2021, where he is currently pursuing the master's degree. He is a Researcher with the Mobility AI Center, Samara National Research University. His research interests include image processing, computer vision, neural networks, and hyperspectral images.



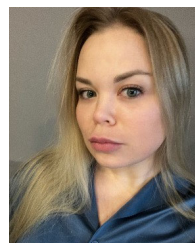
VALERIY LOBANOV was born in 1998. He received the degree from Adygeya State University, in 2022. He is currently pursuing the master's degree with Samara National Research University. He is a Laboratory Assistant with the Mobility AI Center, Samara National Research University. His research interests include image processing, convolutional neural networks, and hyperspectral images.



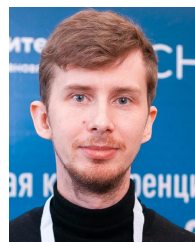
SVETLANA ILLARIONOVA received the bachelor's and master's degrees in computer science from Lomonosov Moscow State University, Moscow, Russia, in 2017 and 2019, respectively, and the Ph.D. degree in computer science from the Skolkovo Institute of Science and Technology, Moscow, in 2023. Her research interests include computer vision, deep neural networks, and remote sensing.



VLADIMIR PODLIPNOV was born, in 1987. He is currently an Engineer with the Lab-35, Samara National Research University. His research interests include mathematical modeling, electron-beam lithography, optimization of etching procedures in microelectronics, and diffractive optics and techniques for surface processing and inspection.



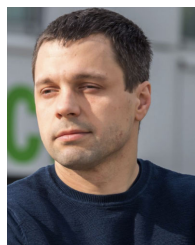
YULIYA VYBORNOVA was born in 1993. She received the degree in information security from Samara National Research University, in 2015. She defended the thesis for the degree of candidate of technical sciences, in 2019. She is currently a Senior Researcher with the Mobility AI Center, Samara National Research University. Her research interests include computer vision, image processing, cryptography, steganography, and digital watermarking.



DMITRII SHADRIN was born in 1992. He received the M.S. degree in applied physics and mathematics from Moscow Institute of Physics and Technology (MIPT), in 2016, and the Ph.D. degree in data science from the Skolkovo Institute of Science and Technology (Skoltech), Russia, in 2020. He is currently the Head of Direction with Skoltech. His research interests include data processing, modeling of physical and bioprocesses in closed artificial growing systems, machine learning, and computer vision. He involved in the development of approaches for monitoring and modeling of the carbon footprint at Skoltech. He is responsible for the experimental research and several projects in the research center in artificial intelligence in the direction of optimization of management decisions to reduce carbon footprint.



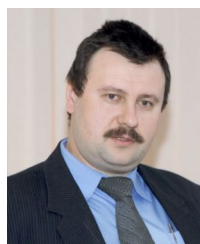
ANDREY RASTORGUEV was born in 1984. He received the degree in optics from the Physics Department, Samara State University, in 2006. In 2023, he defended the Ph.D. thesis on the topic: “Calculation of the characteristics of an onboard optical hyperspectrometer based on the Offner scheme.” He is currently a Research Fellow with the Mobility AI Center, Samara National Research University. His research interests include image processing, neural networks, earth remote sensing, and modeling of hyperspectral equipment.



EVGENY BURNAEV was born in 1983. He received the degree from Moscow Institute of Physics and Technology, in 2006, the degree in science from the Institute for Information Transmission Problems, in 2008, and the Doctor of Physical and Mathematical Sciences degree from Moscow Institute of Physics and Technology, in 2022.

Since 2008, he has been with the Institute for Information Transmission Problems, as the Head of the Data Analysis and Predictive Modeling Laboratory. He is currently the Director of the Skoltech Applied AI Center. His current research include developing new algorithms in machine learning and artificial intelligence, such as deep networks for an approximation of physical models, generative modeling, and manifold learning, with applications to computer vision and 3D reconstruction, and neurovisualization. The results are published in top computer science conferences (ICML, ICLR, NeurIPS, CVPR, ICCV, and ECCV) and journals.

Prof. Burnaev was honored with many awards for his research, including Moscow Government Prize for Young Scientists (2017), the Ilya Segalovich Yandex Science Prize “Best Research Director of Postgraduate Students in the field of computer sciences” (2020), and the Best Paper Award for the research on modeling of point clouds and predicting properties of 3D shapes at the International Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR) (2020).



ROMAN SKIDANOV was born in 1973. He received degree (Hons.) in physics from Samara State University (SamGU), in 1990, and the Doctor of Physical and Mathematical Sciences degree, in 2008.

He is currently the Chief Research Scientist of the Laboratory of Micro- and Nanotechnology, “Institute for Image Processing Systems–Samara” division of the Kurchatov Complex of Crystallography and Photonics, National Research Center “Kurchatov Institute.” He is also a Professor with the Department of Technical Cybernetics, Samara University. He is a specialist in diffraction optics, mathematical modeling, image processing, and nanophotonics. His scientific work includes more than 300 articles and nine monographs.



ARTEM NIKONOROV was born in 1979. He received the degree from Samara National Research University, in 2002, and the Ph.D. and Dr.-Sc. degrees, in 2005 and 2016, respectively. He is currently the Director of the Mobility AI Center, Samara National Research University. He has more than 90 publications and two monographs. His research interests include pattern recognition and image analysis, system identification, parallel and distributed programming, and GPGPU programming.

...