Dual-Stream 2D and 3D-SE-ResNet Architectures for Crop Mapping Using EnMAP Hyperspectral Time-Series

László Mucsi *, Márkó Sóti, Dorottya Litkey-Kovács, János Mészáros, Dóra Vigh-Szabó, Elemér Szalma Jr., Zalán Tobak, József Szatmári

Posted Date: 3 February 2026

doi: 10.20944/preprints202602.0202.v1

Keywords: deep learning; spatial-spectral classification; spatial cross-validation; imaging spectroscopy; precision agriculture; phenology

 doi:10.20944/preprints202602.0202.v1

Article

Dual-Stream 2D and 3D-SE-ResNet Architectures for Crop Mapping Using EnMAP Hyperspectral Time-Series

László Mucsi [1,*,†], Márkó Sóti [2,†], Dorottya Litkey-Kovács [3], János Mészáros [4], Dóra Vigh-Szabó [5], Elemér Szalma, Jr. [6], Zalán Tobak [1] and József Szatmári [1]

[1] Department of Atmospheric and Geospatial Data Science, University of Szeged, Egyetem str. 2, H-6722, Szeged, Hungary
[2] Doctoral School of Geosciences, University of Szeged, Egyetem Str. 2, H-6722 Szeged, Hungary
[3] Lajtamag Ltd. Mosonudvar, Bereki str. 1, H-9246
[4] Institute for Soil Sciences, HUN-REN Centre for Agricultural Research, Fehérvári út.132–144, 1116 Budapest, Hungary
[5] Middle Tisza District Water Directorate, Boldog S. I. Blvd 4, H-5000 Szolnok, Hungary
[6] Cereal Research Institute, Alsó Kikötő sor 9. H-6726 Szeged
* Correspondence: mucsi.laszlo@szte.hu
† These authors contributed equally to this work.

Highlights

What are the main findings?

- Dual-stream 2D and 3D-SE-ResNet architectures achieve superior crop classification accuracy (OA > 97%) using only three EnMAP acquisitions, outperforming pixel-based baselines.
- Feature importance analysis confirms the critical role of the SWIR domain (1000–2450 nm) in resolving phenological ambiguities during early soil-dominated and late senescence stages.

What are the implications of the main findings?

- The study confirms that advanced deep learning architectures can compensate for limited temporal sampling in tasking-based hyperspectral missions such as EnMAP.
- The proposed dual-stream framework supports the operational use of spaceborne imaging spectroscopy for large-scale agricultural monitoring and provides guidance for future hyperspectral mission design and analysis.

Abstract

Deep learning-based crop mapping from hyperspectral satellite data offers immense potential for capturing subtle phenological differences, yet leveraging sparse time-series remains a major methodological challenge. This study evaluates the EnMAP sensor for identifying nine major crop types in the intensive agricultural landscape of Southeastern Hungary. We utilized a limited time series (November, March, August) to benchmark two modeling strategies: a single-date Dual-Stream Spatial-Spectral 2D-CNN (DSS-2D) and a multi-temporal 3D-SE-ResNet. Model performance was assessed using parcel-level spatial cross-validation to ensure realistic accuracy estimates and reduce spatial autocorrelation bias. Results demonstrate that the DSS-2D model achieved superior single-date accuracy (OA > 97%), significantly outperforming pixel-based baselines. Furthermore, the multi-temporal 3D-SE-ResNet achieved a robust seasonal accuracy of 92.9%, effectively compensating for temporal sparsity by exploiting the deep spectral information of the SWIR domain. The study confirms that treating hyperspectral data as a 3D volume enables the extraction of phenological traits even from limited observations. These findings provide a strong proof-of-concept for the operational feasibility of future missions like Copernicus CHIME for continental-scale food security monitoring.

 doi:10.20944/preprints202602.0202.v1

Keywords: deep learning; spatial-spectral classification; spatial cross-validation; imaging spectroscopy; precision agriculture; phenology

1. Introduction

Accurate and timely crop type mapping is a cornerstone of modern agricultural monitoring, supporting food security assessments, yield forecasting, subsidy control, and the implementation of sustainable land management practices. Satellite-based remote sensing has long been recognized as a key technology for large-scale crop mapping due to its synoptic coverage and temporal consistency. In recent years, the increasing availability of multi-temporal satellite observations has shifted research attention toward exploiting crop phenology as a discriminative signal, particularly for spectrally similar crop types that are difficult to separate using single-date imagery alone [1].

Multispectral satellite missions such as Sentinel-2 have proven highly valuable for operational crop mapping owing to their high revisit frequency and global coverage [2–4]. Nevertheless, the limited number and width of spectral bands constrain their ability to capture subtle biochemical and structural differences between crops, especially during key phenological stages or under intensive agricultural management [5,6]. Imaging spectroscopy offers a significant advantage in this respect by providing contiguous narrow spectral bands across the visible, near-infrared, and shortwave infrared regions, enabling a more detailed characterization of vegetation traits related to chlorophyll content, canopy structure, and water status [7–9].

The Environmental Mapping and Analysis Program (EnMAP) mission represents a major milestone in spaceborne hyperspectral Earth observation. EnMAP delivers 224 contiguous spectral bands between 420 and 2450 nm at a spatial resolution of 30 m, providing unprecedented spectral fidelity for terrestrial applications [10]. Unlike multispectral missions, EnMAP operates as a scientific, tasking-based satellite rather than a continuous global monitoring system [11]. As a consequence, EnMAP time series are typically sparse and irregular, consisting of a limited number of phenologically relevant acquisitions [12]. This acquisition paradigm fundamentally alters the crop mapping problem, requiring methods that can extract maximal information from a small number of hyperspectral observations while remaining robust to phenological variability and incomplete temporal sampling [13].

In parallel with advances in sensor technology, deep learning has become the dominant paradigm for remote sensing image analysis. Convolutional Neural Networks (CNNs) have demonstrated superior performance over traditional machine learning approaches by learning hierarchical feature representations directly from the data. Early hyperspectral crop mapping studies relied primarily on spectral-only classifiers such as Random Forests or one-dimensional CNNs [14]. While effective in many cases, these models neglect spatial context and are therefore sensitive to mixed pixels, sparse canopy cover, and soil background effects, particularly during early growth stages and post-harvest periods.

To address these limitations, spatially explicit CNN architectures based on two-dimensional convolutions have been introduced, enabling the joint exploitation of spectral and spatial information within local neighborhoods. Such models have demonstrated clear benefits in agricultural landscapes characterized by large, homogeneous parcels, where spatial texture and field structure provide important discriminatory cues [6]. More recently, three-dimensional CNNs have been proposed to explicitly model spectral–spatial–temporal dependencies by treating image time series as volumetric data. Although promising, most existing 3D CNN studies have focused on dense multispectral time series, leaving their applicability to sparse spaceborne hyperspectral observations largely unexplored, [15].

Another critical yet often overlooked aspect of crop mapping studies is the validation strategy. Pixel-level random sampling can lead to overly optimistic performance estimates due to strong spatial autocorrelation within agricultural fields, an effect that is particularly pronounced for deep learning models. Parcel-level spatial cross-validation has therefore been advocated as a more realistic

 doi:10.20944/preprints202602.0202.v1

evaluation framework for agricultural applications [16,17]. However, despite these recommendations, parcel-based validation remains underutilized in hyperspectral deep learning studies.

Against this backdrop, the present study investigates the potential of sparse EnMAP hyperspectral time-series data for parcel-level crop mapping in a large-scale, operational agricultural environment. Focusing on an intensively managed estate in southeastern Hungary characterized by oversized agricultural parcels, we systematically evaluate deep learning architectures of increasing complexity, ranging from spectral-only baselines to spatial–spectral 2D CNNs and spatio-temporal 3D CNNs with Squeeze-and-Excitation (SE) residual connections. Both single-date, phenology-specific classifications and a seasonal multi-temporal mapping scheme are considered to reflect realistic operational use cases.

A central methodological contribution of this work is the evaluation of a dual-stream spatial–spectral 2D CNN architecture that disentangles texture learning and pixel-wise spectral feature extraction through parallel convolutional pathways. Complementing this approach, a 3D-SE-ResNet model is employed to assess whether joint spectral–spatial–temporal modeling provides additional benefits when only a limited number of hyperspectral acquisitions are available. All models are evaluated using a strict parcel-level spatial cross-validation framework to ensure unbiased and operationally relevant performance estimates.

The specific objectives of this study are to: (i) quantify the discriminative power of single-date EnMAP hyperspectral imagery across key phenological stages; (ii) compare spectral-only, spatial–spectral 2D, and spatio-temporal 3D deep learning architectures under a consistent parcel-level validation scheme; (iii) assess the added value of sparse hyperspectral time-series integration for seasonal crop mapping; The remainder of the paper is structured as follows. Section 2 describes the study area, data, methods, and validation strategy. Section 3 presents the classification results, followed by a discussion of the findings and their implications for operational agricultural monitoring.

2. Materials and Methods

2.1. Study Area

The study area is situated in Southeastern Hungary, within the administrative boundaries of Mezőhegyes (46°19′N, 20°49′E), adjacent to the Romanian border (Figure 1). The region holds a historical significance in agricultural production. Today, the area is managed by the Nemzeti Ménesbirtok és Tangazdaság Zrt., which operates Hungary's largest contiguous agricultural estate. A distinctive feature of this landscape is the prevalence of oversized agricultural parcels (>100 ha), locally referred to as "giant fields." These vast, homogeneous plots are significantly larger than the national average (28 ha in 2023). Combined with fertile chernozem soils, they create an optimal environment for intensive cereal and oilseed cultivation. The large-scale spatial continuity of these fields makes the area particularly suitable for deep learning-based remote sensing analysis.

The dataset for this study was derived from 106 agricultural plots, representing a combined area of 5,739 hectares. The crop composition within these selected fields reflects the estate's diverse portfolio. While the area is particularly renowned for hybrid maize seed production, the studied parcels also encompass significant stands of winter wheat, winter barley, sunflower, alfalfa, and rapeseed, alongside maize cultivated for both silage and grain (fodder) purposes.

 doi:10.20944/preprints202602.0202.v1

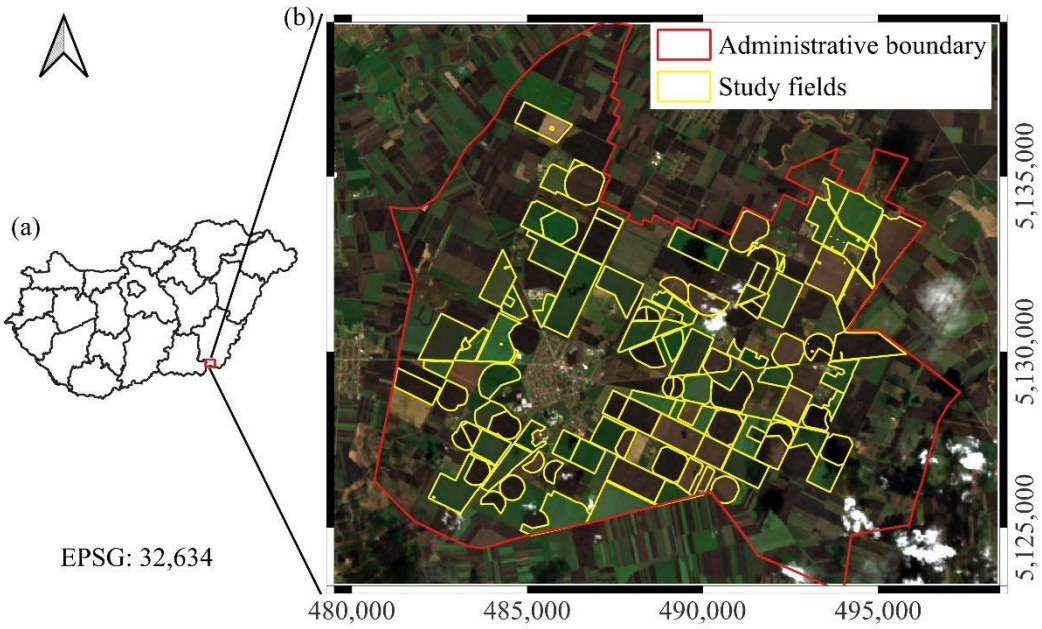


Figure 1. Overview of the study site. (a) Location of the study area within Hungary; (b) The administrative boundary of Mezőhegyes and the investigated study fields, displayed on an EnMAP hyperspectral true color composite (acquired on 13 March 2025). Grid spacing represents 5 km intervals.

2.2. Hyperspectral Dataset

The study utilizes hyperspectral imagery acquired by the EnMAP satellite [10,18]. designed for high-fidelity Earth observation, the sensor covers the spectral range from 420 to 2,450 nm across 224 contiguous bands, providing detailed spectral information in both the visible-near infrared (VNIR) and shortwave infrared (SWIR) domains. The data features a 30 m spatial resolution and a 30 km swath width, enabling the precise discrimination of surface materials and vegetation types [11].

For the 2025 growing season, three acquisitions were available for the study area: 21 November 2024, 13 March 2025, and 10 August 2025 (Figure 2) according to the author's proposal (Proposal No. A00001-P00875) to EnMAP. As EnMAP operates as a scientific mission with on-demand tasking capabilities rather than continuous global monitoring, the availability of cloud-free imagery for specific time windows is limited. Consequently, these three dates represent the only high-quality observations obtainable during the phenological cycle. It should be noted that the image acquired on 13 March was affected by minor scattered cloud cover; however, the impacted areas were masked out, and the remaining clear pixels provided sufficient coverage for the classification of winter crops.

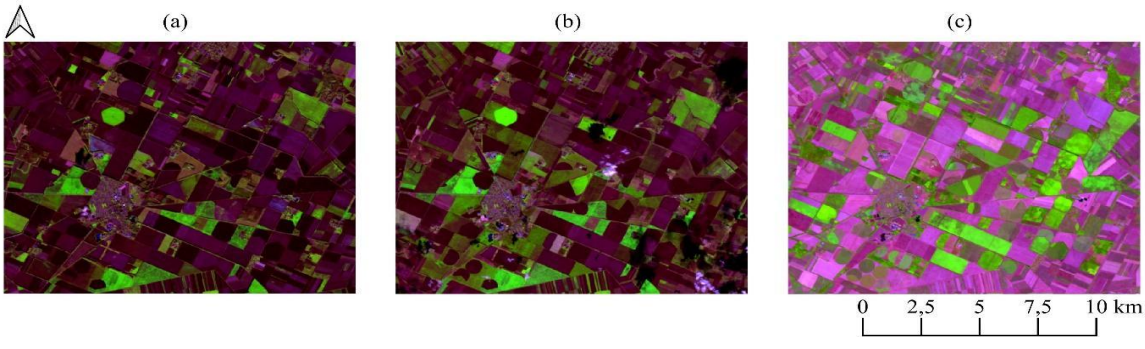


Figure 2. False color composite ENMAP images (R: 1600 nm, G: 850 nm, B: 660 nm) acquired on (a) 21 November 2024, (b) 13 March 2025, and (c) 10 August 2025. Vegetation appears in green, while bare soil is purple. The

 doi:10.20944/preprints202602.0202.v1

variation in purple tones reflects soil moisture: moist soil appears dark purple (Nov, Mar), whereas dry soil and stubble appear light purple (Aug).

2.3. Reference Data and Classification Strategy

2.3.1. Reference Data

The ground truth dataset was constructed using high-precision vector polygons representing the 106 investigated agricultural parcels (Figure 3). These geospatial data were obtained directly from the digital farm management system of the Nemzeti Ménesbirtok és Tangazdaság Zrt. A distinct advantage of this dataset is its reliability, as the parcel boundaries reflect official cadastral limits and the attribute data derive from the estate's actual sowing plans and harvest records. This authoritative source eliminates the uncertainty often associated with visual interpretation, ensuring spatially accurate and thematically correct reference labels for both training and validation.

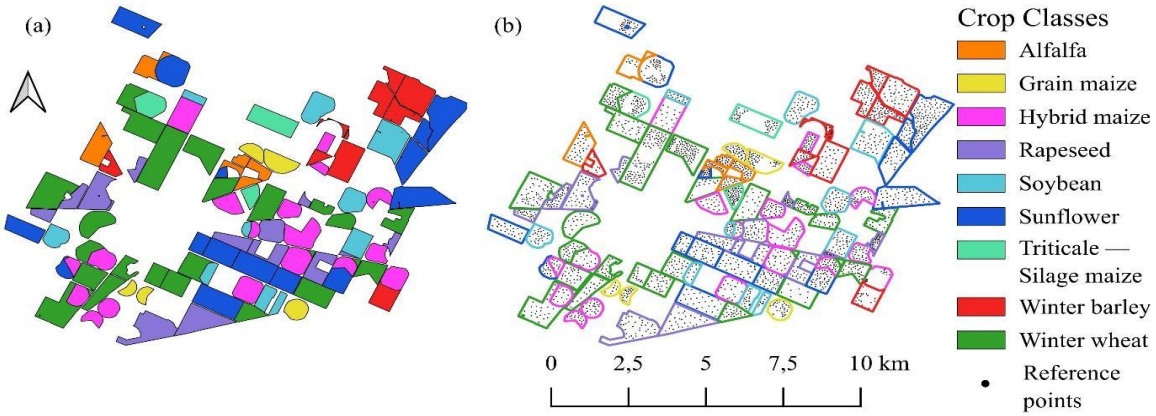
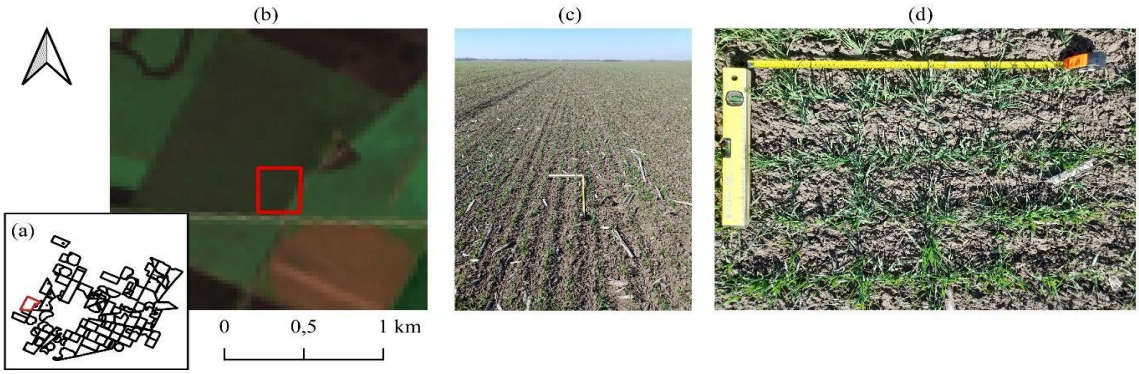


Figure 3. Reference data used in the study. (a) Seasonal crop map derived from official records ("Triticale – Silage Maize" indicates double cropping). (b) Visualization of the sampling strategy showing ground truth points positioned within parcel interiors to avoid edge effects.

A manual sampling strategy was employed to generate the reference dataset (Figure 3). The sampling points were carefully positioned within the parcel interiors to strictly avoid edge effects and mixing with adjacent land cover types. To ensure the model's robustness, the selection process aimed to capture the natural intra-field spectral variability of the crops; therefore, the dataset includes both spectrally homogeneous pixels and those exhibiting typical heterogeneity. However, pixels dominated by soil background noise or ambiguous spectral signatures (outliers) were explicitly excluded to maintain data quality. Crucially, the labeling was performed in a multi-temporal manner: for each spatial sample point, the actual land cover class was recorded independently for all three acquisition dates, thereby capturing the complete phenological status and land use changes throughout the observed timeline. Finally, the reliability of the ground truth points was verified using high-resolution PlanetScope imagery and synchronized field surveys (Figure 4).



 doi:10.20944/preprints202602.0202.v1

Figure 4. Multiscale validation of a winter wheat field (March 2025). (a) Location, (b) EnMAP satellite view, (c) landscape photo, (d) nadir view with measuring tape.

2.3.2. Classification Schemes

To assess the capabilities of the EnMAP hyperspectral imagery, two distinct classification strategies were employed based on the temporal characteristics of the vegetation. The first, a date-specific approach, was designed to reflect the instantaneous phenological status of the fields at each acquisition date. Since the spectral appearance of the landscape changes dynamically, class definitions were adjusted accordingly. For the November 2024 and March 2025 datasets, the analysis focused on active winter crops (winter wheat, winter barley, triticale, rapeseed) and perennials (alfalfa), while areas reserved for spring-sown crops were spectrally characterized as bare soil (Table 1). Conversely, in August 2025, the target shifted to summer crops (hybrid maize, grain maize, silage maize, sunflower, soybean, alfalfa); by this time, winter crops had been harvested, resulting in the reclassification of their locations as bare soil to represent stubble or tilled land.

Table 1. Classification scheme and reference dataset statistics. The table indicates the presence of each crop type during the specific acquisition dates (Nov, Mar, Aug), the number of reference parcels, the total area in hectares, and the number of reference samples (pixels) selected for the modeling process.

Class Name	Nov 2024	Mar 2025	Aug 2025	No. of Fields	Area (ha)	No. of Samples
Alfalfa	✓	✓	✓	6	220	122
Winter wheat	✓	✓	–	24	1532	581
Winter barley	✓	✓	–	10	499	228
Triticale	✓	✓	–	3	203	139
Rapeseed	✓	✓	–	14	667	316
Grain maize	–	–	✓	5	181	134
Hybrid maize	–	–	✓	18	798	360
Silage maize	–	–	✓	3	203	139
Sunflower	–	–	✓	15	1113	501
Soybean	–	–	✓	9	500	238

Complementing the single-temporal analysis, a seasonal crop (multitemporal) classification scheme was defined to map the dominant economic crop of the 2025 growing season. This approach integrates temporal information to eliminate the transient “bare soil” class inherent to the single-date assessments. In this scheme, every pixel is assigned to its primary crop type, regardless of whether it was harvested in early summer or autumn. The resulting thematic map encompasses the complete portfolio of cultivated species, including winter wheat, winter barley, triticale, rapeseed, hybrid maize, grain maize, silage maize, sunflower, soybean, and alfalfa.

2.3.3. Phenological Basis of the Classification

The classification logic is grounded in the available hyperspectral time series, correlating the spectral reflectance properties with the specific development stages of the crops. The phenological status at the time of the three acquisitions was determined using the BBCH scale [19], as illustrated in Figure 5, and substantiated by the mean spectral profiles presented in Figure 6.

 doi:10.20944/preprints202602.0202.v1

	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Alfalfa	97	97	97	09	10-19	31-35	51-61	69	15-51	61-65
Grain maize	-	-	-	-	-	00-09	11-16	19-38	51-65	71-83
Hybrid maize	-	-	-	-	-	00-09	11-16	19-38	51-65	71-83
Rapeseed	16-19	19	19	19-25	30-50	51-65	67-75	78-87	89-99	-
Soybeans	-	-	-	-	-	00-09	10-18	51-63	65-73	75-89
Sunflower	-	-	-	-	-	00-09	12-19	30-39	51-69	71-89
Triticale	22-25	26-29	29	29	30-36	37-49	51-69	71-89	92-99	-
Winter barley	21-25	25-29	29	29	30-34	35-47	49-69	71-89	92-99	-
Winter wheat	13-21	21-25	25	25-29	30-32	32-37	39-59	61-77	81-99	-

BBCH code	Phenological stage
0-09	Germination and emergence
10-29	Leaf development
30-39	Stem elongation
40-49	Advanced vegetative growth
50-59	Inflorescence emergence / Flower bud formation
60-69	Flowering
70-89	Fruit development and grain filling
90-99	Maturity and senescence / dormant period

Figure 5. Phenological calendar of the studied crops based on the BBCH scale. Colors indicate major growth stages, while the months highlighted in red correspond to the satellite image acquisition dates.

The acquisition on November 21, 2024, coincides with the early vegetative phase of winter crops. At this stage, winter wheat, barley, rapeseed, and triticale (BBCH 19–25) exhibit active photosynthetic activity, resulting in a distinct vegetation signal (Figure 6a). This allows for their spectral discrimination from parcels reserved for spring sowing, which are characterized by the reflectance features of bare soil or crop residues.

By the time of the March 13, 2025 acquisition, winter crops reach the stem elongation phase (BBCH 30–35). This stage is characterized by rapid biomass accumulation and canopy closure, providing a strong green vegetation reflectance (Figure 6b). In contrast, the fields designated for summer crops (maize, sunflower, soybean) are in the seedbed preparation phase, where the spectral response is dominated by the soil background. This phenological divergence creates the maximum contrast between winter and summer cropping systems.

The August 10, 2025 image documents the phenological inversion of the landscape. Summer crops are in their reproductive or ripening stages (BBCH 65–80), showing high biomass and vigorous spectral reflectance. Conversely, winter crops have been harvested, presenting the spectral signature of dry stubble or ploughed soil. Crucially, the timing of this acquisition enables the detection of double cropping: while standard cereal stubble is spectrally inactive, fields re-sown with silage maize show a renewed vegetative signal (BBCH 13–15), allowing for their identification based on active chlorophyll reflectance (Figure 6c).

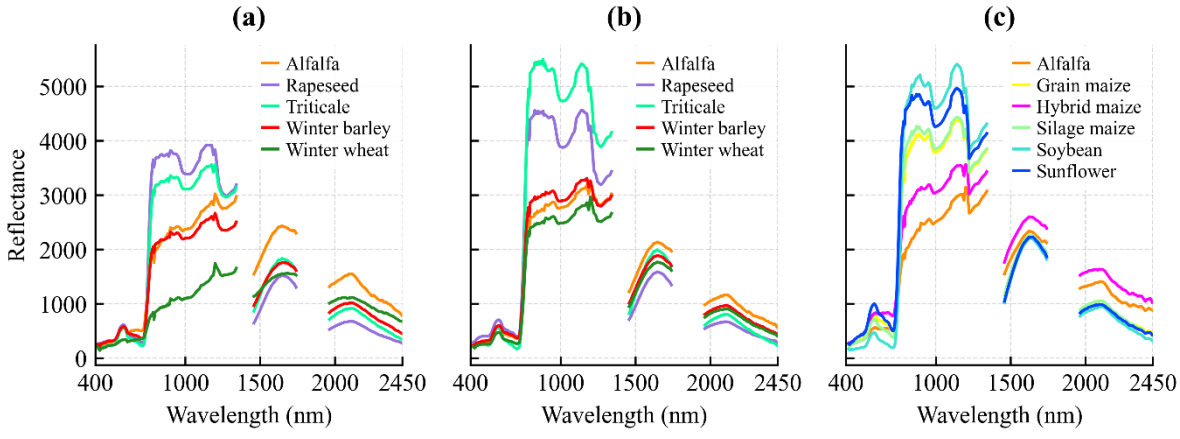


Figure 6. Mean spectral reflectance profiles of the classified crops for (a) Nov 21, 2024, (b) Mar 13, 2025, and (c) Aug 10, 2025.

 doi:10.20944/preprints202602.0202.v1

2.4. Machine Learning and Deep Learning Implementation

The classification workflow was implemented in a high-performance Python 3.x environment using the TensorFlow 2.x deep learning framework (Keras API). Initial spatial processing and reference data preparation were conducted using QGIS 3.4 extended with the EnMapBox 3.16 plugin. Data standardization, performance metrics, and the baseline Random Forest classifiers were implemented using the Scikit-learn library. All experiments were executed in a GPU-accelerated environment to handle the computational load of the hyperspectral data cubes. To ensure the reproducibility of the results, a global deterministic seed (seed=42) was applied to all random number generators prior to the training and validation procedures.

2.5. Single-Date Classification Models

To evaluate the spectral discriminative power of the EnMAP imagery at specific phenological stages, three distinct modeling approaches were applied independently to each acquisition date (November, March, and August).

2.5.1. Random Forest (RF) – Baseline

As a robust benchmark for shallow learning, a Random Forest classifier was implemented. The model processes the input as 1D spectral vectors (1 × 219) without considering spatial neighborhood information. The ensemble was configured with 500 decision trees to ensure higher stability compared to standard implementations. Trees were grown to their full depth using the Gini impurity criterion, and bootstrap sampling was enabled to improve generalization. [20]

2.5.2. 1D Convolutional Neural Network (1D-CNN)

To assess the benefits of deep spectral feature extraction, a 1D-CNN was developed [21]. Unlike the RF, this model utilizes convolutional filters to learn hierarchical features directly from the continuous spectral curve. The architecture consists of a sequence of Conv1D layers (32 and 64 filters) with kernel sizes ranging from 3 to 5, designed to capture local spectral absorption features. Each convolutional block is followed by Batch Normalization and ReLU activation, concluding with a dense layer for classification.

2.5.3. Dual-Stream Spatial-Spectral 2D CNN (DSS-2D)

The proposed DSS-2D architecture is designed to disentangle and independently learn the spatial texture and spectral material signatures before fusion. The network accepts 3D input patches of size 5 × 5 × 219 ($H \times W \times C$) and processes them through two parallel, independent streams:

1. Spatial Branch: This stream utilizes a standard 2D Convolutional layer with 3 × 3 kernels (64 filters, padding='same'). By operating on the local neighborhood, this branch is responsible for capturing texture features, edges, and inter-pixel spatial dependencies within the 5 × 5 window.

2. Spectral Branch: This stream employs 1×1 point-wise convolutions (64 filters). This operation acts as a pixel-wise spectral feature extractor. It projects the high-dimensional spectral vector (219 bands) into a lower-dimensional feature space without mixing information from neighboring pixels, thereby preserving the purity of the spectral signature:

$$y_{i,j} = \sigma\left(b + \sum_{c=1}^{C_{in}} \sum_{u=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} \sum_{v=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} W_{u,v,c} \cdot x_{i+u,j+v,c}\right)$$

where:

- $y_{i,j}$: the output value of the feature map at spatial position (i, j).
- $\sigma(\cdot)$: the non-linear activation function (ReLU).
- b: the bias term.
- c: the index of the input channel (ranging from 1 to C)
- u, v: the spatial indices of the kernel relative to the center position (ranging from - $[K/2]$ to $[K/2]$)

 doi:10.20944/preprints202602.0202.v1

- C_{in}: the number of input channels.
- W: the learnable weights of the kernel.
- x: the input feature map value at the given position.
- K: the kernel spatial dimension, where K=3 for the Spatial Branch (capturing local texture and neighborhood context) and K=1 for the Spectral Branch (performing pixel-wise spectral feature extraction).

Both branches utilize BN and the ReLU activation function. The feature maps from the spatial and spectral streams are then concatenated along the channel dimension. To prevent overfitting and reduce the total number of parameters, the fused features are spatially compressed using Global Average Pooling (GAP), which transforms the 5 × 5 feature maps into a 1D vector. This vector is fed into a fully connected Dense layer with 256 neurons (followed by a Dropout of 0.5) to learn high-level non-linear combinations. Finally, a Softmax output layer produces the class probabilities.

2.6. Multi-Temporal Seasonal Classification Models

To evaluate the added value of temporal information, a second set of models was developed to perform classification using the full seasonal time series (November, March, August) simultaneously.

2.6.1. Multi-Temporal Random Forest (Seasonal RF)

A baseline multi-temporal classifier was implemented by extending the Random Forest approach. The spectral bands from all three acquisition dates were stacked into a single feature vector (3 × 219 = 657 features per pixel). The model hyperparameters (500 trees) were kept identical to the single-date RF to ensure a direct comparison of the temporal integration effect.

2.6.2. 3D-ResNet with Squeeze-and-Excitation (3D-SE-ResNet)

To fully leverage the complex spatio-temporal structure of the EnMAP data, a 3D Convolutional Neural Network was designed. Unlike the 2D approach, this architecture treats the time series as the third dimension of the input volume, processing 4D patches of size 3 × 5 × 5 × 219 (Time × Height × Width × Bands) (Table 3). The backbone is based on the ResNet architecture adapted for 3D data, where the fundamental operation is the 3D convolution [22]. This allows the model to simultaneously capture local spatial interactions and temporal evolution:

$$v_{i,j,t}^{k} = \sigma\left(b^{k} + \sum_{c=1}^{C_{in}} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} w_{p,q,r,c}^{k} \cdot x_{i+p,j+q,t+r,c} \right)$$

To prevent gradient vanishing and enable deeper feature extraction, identity shortcuts (residual connections) are employed. The output of a residual block Y is defined as the summation of the input X and the learned residual mapping

$$F\colon Y = F(X, \{W_i\}) + X$$

To explicitly model the interdependencies between spectral bands and temporal phases, Squeeze-and-Excitation (SE) blocks were integrated within the residual units. This mechanism adaptively recalibrates channel-wise feature responses in two steps: Squeeze (global compression) and Excitation (adaptive weighting):

$$z_c = \frac{1}{H \times W \times T} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{t=1}^{T} u_{i,j,t,c}\,, \qquad \widetilde{x_c} = \sigma\big(W_2 \delta(W_1 z_c)\big) \cdot u_c$$

where:

- $v_{i,j,t}^{k}$: the output value at spatial position (i, j) and temporal index t.
- R: the temporal kernel depth.
- $F(.)$: the residual mapping function.
- z_c: the global average statistic for the c-th channel.

 doi:10.20944/preprints202602.0202.v1

- u: the feature map before recalibration.
- $\widetilde{x_c}$: the final recalibrated feature map.
- W_1, W_2: learnable weights of the SE block.
- C_{in}: the number of input channels (219 bands).

Table 2. Summary of the baseline and deep learning models, including input shapes and architectural specifications used in the study.

Model Name	Type	Input Shape (H x W x C)	Architecture Details
RF	Baseline (Shallow)	1 x 219	Ensemble of 500 trees; Gini impurity; Bootstrap sampling enabled. No spatial context.
Seasonal RF	Baseline (Multi-temporal)	1 x 657	Stacked input of 3 dates (3 x 219 bands). Same configuration as single-date RF (500 trees).
1D-CNN	Deep Learning (Spectral)	1 x 219 x 1	3 distinct Conv1D layers (filters: 32, 64); Kernel sizes: 3-5; Batch Normalization + ReLU; MaxPooling1D.
DSS-2D	Deep Learning (Spatial-Spectral)	5 x 5 x 219	**Dual-Stream Architecture:** • **Spatial Branch:** 3x3 Conv2D (Context & texture learning) • **Spectral Branch:** 1x1 Conv2D (Pixel-wise spectral extraction) • **Fusion:** Concatenation → Global Average Pooling (GAP) → Dense (256) → Softmax
3D-SE-ResNet	Deep Learning (Spatio-Temporal)	3 x 5 x 5 x 219	**3D-CNN Backbone:** • 4D Input Tensor (Time, H, W, Bands) • ResNet blocks with Conv3D layers • **Squeeze-and-Excitation (SE)** blocks for channel-wise attention.

2.7. Training Protocol

All deep learning models (1D-CNN, DSS-2D, and 3D-SE-ResNet) were trained using a unified protocol (Table 3). The Adam optimizer was employed with an adaptive learning rate initialized at 1×10^{-4}. The Sparse Categorical Crossentropy served as the loss function. Models were trained for 40 to 60 epochs with a batch size of 16 or 32, depending on the memory requirements of the architecture. An Early Stopping mechanism was implemented (patience = 8–12 epochs), which terminated training if the validation loss ceased to improve, automatically restoring the best model weights.

Table 3. Hyperparameter settings and training configuration employed for the deep learning models.

Reproducibility	Random Seed	Global deterministic seed (seed=42)
Optimization	Optimizer	Adam (Adaptive Moment Estimation)
	Learning Rate	Initial: 1e-4 (0.0001)

 doi:10.20944/preprints202602.0202.v1

	Loss Function	Sparse Categorical Crossentropy
Training	Batch Size	16 – 32 (depending on model memory usage)
	Epochs	40 – 60
Regularization	Early Stopping	Patience: 8–12 epochs (Monitor: Validation Loss)
	Techniques	Dropout (0.3–0.5), Batch Normalization
Augmentation	Geometrical (2D/3D)	Random Rotations: 90°, 180°, 270° (Applied to DSS-2D and 3D-SE-ResNet)

2.8. Validation

To address the high spatial autocorrelation of agricultural imagery and prevent data leakage, a strict spatial field-based splitting strategy was implemented. Unlike standard random sampling, this approach assigns all pixels within a specific agricultural parcel exclusively to the same partition, ensuring the complete spatial independence of the testing set within the proposed framework.

The evaluation followed a repeated stratified k-fold scenario, consisting of 5 spatial splits, each repeated 5 times with different initialization seeds (totaling 25 independent runs) (Figure 7). To ensure rigorous comparability, all single-date models were evaluated on identical cross-validation splits. Similarly, the multi-temporal models shared a consistent set of partitions, which were distinct from the single-date splits. In every iteration, after assigning fields to the test partition, exactly 50 samples per class were randomly selected from these current test fields to form the final balanced evaluation set.

For each run, feature standardization was fit solely on the training data to prevent leakage, and geometrical augmentation (random rotations) was applied to the training patches. During the optimization phase, the training data was further divided into an 80:20 ratio, where 20% of the samples served as an internal validation set to monitor performance and trigger Early Stopping.

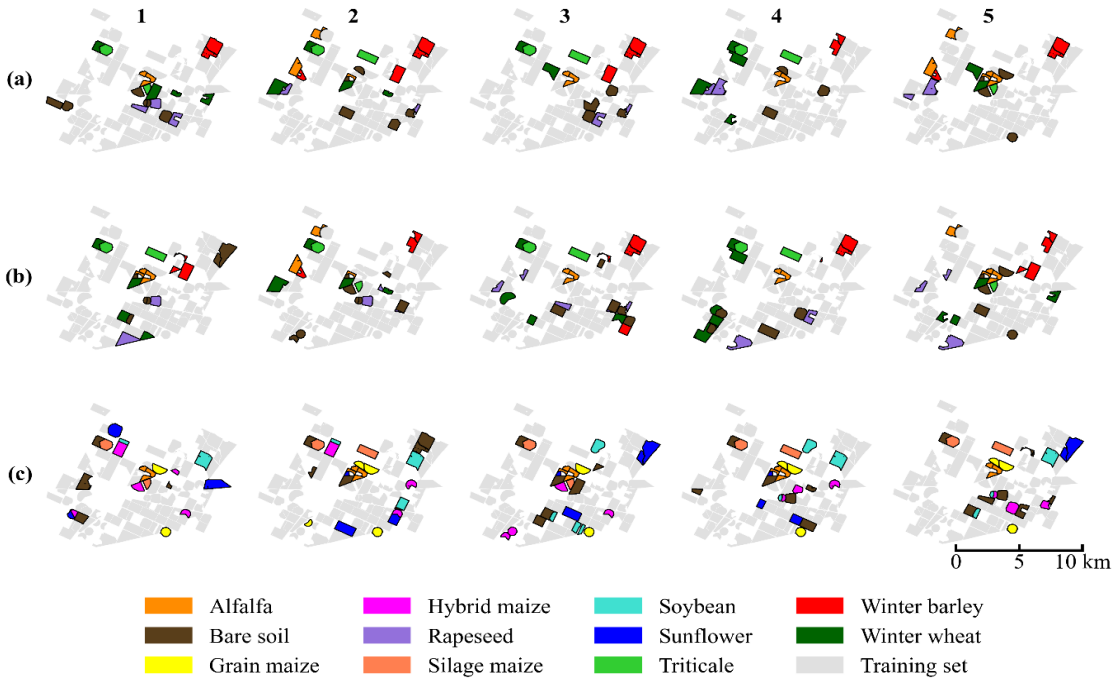


Figure 7. Spatial distribution of training (gray) and testing (colored) samples across the five random scenarios (Columns 1–5) and three datasets: (a) November, (b) March, and (c) August. Due to the limited availability of parcels in certain minority crop classes, some specific fields appear in the test set across multiple splits to ensure adequate representation.

 doi:10.20944/preprints202602.0202.v1

2.8.1. Accuracy Assessment Metrics

To quantitatively assess classification performance, standard accuracy metrics were derived from the population error matrix of size $C \times C$, where C represents the number of crop classes. The diagonal elements $n_{(ii)}$ indicate the number of correctly classified samples for class i, while off-diagonal elements represent misclassifications. The Overall Accuracy (OA) was calculated to represent the proportion of correctly classified samples relative to the total number of evaluation samples:

$$OA = \frac{\sum_{i=1}^{C} n_{ii}}{N}$$

where N is the total number of test samples. While OA provides a global performance measure, it may not fully reflect the model's ability to distinguish individual crop types, particularly in imbalanced scenarios. Therefore, Producer's Accuracy (PA) (Recall) and User's Accuracy (UA) (Precision) were computed for each class to quantify omission and commission errors, respectively:

$$PA_i = \frac{n_{ii}}{n_{+i}}, \quad UA_i = \frac{n_{ii}}{n_{i+}}$$

where n_{+i} denotes the total number of ground truth samples for class i (column total), and n_{i+} represents the total number of samples predicted as class i (row total). In the results section, the False Discovery Rate (FDR) is also reported to explicitly highlight commission errors $(FDR_i = 1 - UA_i)$.

The F1-Score, defined as the harmonic mean of PA and UA, was employed to evaluate the classification performance for each individual crop class:

$$F1_i = 2 \times \frac{PA_i \times UA_i}{PA_i + UA_i}$$

For the multi-temporal model comparison, the Macro-averaged F1-Score was used to ensure that all classes contribute equally to the final score, regardless of sample size variations in the training set.

3. Results

3.1. Performance of Single-Date Classification Models

3.1.1. Efficiency Assessment of Single-Date Classification

This study addresses the critical operational challenge of single-date mapping, determining the immediate land cover status at specific phenological moments without reliance on full temporal sequences. The primary objective of this experimental phase was to evaluate how effectively the proposed DSS-2D architecture can extract identifying features from a single observation. This capability offers a "real-time" alternative to retrospective seasonal monitoring.

The comparative analysis demonstrates that the DSS-2D model provides a highly reliable solution for snapshot classification, as it consistently outperforms both the RF baseline and the spectral-only 1D-CNN across all acquisition dates. Pixel-based methods proved highly sensitive to the lack of temporal context and struggled to resolve spectral ambiguities in both early and late phenological stages. In contrast, the DSS-2D model successfully compensated for this limitation by integrating spatial texture features.

In terms of algorithmic stability, the proposed method exhibited remarkable robustness across the 25 independent experimental runs. The standard deviation of the OA remained below 2.0% for the acquisitions with developed canopy cover, specifically ±1.4% in March and ±1.9% in August. In the early vegetative phase for autumn seeded crops (November), the deviation was slightly higher (±2.9%). This increased variability reflects the inherent difficulty of detecting sparse vegetation signals against a dominant soil background, yet it remains significantly lower than the volatility observed in the 1D-CNN baseline (e.g., ±6.2% in August).

3.1.2. November Acquisition (Early Vegetative Phase)

 doi:10.20944/preprints202602.0202.v1

The November acquisition (BBCH 13-21) served to test the ability of the models to identify crops during the early emergence phase. At this stage, the spectral signal is overwhelmingly dominated by the soil background. Winter crops have a low LAI and are spectrally similar to bare soil or stubble fields. This ambiguity caused the pixel-based 1D-CNN to struggle significantly, and it achieved a Recall of only 74.8% for the Winter Wheat class. Although the RF baseline proved more resilient to this spectral mixing with a Recall of 84.8%, it still failed to match the detection capability of the proposed method. This comparison highlights the limitations of pure spectral monitoring in the early season, which often leads to a systematic underestimation of the cultivated area.

The DSS-2D model successfully mitigated this issue by leveraging the spatial branch of its architecture. Even in the absence of a closed canopy, the convolutional layers effectively detected the faint, periodic row structures and tillage patterns characteristic of emerging vegetation. Consequently, the model improved the Winter Wheat Recall to 86.1% and achieved a Global OA of 97.2%. Furthermore, for Silage Maize, which is present as stubble or bare soil at this time, the model maintained an FDR of just 0.5%. These results confirm that for operational tasks requiring early-season mapping, spatial features provide discriminative information that spectral data alone cannot supply.

3.1.3. March Acquisition (Stem Elongation and Canopy Closure)

The March acquisition (BBCH 30-39) represented the active vegetative phase where winter crops reached stem elongation and maximum canopy closure. The challenge here shifted from background noise to the spectral similarity between biologically related species. The RF baseline struggled to differentiate Winter Barley from Winter Wheat and achieved a Recall of 83.6% for the barley class. The spectral signatures of these two cereals are nearly identical in the visible and near-infrared domains during this intensive vegetative phase, leading to “salt-and-pepper” noise in the pixel-based classification maps

The DSS-2D model resolved this ambiguity entirely and achieved a near-perfect OA of 99.2%. Error decomposition analysis revealed an Allocation Disagreement of 0.0%, implying that the model made virtually no spatial errors in boundary detection. Specifically, it achieved an FDR of 0.0% and 96.3% Recall for Winter Barley. This demonstrates that at this phenological stage, the distinct canopy textures of barley and wheat allow for their precise separation. These textural differences are likely related to variations in sowing density and leaf orientation, making the March acquisition an optimal window for differentiating winter cereals using spatial-spectral methods.

3.1.4. August Acquisition (Reproductive Phase and Maturity)

The August acquisition (BBCH 70-89) tested the models on the most taxonomically complex task. This involved separating fully developed maize varieties, including Silage Maize, Grain Maize, and Hybrid Maize, during the peak growing season (Figure 8). These crops share almost identical spectral traits but differ in agronomic management and canopy architecture, such as planting density and tassel distribution. Standard spectral classifiers failed to reliably distinguish these crops. The RF model yielded an FDR of 12.3% for Silage Maize, meaning that more than one in ten pixels classified as silage maize was incorrect.

In strong contrast, the DSS-2D model demonstrated the decisive advantage of spatial-spectral learning. It achieved a global OA of 98.9% and reduced the FDR for Silage Maize to a negligible 1.9%. This performance leap, which reduced the error rate by a factor of six compared to the baseline, confirms that the convolutional features successfully encoded the textural differences associated with crop height and structure. The results prove that accurate discrimination of agronomically distinct but spectrally similar varieties is achievable with single-date imagery, provided that spatial context is adequately modeled.

 doi:10.20944/preprints202602.0202.v1

Table 4. Performance assessment of single-date models across distinct phenological phases. The table summarizes the Overall Accuracy (OA) and the specific detection capability (Recall, FDR) for the critical classes.

Acquisition	Model	OA	Critical Class	Recall (PA)	FDR
November	RF	96.1 ± 3.4%	Winter Wheat	84.8%	6.2%
	1D-CNN	94.3 ± 5.5%		74.8%	5.7%
	DSS-2D	**97.2 ± 2.9%**		**86.1%**	**1.4%**
March	RF	95.3 ± 4.0%	Winter Barley	83.6%	8.3%
	1D-CNN	96.9 ± 3.2%		87.6%	1.3%
	DSS-2D	**99.2 ± 1.4%**		**96.3%**	**0.0%**
August	RF	93.7 ± 3.7%	Silage Maize	92.8%	12.3%
	1D-CNN	89.3 ± 6.2%		86.2%	12.0%
	DSS-2D	**98.9 ± 1.9%**		**99.4%**	**1.9%**

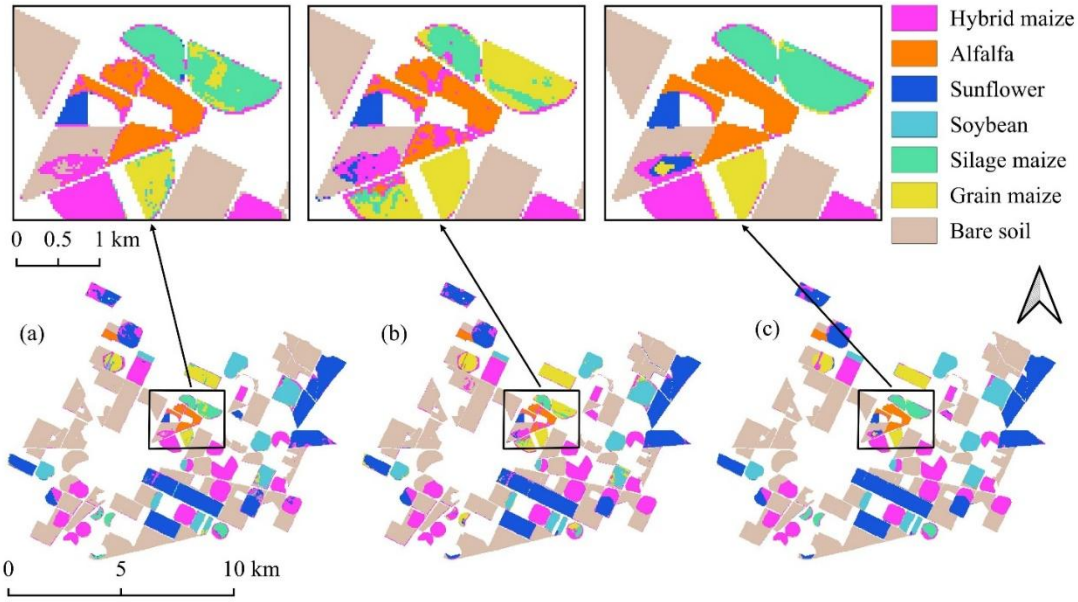


Figure 8. Classification maps derived from the August imagery using RF, 1D-CNN, and DSS-2D methods.

3.2. Multi-Temporal Classification Results

3.3.1. Global Accuracy Assessment

The classification performance of the multi-temporal models was evaluated using the complete three-date EnMAP dataset to determine the efficacy of deep learning in handling high-dimensional spectral time series. The proposed 3D-SE-ResNet achieved an Overall Accuracy (OA) of 92.9%, demonstrating a substantial improvement over the Seasonal RF baseline, which achieved 84.5% (Table 5.). This performance gap highlights the limitations of traditional machine learning when applied to hyperspectral data cubes. The Seasonal RF, processing a stacked feature vector of 657 variables, exhibited signs of the "curse of dimensionality," where the high correlation between contiguous spectral bands introduced significant spectral redundancy, hampering the classifier's generalization capability. In contrast, the 3D-SE-ResNet effectively utilized the spatio-temporal volume, with the Squeeze-and-Excitation mechanisms successfully prioritizing relevant spectral channels while suppressing noise.

 doi:10.20944/preprints202602.0202.v1

Table 5. Multi-temporal classification performance. Comparison of Global Overall Accuracy and class-specific metrics (Recall, FDR) for the Seasonal RF baseline and the 3D-SE-ResNet model.

Model	OA	Critical Class	Recall (PA)	FDR
Seasonal RF	84.5 ± 6.2%	Silage Maize	90.7%	4.4%
		Soybean	78.6%	17.9%
		Winter Barley	78.6%	2.3%
3D-SE-ResNet	92.9 ± 4.2%	Silage Maize	99.6%	4.6%
		Soybean	93.2%	2.8%
		Winter Barley	98.4%	0.9%

3.3.2. Class-Specific Evaluation

The superiority of the deep learning architecture is most pronounced in crops with distinct but complex phenological trajectories. For Silage Maize, the 3D-SE-ResNet achieved a near-perfect Recall of 99.6%, whereas the Seasonal RF reached only 90.7%. The contrast was even sharper for Soybean, where the baseline struggled with severe confusion, resulting in a critically high FDR of 17.9%. This means that nearly one in five pixels classified as soybean by the Seasonal RF was incorrect. The deep model reduced this error rate to a negligible 2.8%, accurately capturing the rapid vegetative development characteristic of broad-leaved summer crops.

Furthermore, the analysis of Winter Barley provides insight into the models' handling of "presence-absence" temporal patterns, as the crop is already harvested by the August acquisition. The Seasonal RF failed to robustly model this transition, resulting in a low Recall of 78.6%. The 3D-SE-ResNet, however, effectively encoded the senescence and harvest signal, achieving a Recall of 98.4%. This confirms that the spatio-temporal filters can adaptively learn diverse phenological profiles, distinguishing between active vegetation and stubble within the same growing season better than the stacked-vector approach.

3.3. Relationship Between Spectral Importance and Biophysical Development

The dynamic shift in feature importance across the growing season confirms that the decision boundaries are governed by the transition from soil-dominated to vegetation-dominated signals. The model tracks key biophysical parameters that characterize each phenological stage by leveraging specific EnMAP bands.

In the November acquisition, the combined importance of SWIR II (30.1%) and SWIR III (10.9%) accounts for 41% of the decision weight (Figure 9). This high significance is a direct consequence of the sparse canopy cover typical of winter cereals during early establishment. At this stage, the classification is primarily driven by soil moisture and organic matter content, which are sensed through the specific absorption properties of the long-wave infrared spectrum. This reliance on soil-background contrast allows the model to differentiate fields even when the winter wheat and barley are only in the initial leaf development phase.

In the March image, the model reallocates its attention toward the VIS (34.6%) and NIR (26.1%) regions, which together account for 60.7% of the total importance. This shift corresponds to the rapid increase in chlorophyll concentration and above-ground biomass during the stem elongation phase. The peak significance of the VIS spectrum indicates that the model distinguishes winter wheat from winter barley based on subtle variations in pigment composition and nitrogen uptake. Simultaneously, the increased NIR contribution reflects the development of the leaf cellular structure, which serves as a proxy for canopy density and vigor as the crops move toward their peak vegetative state.

 doi:10.20944/preprints202602.0202.v1

In the August dataset, the hyperspectral SWIR I (18.8%) range becomes particularly critical. As summer crops reach maturity and senescence begins, the model monitors the water-to-dry-matter ratio within the plant tissues. The NIR importance of 27.5% is coupled with SWIR I signals to track changes in leaf dry matter, cellulose, and lignin content. For maize varieties, this transition zone is essential for detecting the onset of grain hardening and dry down. The fact that SWIR III importance reaches its minimum of 7.7% during this period confirms that the model effectively ignores soil background noise when the water-rich, dense vegetation is at its physiological peak.

The temporal progression from soil mineralogy in November to pigment and biomass in March and finally to dry matter and structure in August validates that the classification is driven by actual physiological changes. The results prove that the model successfully internalizes the seasonal physics of the agricultural landscape to ensure robust performance across diverse crop types.

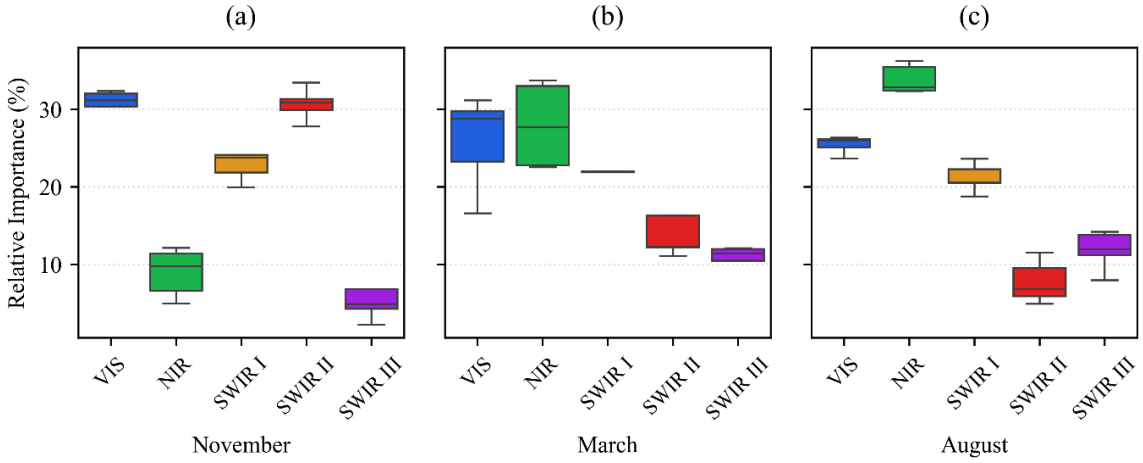


Figure 9. Relative importance of spectral regions for crop classification at three distinct phenological stages using the DSS-2D model and EnMAP hyperspectral data. (a) November; (b) March; (c) August. The spectral bands are VIS (400–700 nm), NIR (700–1000 nm), SWIR I (900–1390 nm), SWIR II (1480–1760 nm), and SWIR III (1950–2450 nm). The boxplots illustrate the distribution of feature importance across 25 model iterations, where the horizontal black line denotes the median and the box represents the interquartile range.

3.4. Computational Efficiency and Training Convergence

To assess the operational feasibility of the proposed architectures, we benchmarked their training stability and computational costs (Figure 10). The validation accuracy curves (Figure 10a) demonstrate that the spatial-spectral DSS-2D model achieves the fastest convergence, stabilizing at high accuracy (>98%) within fewer epochs compared to the pixel-based methods. The 1D-CNN baseline exhibits signs of saturation at a lower accuracy level, confirming its limited capacity to handle spectral variability without spatial context. The multi-temporal 3D-SE-ResNet requires a longer training phase to learn complex phenological features but achieves stable convergence, effectively avoiding overfitting despite the higher parameter count.

 doi:10.20944/preprints202602.0202.v1

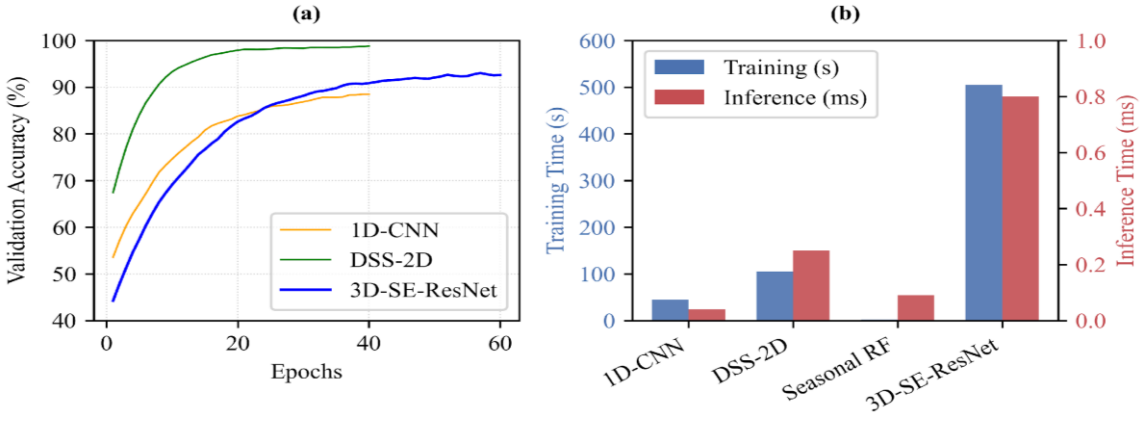


Figure 10. Assessment of model efficiency. (a) Mean validation accuracy curves showing training convergence. (b) Comparison of average training times (blue bars, left axis) and inference times per sample (red bars, right axis).

In terms of computational load (Figure 10b), a significant trade-off is observed. The Seasonal RF baseline is computationally negligible during training (< 2 s), whereas the deep learning models require significantly higher resources. The 3D-SE-ResNet is the most intensive, requiring approximately 505 s per fold due to the complexity of 3D convolutions. However, for operational deployment, inference time is the decisive metric. While the RF remains the fastest (~0.09 ms/sample), the proposed 3D-SE-ResNet maintains a feasible throughput of ~0.80 ms/sample. This confirms that while deep learning models demand substantial training resources, their inference speed is sufficiently fast for the operational mapping of large-scale EnMAP scenes, justifying the additional computational cost for the gained classification accuracy.

4. Discussion

The primary objective of this study was to evaluate the suitability of EnMAP hyperspectral imagery for agricultural mapping in an intensive production environment. The results confirm our working hypotheses regarding the superiority of spatial-spectral deep learning and the critical role of the SWIR domain. This section interprets the findings in the context of algorithmic performance, spectral discriminability, and operational feasibility under data-constrained conditions.

4.1. Superiority of Spatial-Spectral Deep Learning over Pixel-Based Methods

Our first hypothesis posited that CNNs would outperform traditional baseline models. The results strongly corroborate with the DSS-2D model consistently achieving higher accuracy (OA > 97%) compared to the RF and 1D-CNN baselines across all acquisition dates.

The failure of the 1D-CNN to surpass even the RF baseline in certain scenarios offers a critical insight into the relationship between model architecture and sample size. Deep learning models typically require vast amounts of labeled data to generalize effectively. In our study, the training dataset was inherently limited. The 1D-CNN, relying solely on spectral vectors, lacked the capacity to generate sufficient feature variance from the limited samples, leading to overfitting and sensitivity to spectral noise.

In contrast, the success of the DSS-2D architecture can be attributed to its ability to exploit spatial context and data augmentation. By processing 5×5 patches, the model learned textural features, such as row spacing and canopy roughness, which are more robust against intra-class spectral variability than raw reflectance values. Furthermore, the 2D nature of the input allowed for geometrical augmentation, effectively multiplying the training variance without requiring additional ground truth collection. This confirms that for operational mapping with limited reference data, integrating spatial dimensions is not merely an accuracy booster but a prerequisite for the stable convergence of deep learning models.

 doi:10.20944/preprints202602.0202.v1

4.2. The Critical Role of the SWIR Domain (1000–2450 nm)

The feature importance analysis validates the hypothesis that the extended spectral range of EnMAP provides identifying information that standard multispectral sensors cannot fully capture. While visible and NIR bands drove the classification during the peak vegetative stage, the SWIR ranges proved indispensable during the early and late phenological phases.

Specifically, the high contribution of SWIR II and III in November highlights the importance of characterizing soil background properties, including moisture and organic matter, to differentiate sparse winter crops from bare soil. Similarly, the dominance of SWIR I in August confirms that monitoring ligno-cellulose content and crop water stress is essential for separating spectrally similar maize varieties during senescence. These findings underscore that for accurate crop monitoring outside the peak greenness window, the 1000–2450 nm spectral domain is a decisive factor, justifying the use of hyperspectral sensors over cheaper multispectral alternatives.

4.3. Seasonal Mapping with Sparse Temporal Sampling

A key operational question was whether a seasonal land cover map could be derived from a limited number of hyperspectral acquisitions, as opposed to the dense time series required by multispectral approaches. Our results demonstrate that the 3D-SE-ResNet achieved a high overall accuracy (92.9%) using only three images.

This challenges the prevailing paradigm that high temporal resolution is strictly necessary for crop classification. The Seasonal RF baseline struggled with the "curse of dimensionality" when stacking highly correlated bands, which diluted the temporal signal. However, the 3D deep learning model successfully compensated for the temporal sparsity by leveraging the extreme spectral depth of the EnMAP data. The Squeeze-and-Excitation mechanism played a pivotal role here, allowing the model to adaptively focus on the most relevant spectral features for each timestamp. This suggests that spectral resolution can effectively substitute for temporal frequency, enabling accurate seasonal mapping even in cloud-prone regions where continuous monitoring is not feasible.

4.4. Limitations and Future Directions

While the proposed DSS-2D and 3D-SE-ResNet models demonstrated robustness, the study was conducted on large, homogeneous agricultural parcels. The spatial smoothing effect of the 2D/3D convolutions is highly advantageous in such environments but may lead to boundary artifacts in fragmented landscapes with smallholder farming. Future research should assess the transferability of these architectures to heterogeneous agricultural systems.

A significant limitation of the current study stems from the experimental nature of the EnMAP mission. Unlike operational systems, EnMAP currently provides limited on-demand observation capacity, resulting in a dataset where critical phenological transitions in late spring and early summer were not captured. However, the fact that the seasonal model achieved high accuracy despite this temporal sparsity is highly encouraging. This serves as a strong validation for the upcoming CHIME mission [23], which is designed to provide similar hyperspectral specifications but with the systematic revisit frequency required for operational monitoring.

Additionally, the reliance on a single growing season limits the assessment of inter-annual generalization. Future work will focus on integrating transfer learning techniques to adapt the pre-trained EnMAP models to subsequent years, reducing the need for annual ground truth collection.

5. Conclusions

This study provided a comprehensive evaluation of EnMAP hyperspectral imagery for agricultural land cover mapping, demonstrating that deep spectral-spatial feature extraction can effectively overcome the limitations of data sparsity. Our findings lead to three primary conclusions regarding the operational implementation of spaceborne imaging spectroscopy.

 doi:10.20944/preprints202602.0202.v1

First, the integration of spatial context is a prerequisite for robust hyperspectral classification when training data is limited. The comparative analysis revealed that pixel-based methods, including 1D-CNN and RF, are susceptible to overfitting and spectral noise. In contrast, the proposed DSS-2D and 3D-SE-ResNet architectures successfully exploited texture and pattern recognition, utilizing geometrical augmentation to achieve high stability (OA > 97%) even with restricted ground truth. This confirms that treating hyperspectral data as a 3D volume rather than a 1D vector is critical for algorithmic convergence.

Second, the study validates the distinct biophysical value of the full 420–2450 nm spectral range. While standard multispectral sensors are sufficient for monitoring peak greenness, the SWIR domain proved indispensable for differentiating crops during critical phenological transitions. Specifically, the SWIR bands were essential for characterizing soil background properties in the early season and identifying senescence markers in the late season. This suggests that future agricultural monitoring systems must prioritize the inclusion of the 1000–2450 nm range to resolve spectral ambiguities between biologically similar crop types.

Finally, we demonstrated that high spectral resolution can compensate for low temporal frequency. Despite the sparse availability of EnMAP acquisitions, the 3D deep learning model successfully reconstructed the seasonal crop map with an accuracy of 92.9%. This finding has significant implications for the upcoming Copernicus CHIME mission. It serves as a proof-of-concept that once the systematic revisit capability of CHIME is combined with the deep learning architectures proposed herein, it will be feasible to deliver continuous, high-precision agricultural services on a continental scale, transcending the current limitations of experimental tasking missions.

Author Contributions: Conceptualization, L.M, Z.T. and J.S; methodology, L.M and M.S.; software, M.S.; validation, M.S, J.S.; formal analysis, L.M, M.S.; investigation, M.S, J.M and E.S.; resources, L.M and J.M; data curation, M.S, D.V and D.L; writing—original draft preparation, M.S.; writing—review and editing, L.M.; visualization, M.S., Z.T.; supervision, L.M.; project administration, L.M.; funding acquisition, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NKFIH-National Research and Innovation Office (NKFIH ADVANCED), grant no. 149686. Project title: Definition and Geospatial Big Data Analysis of Spectral Fingerprints of Healthy and Unhealthy Crops Using Fused Hyperspectral Multi-temporal Satellite and Field Remotely Sensed Data and Deep Learning Methods.

Data Availability Statement: The ENMAP images were acquired according to the accepted scientific proposal of Laszlo Mucsi (Proposal No. A00001-P00875). All used ENMAP images available for download at the EOC Geoservice are accessible using a unified Geoservice Account. The agricultural reference data are owned by Mezőhegyesi Ménesbirtok Zrt.

Acknowledgments: The authors would like to thank Mezőhegyesi Ménesbirtok Zrt. for providing the agricultural reference data used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

1D-CNN	1D Convolutional Neural Network
BN	Batch Normalization
DSS-2D	Dual-Stream Spatial-Spectral 2D CNN
EnMAP	Environmental Mapping and Analysis Program
FDR	False Discovery Rate
GAP	Global Average Pooling
LAI	Leaf Area Index
NIR	Near-Infrared

 doi:10.20944/preprints202602.0202.v1

OA	Overall Accuracy
PA	Producer's Accuracy
ReLU	Rectified Linear Unit
RF	Random Forest
SE	Squeeze-and-Excitation
SWIR	Shortwave Infrared
UA	User's Accuracy

References

1. M. Belgiu and O. Csillik, 'Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis', *Remote Sens.*, vol. 10, no. 1, p. 125, 2018, doi: 10.3390/rs10010125.
2. M. Buchhorn, M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, 'Copernicus global land cover layers-collection 2', *Remote Sens.*, vol. 12, no. 6, 2020, doi: 10.3390/rs12061044.
3. D. K. Bolton, J. M. Gray, E. K. Melaas, M. Moon, L. Eklundh, and M. A. Friedl, 'Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery', *Remote Sens. Environ.*, vol. 240, 2020, doi: 10.1016/j.rse.2020.111685.
4. L. Blickensdörfer, M. Schwieder, D. Pflugmacher, C. Nendel, S. Erasmi, and P. Hostert, 'Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany', *Remote Sens. Environ.*, vol. 269, 2022, doi: 10.1016/j.rse.2021.112831.
5. J. G. P. W. Clevers and A. A. Gitelson, 'Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3', *Int. J. Appl. Earth Obs. Geoinformation*, vol. 23, pp. 344–351, Aug. 2013, doi: 10.1016/j.jag.2012.10.008.
6. L. Homolová, Z. Malenovský, J. G. P. W. Clevers, G. García-Santos, and M. E. Schaepman, 'Review of optical-based remote sensing for plant trait mapping', *Ecol. Complex.*, vol. 15, pp. 1–16, Sep. 2013, doi: 10.1016/j.ecocom.2013.06.003.
7. D. Haboudane, J. Miller, N. Tremblay, P. Zarco-Tejada, and L. Dextraze, 'Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture', *Remote Sens. Environ.*, vol. 81, pp. 416–426, Aug. 2002, doi: 10.1016/S0034-4257(02)00018-4.
8. Y. Guo et al., 'Comparison of different machine learning algorithms for predicting maize grain yield using UAV-based hyperspectral images', *Int. J. Appl. Earth Obs. Geoinformation*, vol. 124, p. 103528, Nov. 2023, doi: 10.1016/j.jag.2023.103528.
9. P. Singh, P. K. Srivastava, P. K. Jha, J. Verrelst, P. N. Singh, and R. Prasad, 'Retrieval of crop traits using PROSAIL-based hybrid radiative transfer model and EnMAP hyperspectral data', *Int. J. Appl. Earth Obs. Geoinformation*, vol. 143, p. 104769, Sep. 2025, doi: 10.1016/j.jag.2025.104769.
10. L. Guanter, H. Kaufmann, and K. Segl, 'The EnMAP spaceborne imaging spectroscopy mission for Earth observation', *Remote Sens.*, vol. 7, pp. 8830–8857, 2015, doi: 10.3390/rs70708830.
11. U. Heiden et al., *The user interface of the EnMAP satellite mission*. 2010. doi: 10.1109/IGARSS.2010.5649217.
12. L. Mucsi et al., 'Assessment of the Effectiveness of Spectral Indices Derived from EnMAP Hyperspectral Imageries Using Machine Learning and Deep Learning Models for Winter Wheat Yield Prediction', *Remote Sens.*, vol. 17, no. 20, p. 3426, Oct. 2025, doi: 10.3390/rs17203426.
13. S. Cooper, A. Okujeni, D. Pflugmacher, S. van der Linden, and P. Hostert, 'Combining simulated hyperspectral EnMAP and Landsat time series for forest aboveground biomass mapping', *Int. J. Appl. Earth Obs. Geoinformation*, vol. 98, 2021, doi: 10.1016/j.jag.2021.102307.
14. M. Immitzer, C. Atzberger, and T. Koukal, 'Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data', *Remote Sens.*, vol. 4, no. 9, pp. 2661–2693, Sep. 2012, doi: 10.3390/rs4092661.
15. Z. Chen, Y. Hao, Q. Liu, Y. Liu, M. Zhu, and L. Xiao, 'Deep Learning for Hyperspectral Image Classification: A Critical Evaluation via Mutation Testing', *Remote Sens.*, vol. 16, no. 24, p. 4695, Dec. 2024, doi: 10.3390/rs16244695.

 doi:10.20944/preprints202602.0202.v1

16. D. R. Roberts et al., 'Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure', *Ecography*, vol. 40, no. 8, pp. 913–929, Aug. 2017, doi: 10.1111/ecog.02881.
17. P. Ploton et al., 'Spatial validation reveals poor predictive performance of large-scale ecological mapping models', *Nat. Commun.*, vol. 11, no. 1, p. 4540, Sep. 2020, doi: 10.1038/s41467-020-18321-y.
18. S. Chabrillat et al., 'The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch', *Remote Sens. Environ.*, vol. 315, p. 114379, Dec. 2024, doi: 10.1016/j.rse.2024.114379.
19. U. Meier, 'Growth stages of mono- and dicotyledonous plants: BBCH Monograph', 2018, doi: 10.5073/20180906-074619.
20. L. Breiman, 'Random Forests', *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
21. W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, 'Deep Convolutional Neural Networks for Hyperspectral Image Classification', *J. Sens.*, vol. 2015, pp. 1–12, 2015, doi: 10.1155/2015/258619.
22. H. Firat, M. E. Asker, M. Bayindir, and D. Hanbay, '3D residual spatial–spectral convolution network for hyperspectral remote sensing image classification', *Neural Comput. Appl.*, vol. 35, pp. 1–19, Oct. 2022, doi: 10.1007/s00521-022-07933-8.
23. J. Nieke et al., The copernicus hyperspectral imaging mission for the environment (CHIME): an overview of its mission, system and planning status. 2023. doi: 10.1117/12.2679977.