agriculture

MDPI

Article

One-Versus-Rest Crop Classification from Multi-Temporal Sentinel-2 Imagery Using a Hybrid CNN–LSTM Architecture and Explainable GeoAI

Vasile Adrian Nan [1], Gheorghe Badea [1,*], Anca Patricia Grădinaru [1], Ana Cornelia Badea [1], Mihai Valentin Herbei [2] and Lucian Octavian Dragomir [2]

[1] Faculty of Geodesy, Technical University of Civil Engineering, 020396 Bucharest, Romania; vasile-adrian.nan@phd.utcb.ro (V.A.N.); anca-patricia.gradinaru@utcb.ro (A.P.G.); ana.badea@utcb.ro (A.C.B.)
[2] Faculty of Agriculture, University of Life Sciences "King Mihai I" from Timisoara, 300645 Timișoara, Romania; mihai_herbei@yahoo.com (M.V.H.); gammaz1973@gmail.com (L.O.D.)
* Correspondence: gheorghe.badea@utcb.ro

Abstract

Accurate crop classification from satellite imagery can substantially enhance agricultural monitoring. Recent advances in deep learning have improved crop classification by employing models that take into account both spatial and temporal information contained in satellite image time series. However, the limited interpretability of these models remains a major challenge for their adoption in agricultural applications. This study presents a one-versus-rest classification framework for five crop types (wheat, maize, rapeseed, sunflower and watermelon and melon), in Ialomița County, Romania, using a CNN-LSTM (Convolutional Neural Network–Long Short-Term Memory) architecture and nine Sentinel-2 images acquired throughout the 2025 growing season. Each image sequence consisted of four spectral bands and NDVI (Normalized Difference Vegetation Index), enabling the exploitation of spectral, spatial and temporal information. Model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrices and the area under the receiver operating characteristic curve (ROC-AUC). The models achieved high validation accuracy, with values of 96.01% for wheat, 94.33% for maize, 97.48% for rapeseed, 97.06% for sunflower and 97.06% for watermelon and melon classifications. Explainable AI (XAI) techniques such as Integrated Gradients and SHAP (SHapley Additive exPlanations) were employed to quantify the contribution of individual spectral bands and acquisition dates to the classification decision. The Integrated Gradients analysis showed that for some crops, NDVI was the most influential input feature, accounting for 50.78% of the feature importance in wheat classification, 39.30% in rapeseed classification and 40.01% in watermelon and melon classification, while for other crops, the NIR (Near-infrared) spectral band was the most influential feature, accounting for 29.89% of the feature importance in maize classification and 41.00% in sunflower classification. The results demonstrate that the proposed framework achieves high classification performance while providing meaningful explanations of model predictions, highlighting the importance of both vegetation characteristics and seasonal dynamics in crop classification.

Keywords: crop classification; explainable GeoAI; time series; one-versus-rest classification

Check for updates

Academic Editor: Xiuliang Jin

Received: 7 September 2026
Revised: 26 October 2026
Accepted: 3 November 2026
Published: 8 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Smart farming combines both agricultural science and information and communication technology to enhance farm management, by integrating emerging technologies such as IoT (Internet of Things), Big Data and AI (Artificial Intelligence) [1]. Data analysis for sustainable agriculture is facilitated through a geospatial framework aimed at improving agricultural practices and resource efficiency [2] and detailed crop classification maps play an important role in agricultural monitoring with satellite remote sensing being widely used in this regard [3]. Deep learning (DL) classification algorithms have been developed lately to improve the classification efficiency [4] and multispectral data such as Sentinel-2, Sentinel-3, Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) are commonly used as a base for crop type classification [5]. Although many crop type maps are being produced, the training datasets and methodologies are frequently unavailable [6], limiting transparency and the broader adoption of advanced crop classification approaches. This challenge can be addressed through the use of Explainable Artificial Intelligence (XAI) [7], a method used to mitigate the "black-box" issue of most of the AI models [8] and make AI decisions transparent, understandable and justifiable [9].

The EU (European Union) AI Act is part of a wider package of policy measures meant to support the development of reliable and trustworthy AI systems, guaranteeing safety and strengthening innovation in AI across the EU [10]. It defines four levels of risk for AI systems: minimal risk, limited risk, high risk and unacceptable risk. Crop classification and yield prediction are generally not considered high-risk applications under the AI Act. However, when AI-based crop mapping or monitoring tools are integrated into the safety components of autonomous agricultural machinery, they may fall within the scope of high-risk AI systems. In such scenarios, the ability to understand and justify model predictions becomes increasingly important.

At EU level, the Common Agricultural Policy (CAP) 2023–2027 is meant to support farmers in adopting practices that can lead to more sustainable farming models, with one key indicator being the digitalization of agriculture [11]. In Romania, the CAP strategic plan aims at increasing the economic viability of farms, reducing income disparities between farms and encouraging the competitiveness of the agricultural sector [12]. Romania's agricultural sector is characterized by a highly fragmented farm structure, consisting of a large number of small family farms alongside a relatively small number of large-scale agricultural enterprises. Most small farms cultivate limited land areas primarily for household consumption. With approximately 3.5 million farmers, Romania has the largest farming population in the European Union. However, around 90% of these small farms cover less than 5 hectares each, highlighting the fragmented nature of the country's agricultural landscape. Despite this structural fragmentation, Romania is one of the European Union's leading producers of cereals and ranks as the largest producer of sunflower seeds, honey, and plums [13].

In Romania, the Agency for Payments and Intervention in Agriculture (APIA) maintains a comprehensive database of agricultural parcels and crop declarations submitted annually by farmers for agricultural support applications [14]. As this database is primarily based on annual crop declarations provided by the farmers, the reported information requires continuous verification and monitoring. Combined with the fragmented structure of the Romanian agricultural sector, this creates a need for reliable AI models capable of accurately identifying cop types from remote sensing imagery. Furthermore, the distribution of crop types in Romania is imbalanced, with some crops occupying substantially larger cultivated areas than others. As pictured in Figure 1, the south-eastern region of the country is highly dominated by wheat and maize crops [15]. This imbalance is reflected in the available ground truth data, resulting in uneven class distributions within the training dataset. Such class imbalance can bias conventional multi-class classification

models, reducing their ability to accurately identify less represented crops. Therefore, the aim of this research is to propose a novel one-versus-rest classification method, with five models trained to detect wheat, maize, rapeseed, sunflower and watermelons and melons crops in Ialomița, a county in the south-east region of Romania. Unlike a conventional multi-class CNN, which distinguishes among all crop classes within a single model, the proposed approach trains a separate model for each target crop, distinguishing it from the remaining classes. This approach enables each model to focus on learning discriminative spectral and temporal characteristics of a specific crop. The independent structure of the models provides flexibility for model optimization as individual models can be retrained independently when new data become available or when updates are required for specific crop types. This can reduce the training effort required for updates and facilitates the models' adaptation to changing agricultural conditions.

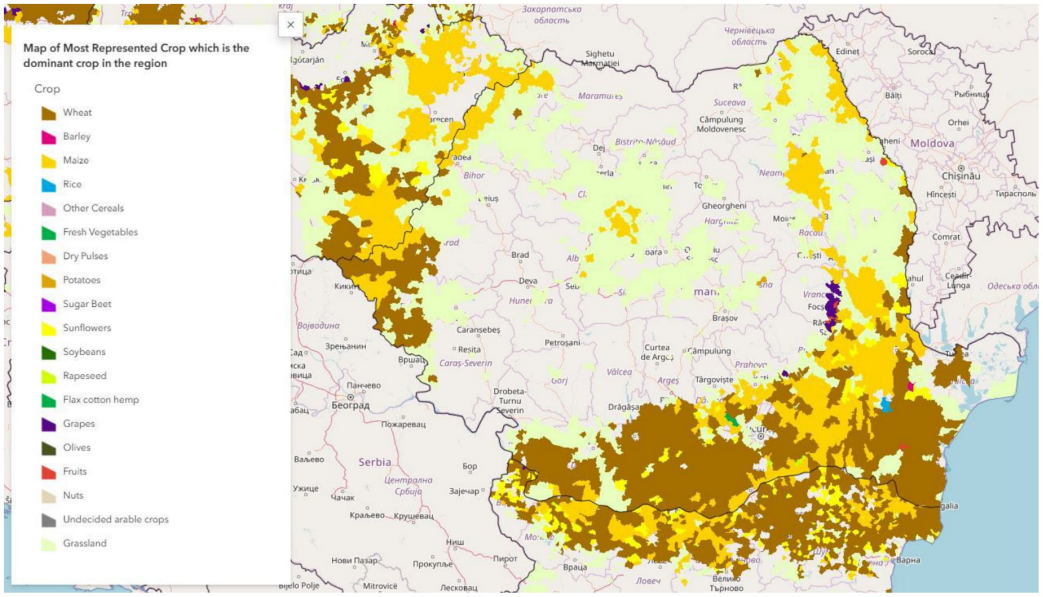


Figure 1. Dominant crops in Romania [15].

2. Literature Review

While numerous studies have investigated the application of AI and remote sensing in agriculture, further research is needed to explore their potential for accurate and automated crop type classification. Accurate crop classification is particularly useful in agricultural monitoring, as it provides information on crop distribution, validates the farmers' declarations and facilitates yield estimation and resource management, contributing to sustainable agricultural planning. AI methods are frequently applied for crop disease recognition based on remote sensing images [16–18] or crop anomaly detection [19]. Other studies regarding AI applications in smart farming include automated variable-rate fertilizer applicators [20], weed detection [21], detecting crop lines and accurately guiding weeding tools [22], crop yield estimation [23] and drought prediction [24].

Several methods have been proposed to solve one important issue of agricultural remote sensing—classifying crop types, through AI classification algorithms such as RF (Random Forest) [25], DNNs (Deep Neural Networks) [4], CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks), MLP (multi-layer perceptron) and transformers [26], with RF being the most used algorithm to solve this task [27]. Selecting the appropriate evaluation metrics for these models plays an important role in accurately assessing model performance and should be guided by the specific objectives and characteristics of the classification task. Argenton et al. proposed a PSE-CNN (Pixel-Set Encoder Convolutional

Neural Network) framework for citrus crop classification, achieving a unified F1-score of 0.704 [28]. A UNet++ architecture was implemented by Yin et al. [29] for winter rapeseed mapping, achieving an overall accuracy of 98.65%. Time series analysis is essential in crop mapping as the seasonal dynamics in crops influence their spectral signatures throughout the growing season, enabling more accurate discrimination between crop types. Some applications rely on multi-temporal data across a range of dates or a pair of dates [30]. Gallo et al. [31] performed a time series analysis for crop mapping using a CNN and data obtained from two growing seasons, achieving a 93% overall accuracy.

The performance of AI models is also influenced by the quality, diversity and representativeness of the training data. Satellite imagery such as Sentinel-2 and Landsat-8/9 [25] can often provide the right spatial resolution for this type of classification tasks, although UAV (Unmanned Aerial Vehicle) imagery is also found in the literature, having the advantage of better spatial resolution and not being limited by cloud coverage [32]. The reliability of ground truth data is also important as it directly influences both model development and evaluation. Collecting accurate field observation is often time-consuming and costly, limiting the availability of large representative datasets for training and validating deep learning models.

Although some studies address the classification issue, there is little research regarding the explainability of these models and challenges related to the uncertainty and data heterogeneity remain major obstacles [7]. Through Explainable AI, tools that support interpreting the results of DL models are created and this approach should be further implemented in the agricultural sector [33]. Some XAI tools are found in the literature in models trained to predict crop yields [34,35] or detect crop disease [34,36,37]. Integrated Gradients (IGs) is a common gradient-based feature attribution method used specifically for image data [38] which quantifies the contribution of each input feature. Other XAI techniques include SHAP (SHapley Additive exPlanations), which provides local and global explanations, or LIME (Local Interpretable Model-agnostic Explanations), that aims at explaining how the model works locally for a specific instance of the model [39].

Existing studies have demonstrated the effectiveness of CNN-LSTM (Convolutional Neural Network–Long Short-Term Memory) or related architectures for crop-type classification using multi-temporal satellite data. However, these studies generally formulate crop mapping as a conventional multiclass problem. This study splits the multiclass problem into crop-specific one-vs-rest binary classification tasks and separate CNN-LSTM models are trained for each crop. This workflow has the potential to facilitate the future inclusion of additional crop types without requiring the complete retraining of the classification model. Furthermore, XAI is incorporated to provide an interpretation of the model predictions for each crop. Table 1 provides a comparison of representative studies addressing crop-type classification using multi-temporal Sentinel-2 data.

Li et al. [40] evaluated five DL architectures using multi-temporal Sentinel-2 data and found that 3D-CNN and ConvLSTM2D achieved the highest crop-classification accuracies by integrating temporal, spectral and spatial information. However, the study used a balanced multiclass dataset and therefore did not address the class-imbalance challenge, nor did it investigate model explainability. Zhao et al. [41] evaluated five deep learning architectures for multi-class crop-type mapping using Sentinel-2 time-series data with missing observations. Although the dataset exhibited substantial differences in sample size among crop types, the study retained a multiclass framework and did not employ a one-versus-rest strategy or XAI. Tufail et al. [42] developed four hybrid architectures using multi-temporal Sentinel-2 spectral bands and vegetation indices for multiclass crop classification. This study did not address the one-vs.-rest classification strategy and the explainability of the models.

Table 1. Related studies.

Study	Sentinel-2 Time Series	CNN-LSTM/ConvLSTM	One-Versus-Rest	Class Imbalance	XAI
Deep Learning Application for Crop Classification via Multi-Temporal Remote Sensing Images	✓	✓	X	X	X
Evaluation of Five Deep Learning Models for Crop Type Mapping Using Sentinel-2 Time Series Images with Missing Information	✓	✓	X	✓	X
Deep Learning Applications for Crop Mapping Using Multi-Temporal Sentinel-2 Data and Red-Edge Vegetation Indices: Integrating Convolutional and Recurrent Neural Networks	✓	✓	X	✓	X
XAI for Early Crop Classification	✓	X	X	✓	✓
This study	✓	✓	✓	✓	✓

While XAI has also been applied in remote-sensing-based crop classification to provide insights into the temporal and spectral factors influencing model predictions, previous studies have predominantly employed other ML (Machine Learning) architectures rather than a crop-specific CNN-LSTM framework. Chan et al. [43] applied Layer-wise Relevance Propagation (LRP) to an attention-based Transformer model for early crop classification using highly imbalanced Sentinel-2 time-series data, with XAI employed to identify the most relevant temporal observations. In contrast, this study employs crop-specific CNN-LSTM models within a one-vs-rest framework, while XAI, through both Integrated Gradients and SHAP methods, is used to interpret both spectral and temporal patterns contributing to individual crop predictions.

Overall, recent advances in Artificial Intelligence and remote sensing have significantly improved crop classification accuracy. However, most existing studies primarily focus on maximizing classification accuracy, while little attention has been given to the interpretability of these models and the transparency of their decision-making processes. Furthermore, challenges related to data uncertainty and the limited availability of reliable ground truth data continue to impact the development of crop classification frameworks. This study addresses these gaps by combining multi-temporal Sentinel-2 imagery with a deep learning architecture, based on ground truth training data provided by official authorities, as well as Explainable AI techniques.

3. Materials and Methods

3.1. Framework for Crop Classification in Ialomița County

The proposed framework is designed to support crop-type identification and agricultural monitoring using multi-temporal satellite imagery in Ialomița county, Romania, an area with an uneven distribution of crops. This framework addresses a classification problem characterized by a severe imbalance in the distribution of crop classes, integrating a one-versus-rest classification strategy that employs CNN-LSTM modeling and class

weighting. The study focuses on five representative crop categories in Ialomița county, Romania: wheat, maize, rapeseed, sunflower and watermelon and melon. These crops were selected to represent both dominant agricultural crops such as maize and wheat and less represented crop classes, such as sunflower, rapeseed and watermelon and melon. XAI is used to provide an interpretable assessment of the model predictions by identifying the spectral and temporal information contributing to the classification of each crop. The data resources required for this framework are represented by multi-temporal satellite imagery and corresponding ground-truth information. The methodological resources include a programming environment required for data processing, model development and evaluation and explainability analysis. The methodological approach consists of a workflow that includes all the steps necessary for model training and evaluation and the resulting outputs are crop-specific classification results, quantitative assessments of the model performance and interpretable information regarding the spectral and temporal characteristics influencing the classification decisions. The framework therefore integrates the consideration of class imbalance, classification performance and model interpretability within a unified methodological approach. The resulting framework (Figure 2) can provide an approach that can support broader agricultural applications, while the generated crop information can serve as a basis for future agricultural decision-support applications.

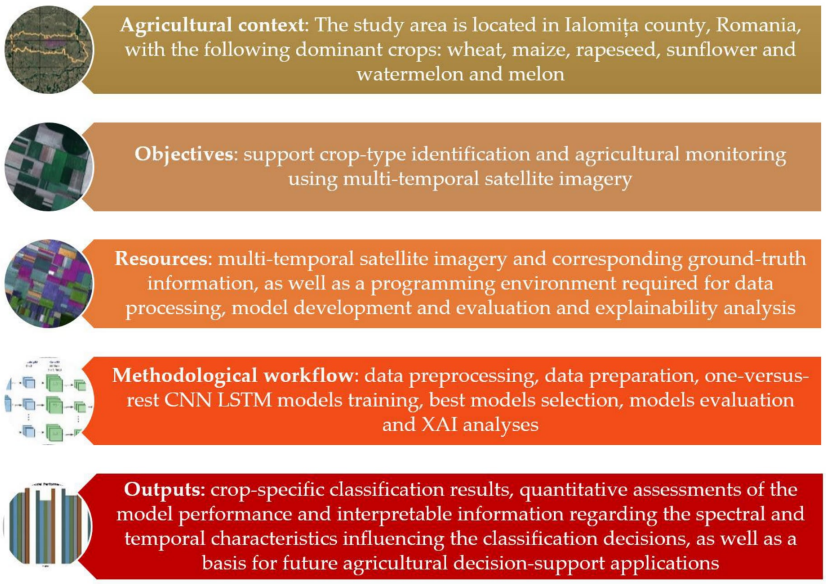


Figure 2. Framework for crop classification in Ialomița county.

3.2. Workflow of the Proposed One-Versus-Rest Explainable GeoAI Framework for Crop Classification

The implemented workflow (Figure 3) began with data acquisition, represented by multi-temporal Sentinel-2 imagery and ground truth data. The next step consisted of data pre-processing, which involved unifying the training data into a common coordinate reference system, parcel extraction, 64 × 64 image patch extraction, band normalization and NDVI computation. Data preparation included splitting the dataset into training and validation subsets using an 80:20 ratio. The 80:20 ratio was chosen as a practical balance between having enough samples for model training and a sufficiently large validation dataset for reliable performance assessment. The resulting split provided 3804 parcels for training and 952 parcels for validation. Class weighting was subsequently applied to reduce the effects of class imbalance. The CNN-LSTM model was trained using binary cross-entropy as the loss function, the Adam optimizer and call-backs such as EarlyStop-

ping, ModelCheckpoint and ReduceLROnPlateau. The best-performing model was selected based on the minimum validation loss and evaluated using accuracy, precision, recall, F1-score, ROC-AUC, Balanced Accuracy, macro F1-score and confusion matrices. Finally, the workflow concluded with Explainable GeoAI (Geospatial Artificial Intelligence) analyses, based on the Integrated Gradients and SHAP methods, to interpret temporal and spectral importance of the input data. Throughout the workflow, QGIS (3.38.3 version) was used for data pre-processing and visualization and Python (3.10.20 version), with TensorFlow (2.10.0 version), NumPy (1.26.4 version), Scikit-learn (1.7.2 version), Ras-terio (1.4.3 version), GeoPandas (1.0.1 version) and Matplotlib (3.10.9 version) libraries, was used for data pre-processing, data preparation, CNN-LSTM model training, best model selection, model evaluation and explainable GeoAI analyses. The individual stages of the workflow are described in detail in the following sections.

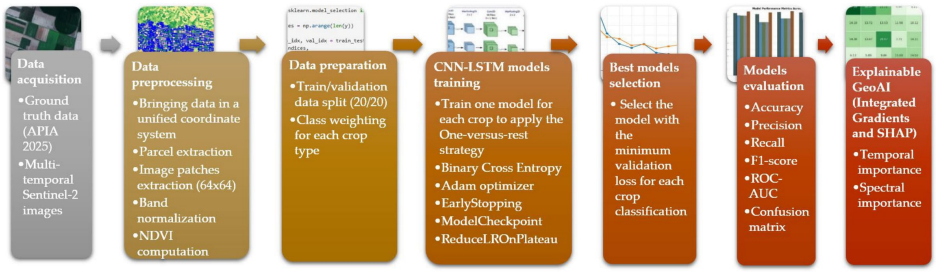


Figure 3. Workflow for the proposed One-Versus-Rest Explainable GeoAI framework for crop classification.

3.3. Study Area

The study was conducted in Ialomița County (Figure 4), located in the southeastern part of Romania, one of the country's most productive agricultural regions. The county is characterized by extensive arable land and flat terrain, providing favorable conditions for large-scale crop cultivation. Ialomița County has a total area of 4453 km^2, of which 3736 km^2 are agricultural land. Of the total cultivated area, cereals account for 66.8%, with maize representing 46.1% of the cereal cultivation area and wheat approximately 43.7% [44].

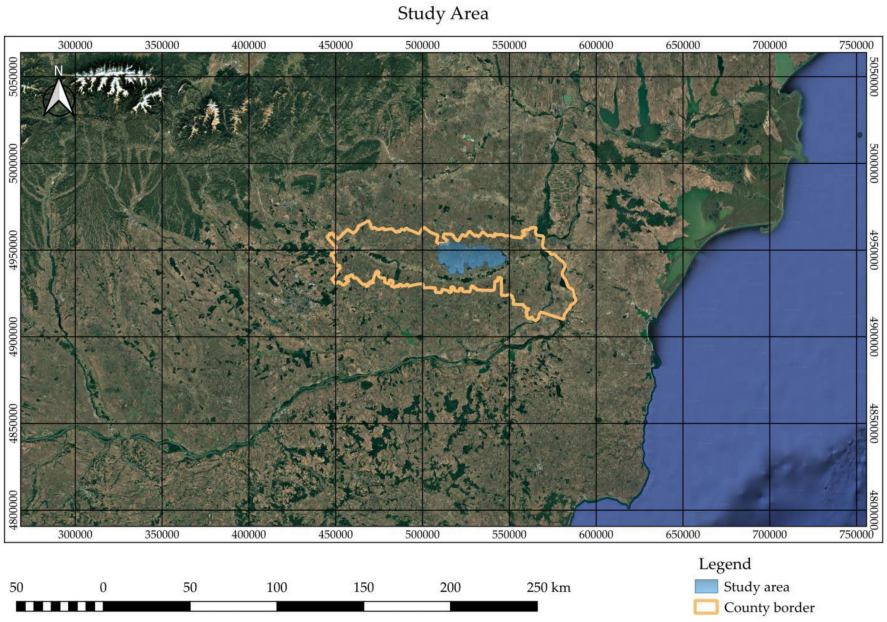


Figure 4. Geographic location of the study area.

3.4. Data Acquisition

The dataset consisted of nine Level-2A Sentinel-2 image acquisitions covering the 2025 growing season, acquired on 29 April, 4 May, 8 June, 23 June, 3 July, 13 July, 25 July, 12 August and 27 August 2025. The Level-2A Sentinel-2 products provide atmospherically corrected surface reflectance data and they were selected to ensure minimal cloud coverage over the study area. The cloud coverage associated with each acquisition is presented in Table 2.

Table 2. Cloud coverage for each acquisition date.

Image No.	Acquisition Date	Cloud Coverage (%)
1	29 April 2025	0.070646
2	4 May 2025	0.007407
3	8 June 2025	0.005877
4	23 June 2025	2.549630
5	3 July 2025	0.001855
6	13 July 2025	0.106167
7	25 July 2025	0.000000
8	12 August 2025	0.001736
9	27 August 2025	0.120421

The ground truth dataset consisted of a shapefile provided by APIA, containing 4756 agricultural parcels and their corresponding crop codes based on farmers' annual declarations. The dataset covered a total area of 4763 ha. Five major crop types were selected for this study (Figure 5). An important characteristic of the dataset is the unequal distribution of the target crop classes. The number of parcels varied among the five crop types, ranging from 344 parcels for watermelon and melon to 1207 parcels for wheat. The one-versus-rest strategy, with a binary CNN-LSTM model trained for each crop was used to reduce the influence of this imbalance. This allowed the contribution of each class to be considered independently. In imbalanced classification problems, several strategies such as Weighted Binary Cross-Entropy [45], Focal Loss [46] or Contrastive Margin [47] can be incorporated to reduce the influence of the majority class. For this study, Binary Cross-Entropy was used as the loss function to penalize deviations between the predicted probabilities and the corresponding true class labels, while class weighting with weights adjusted for each crop was applied during training to mitigate the influence of class imbalance. To ensure a robust evaluation despite the imbalance, model performance was monitored using precision, recall and the F1-score, alongside accuracy.

The remaining parcels corresponded to crop types with a limited number of samples and were therefore excluded from the analysis, as they were insufficiently represented to support accurate model training and evaluation. A qualitative example of the training dataset is shown in Figure 6.

3.5. Data Preparation

The spatial data layers which included the Sentinel-2 images and the polygon parcels representing the underlying training samples were brought into a unified coordinate system (WGS84) to ensure accurate spatial alignment and reliable feature extraction.

For each agricultural parcel, a spatial image patch of 64 × 64 pixels containing 5 spectral bands was extracted from each of the 9 acquisitions. The resulting input therefore represented a multi-temporal and multi-spectral image sequence with dimensions of 9 × 64 × 64 × 5 pixels.

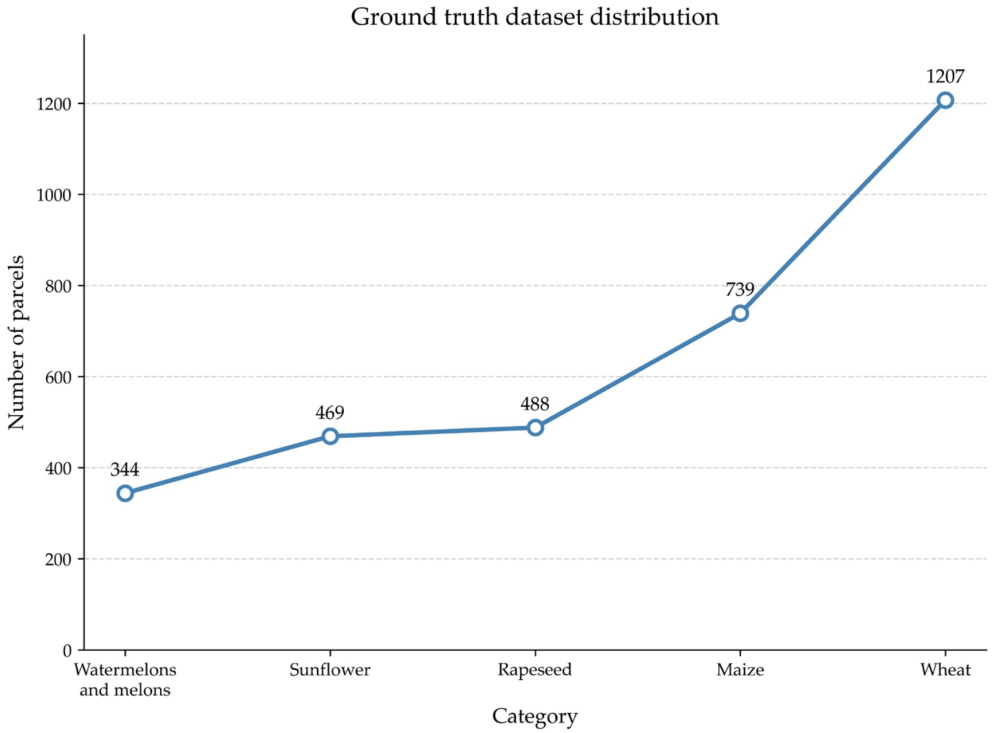


Figure 5. Ground truth dataset distribution.

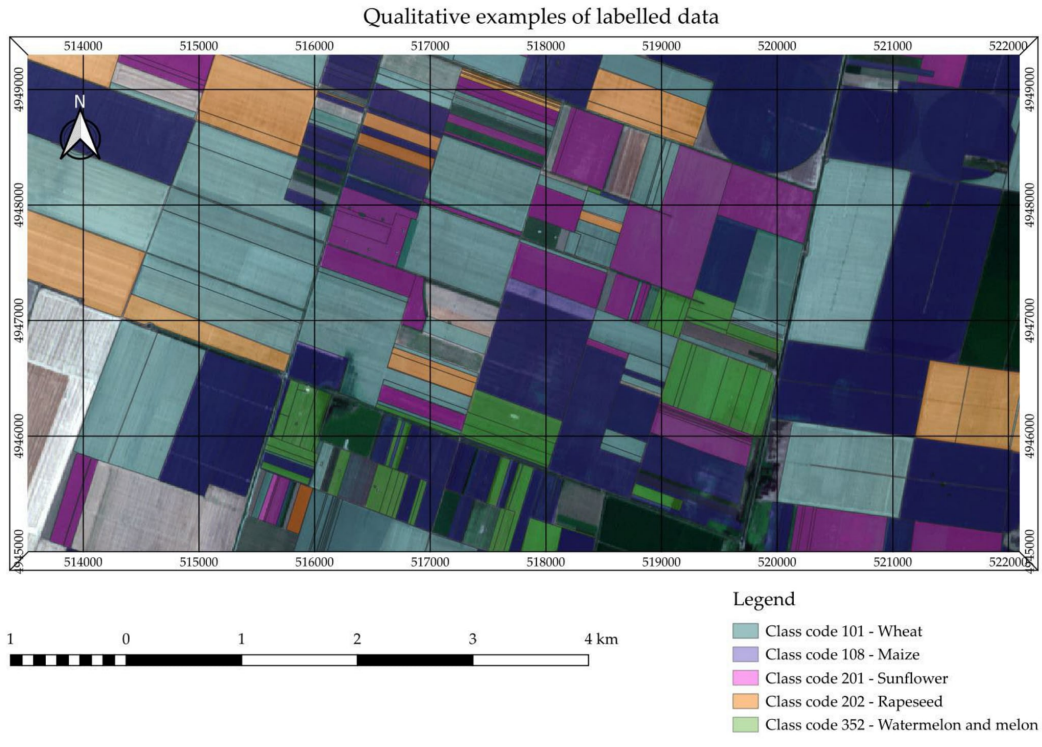


Figure 6. Qualitative examples of labeled data.

Five input features were used for each pixel: four Sentinel-2 spectral bands—red, green, blue and NIR (Near-infrared) and NDVI (Normalized Difference Vegetation Index). The four spectral bands were converted from the original reflectance scaling to decimal reflectance values by dividing the original values by 10,000. NDVI was used as an additional vegetation-related feature.

The data were divided into training and validation subsets using an 80:20 stratified split, ensuring that the class proportions were maintained in both subsets. The split was performed at parcel level and each agricultural parcel was constructed as a single sample containing the complete temporal sequence of Sentinel-2 observations associated with that parcel. Therefore, each parcel was assigned exclusively to the training or validation subset. The split was performed randomly, with a fixed random seed (random_state = 42) used for reproducibility. The number of samples from each class and set is summarized in Table 3.

Table 3. Number of samples from each class and dataset.

Class	Total Number of Samples	Training Samples	Validation Samples
Wheat	1207	965	242
Maize	739	591	148
Rapeseed	488	390	98
Sunflower	469	375	94
Watermelon and melon	344	275	69

3.6. CNN-LSTM Architecture

A Convolutional Neural Network combined with a Long Short-Term Memory network was used to exploit both the spatial characteristics of the image patches and their temporal evolution (Figure 7). The model received an input tensor of dimensions $9 \times 64 \times 64 \times 5$. A sequence of convolutional operations was applied independently to each temporal image using TimeDistributed layers. The first convolutional layer contained 32 filters with a 3×3 kernel and ReLU activation, followed by 2×2 max-pooling. A second convolutional layer contained 64 filters with a 3×3 kernel and was followed by another 2×2 max-pooling operation. The resulting spatial features were flattened independently for each acquisition date and passed to an LSTM layer containing 64 units. The LSTM layer was followed by a dropout layer with a dropout rate of 0.3 and a final fully connected layer with one neuron and sigmoid activation. The CNN-LSTM architecture and associated hyperparameters were selected using a literature-informed approach. The initial architecture was established based on the suitability of CNNs for extracting spatial features from images and LSTM networks for modeling temporal dependencies in multi-temporal satellite observations. The main hyperparameters described in the following section were initially selected based on commonly used configurations reported in the literature for deep learning-based remote-sensing classification. During model development, adjustments were made to the training configuration based on learning behavior. The maximum number of training epochs and the patience setting for early stopping were adjusted based on the observed model performance during training and ReduceLROnPlateau was employed to reduce the learning rate when improvements plateaued. However, a systematic hyperparameter search was not conducted and the final configuration was not established through a comparison of alternative combinations.

3.7. Model Training

The models were trained using the Adam optimizer with an initial learning rate of 0.001 and binary cross-entropy as the loss function. ReduceLROnPlateau monitored validation loss and reduced the learning rate by a factor of 0.5 after five epochs without improvement, with a minimum learning rate of 1×10^{-6}. Accuracy, precision and recall were monitored during training. Training was performed for a maximum of 50 epochs with a batch size of 16. Early stopping was implemented by monitoring validation loss, using a patience of 15 epochs and restoring the model weights corresponding to the lowest validation loss. This

procedure was used to evaluate the reliability of the model's predictions, limit overfitting and retain the model state that achieved the best validation performance. Figure 8 shows a pseudocode description of the algorithm.

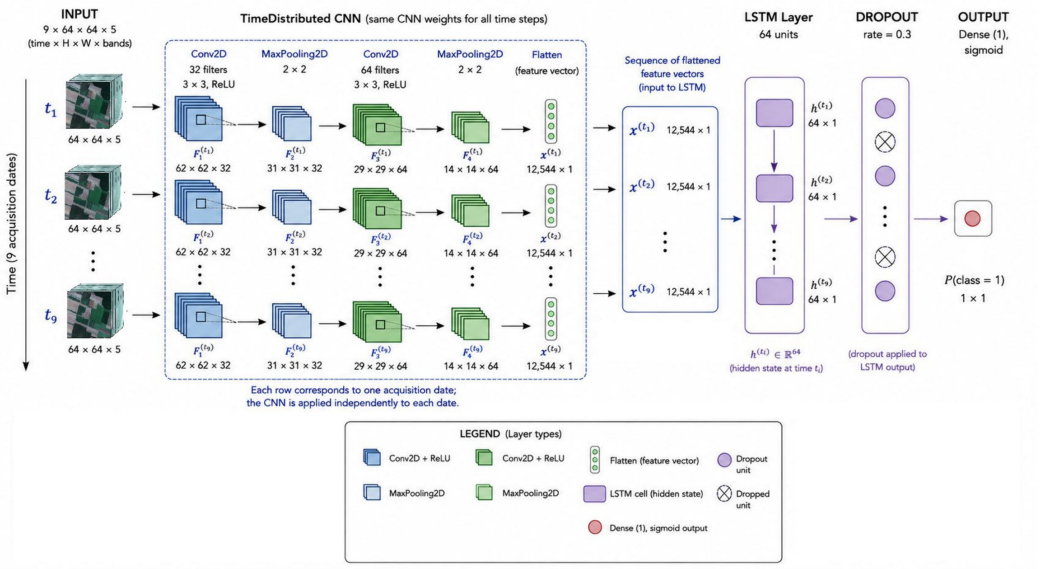


Figure 7. Schematic representation of the proposedCNN-LSTM architecture.

```
DEFINE callbacks:

    EarlyStopping:
        monitor = validation loss
        patience = 15 epochs
        restore weights from the best epoch

    ModelCheckpoint:
        monitor = validation loss
        save model only when validation loss improves

    ReduceLROnPlateau:
        monitor = validation loss
        reduce learning rate by factor = 0.5
        patience = 5 epochs
        minimum learning rate = 1 x 10⁻⁶

EXECUTE model training:

    TRAIN model using:
        training features = X_train
        training labels = y_train

    VALIDATE model using:
        validation data = (X_val, y_val)

    SET training parameters:
        maximum epochs = 50
        batch size = 16
        class weights = training_class_weights

    APPLY:
        callbacks = training_callbacks

    SAVE:
        training history = model_history
```

Figure 8. Pseudocode description of the algorithm.

3.7.1. Wheat

The best validation loss observed during the training was 0.1304. The wheat classification model achieved strong validation performance, with an accuracy of 96.01%, precision of 92.50%, recall of 91.74% and an F1-score of 92.12% at the epoch corresponding to the minimum validation loss. The model also achieved a ROC-AUC of 0.9834, indicating a high discriminative capacity between wheat and non-wheat parcels (Figure 9a).

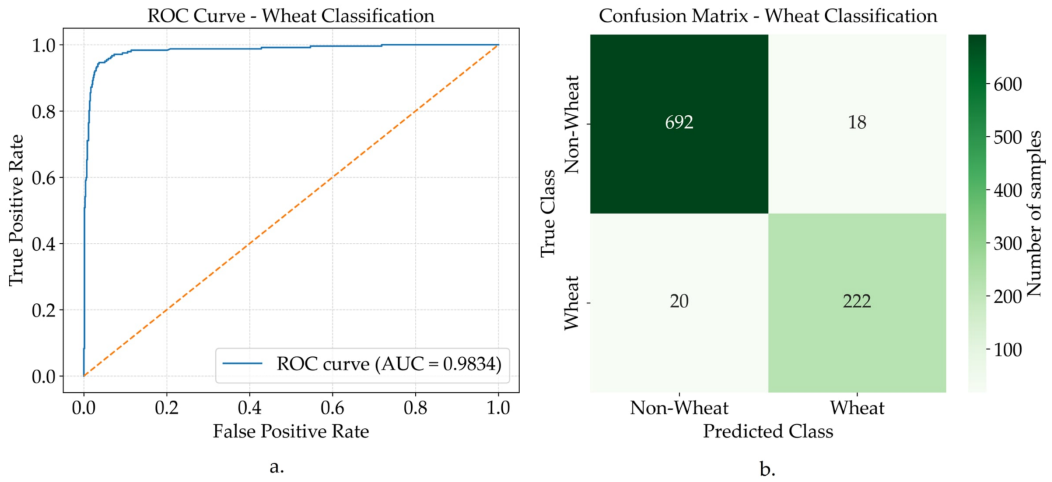


Figure 9. Wheat classification results: (**a**) ROC curve for the wheat classification model; (**b**) wheat classification confusion matrix.

The corresponding model weights were retained through the early-stopping procedure. The resulting confusion matrix (Figure 9b) contained 692 correctly classified non-wheat parcels, 18 false-positive wheat predictions, 20 false-negative wheat predictions and 222 correctly classified wheat parcels. The slightly lower number of false-positive than false-negative predictions indicates that the model was more conservative when predicting the wheat class, contributing to its high precision and F1-score. The high number of correctly classified wheat parcels suggests that the large number of training samples enabled the model to learn spectral and temporal characteristics associated with wheat.

3.7.2. Maize

To account for the difference in representation between maize and the other crops, class weighting was applied during training. The negative class was assigned a weight of 1.0, while the positive maize class was assigned a weight of 1.3. The best validation loss was 0.1754. The model achieved a validation accuracy of 94.33%, validation precision of 81.76%, validation recall of 81.76% and a F1-score of 81.76%. The model achieved a ROC-AUC of 0.9697 (Figure 10a), indicating a high discriminative capacity for distinguishing maize parcels from other agricultural crops.

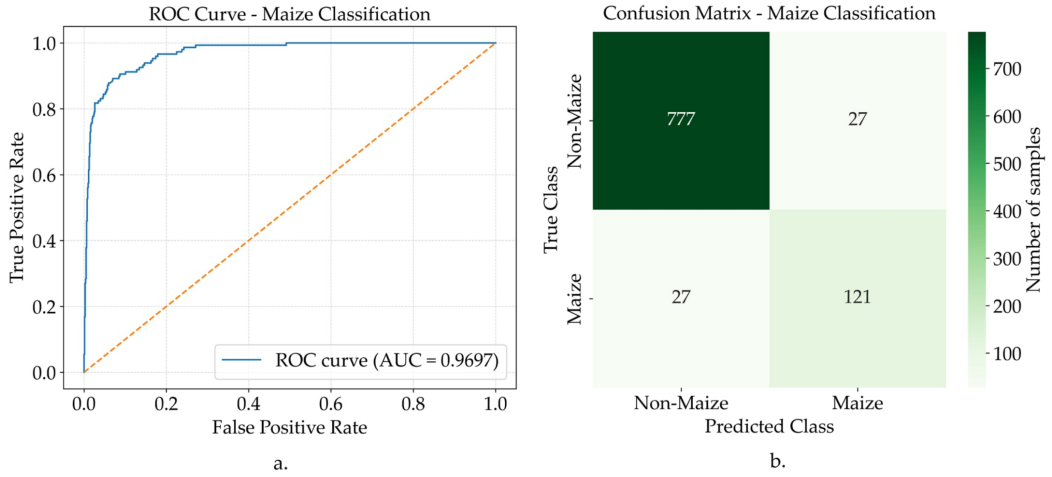


Figure 10. Maize classification results: (**a**) ROC curve for the maize classification model; (**b**) maize classification confusion matrix.

The resulting confusion matrix (Figure 10b) contained 777 correctly classified other-crop parcels, 27 false-positive maize predictions, 27 false-negative maize predictions and 121 correctly classified maize parcels. Although achieving a lower overall performance, the model still correctly classified the vast majority of both maize and non-maize parcels. The equal number of false-positive and false-negative predictions, which also resulted in equal values for precision and recall, indicates that the model did not exhibit a systematic tendency to either over predict or under predict the target crop.

3.7.3. Rapeseed

To address the class imbalance, class weighting was incorporated during model training. Balanced class weights of approximately 0.5571 for the non-rapeseed class and 4.8769 for the rapeseed class were applied, assigning greater importance to the minority rapeseed class during optimization. The minimum validation loss was 0.0897 and the model achieved a validation accuracy of 97.48%, validation precision of 87.76%, and validation recall of 87.76%. The corresponding F1-score was 87.76%. The model also achieved a ROC-AUC of 0.9843 (Figure 11a), indicating a high discriminative capacity for distinguishing rapeseed parcels from other agricultural crops.

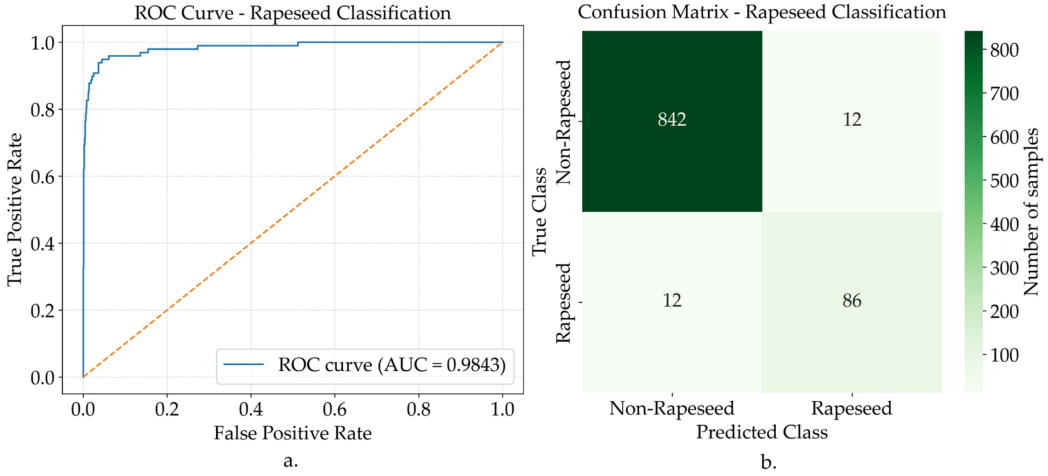


Figure 11. Rapeseed classification results: (**a**) ROC curve for the rapeseed classification model; (**b**) rapeseed classification confusion matrix.

The resulting confusion matrix (Figure 11b) contained 842 correctly classified non-rapeseed parcels, 12 false-positive rapeseed predictions, 12 false-negative rapeseed predictions and 86 correctly classified rapeseed parcels. These indicate that the model accurately distinguished rapeseed from the remaining crop types, with a low number of both false-positive and false-negative predictions.

3.7.4. Sunflower Classification Model Training

To address the class imbalance, class weighting was incorporated during model training. The final class weights were set to 1.0 for the non-sunflower class and 2.3 for the sunflower class, assigning greater importance to the minority sunflower class during optimization. The minimum validation loss was 0.0985 and the model achieved a validation accuracy of 97.06%, validation precision of 85.11% and validation recall of 85.11%. The corresponding F1-score was 85.11%. The model achieved a ROC-AUC of 0.9806 (Figure 12a), indicating a high discriminative capacity for distinguishing sunflower parcels from other agricultural crops.

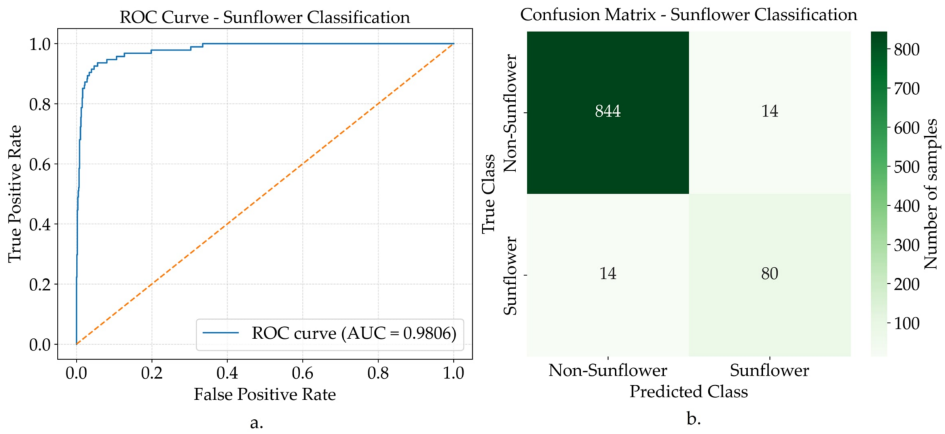


Figure 12. Sunflower classification results: (**a**) ROC curve for the sunflower classification model; (**b**) sunflower classification confusion matrix.

The resulting confusion matrix (Figure 12b) contained 844 correctly classified non-sunflower parcels, 14 false-positive sunflower predictions, 14 false-negative sunflower predictions and 80 correctly classified sunflower parcels. The low number of misclassifications relative to the total number of correctly classified parcels demonstrates the model's ability to reliably distinguish sunflower from the remaining crop types.

3.7.5. Watermelons and Melons

To address the class imbalance, class weighting was incorporated during model training. The final class weights were set to 1.0 for the non-watermelon class and 2.3 for the watermelon and melon class, assigning greater importance to the minority class during optimization. The minimum validation loss was 0.0861 and the model achieved a validation accuracy of 97.06%, validation precision of 76.62% and validation recall of 85.51%. The corresponding F1-score was 80.82%. The model achieved a ROC-AUC of 0.9803 (Figure 13a), indicating a high discriminative capacity for distinguishing watermelon and melon parcels from other agricultural crops.

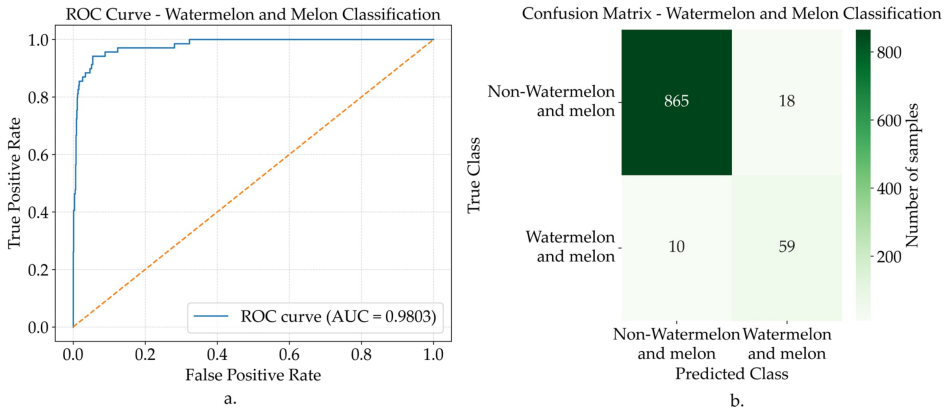


Figure 13. Watermelon and melon classification results: (**a**) ROC curve for the watermelon and melon classification model; (**b**) watermelon and melon classification confusion matrix.

The resulting confusion matrix (Figure 13b) contained 865 correctly classified non-watermelon parcels, 18 false-positive watermelon and melon predictions, 10 false-negative watermelon and melon predictions and 59 correctly classified watermelon and melon parcels. The higher number of false-positive than false-negative predictions indicates that the model was more likely to incorrectly classify other crop types as watermelons and melons rather than to miss actual watermelon and melon parcels. This behavior resulted in

a higher recall than precision, demonstrating that the model was effective at detecting the target crop, while producing a greater number of false-positive predictions. However, the high ROC-AUC value indicates a strong overall discriminative performance, despite the limited number of training samples.

3.8. Explainable GeoAI

The explainable Artificial Intelligence analysis was used to examine the relative contribution of the spectral features and acquisition dates to the crop classification predictions. Integrated Gradients were calculated for all validation parcels, and the resulting attribution values were aggregated across parcels, spatial pixels and the corresponding dimensions of the input data. Absolute attribution values were normalized to percentages, allowing the relative importance of the input features to be compared. Integrated Gradients were selected as the primary explainability method because it is well suited for Deep Neural Networks and provides values for individual input features. In this study, the CNN-LSTM model uses spatio-temporal inputs, consisting of multiple spectral bands and different observation dates. To assess the robustness of this XAI method, a complementary analysis was performed using SHapley Additive exPlanations (SHAP). The analysis was implemented using the GradientExplainer method and was selected because it provides a different formulation of feature attribution, based on the contribution of features relative to a background distribution. The integration of the two methods allows the consistency of the identified spectral and temporal importance patterns to be evaluated across two attribution approaches.

3.8.1. Wheat Classification

The aggregated spectral attribution results (Figure 14) showed that Band 5 (NDVI) was the most influential input feature in wheat classification, accounting for 50.78% of the total attribution according to the IG analysis and for 46.95% according to the SHAP analysis. This was followed by Band 4 (NIR), which accounted for 31.09% (IG) and 33.17% (SHAP). Band 2 contributed 8.86% (IG) and 10.20% (SHAP), while Band 1 and Band 3 had substantially lower relative contributions, accounting for 4.81% (IG)/4.57% (SHAP) and 4.45% (IG)/5.11% (SHAP), respectively. The results therefore indicate that the wheat classification model relied predominantly on NDVI and NIR. Both IG and SHAP showed very similar trends and ranking across all bands, strongly validating the model's reliance on NDVI and NIR.

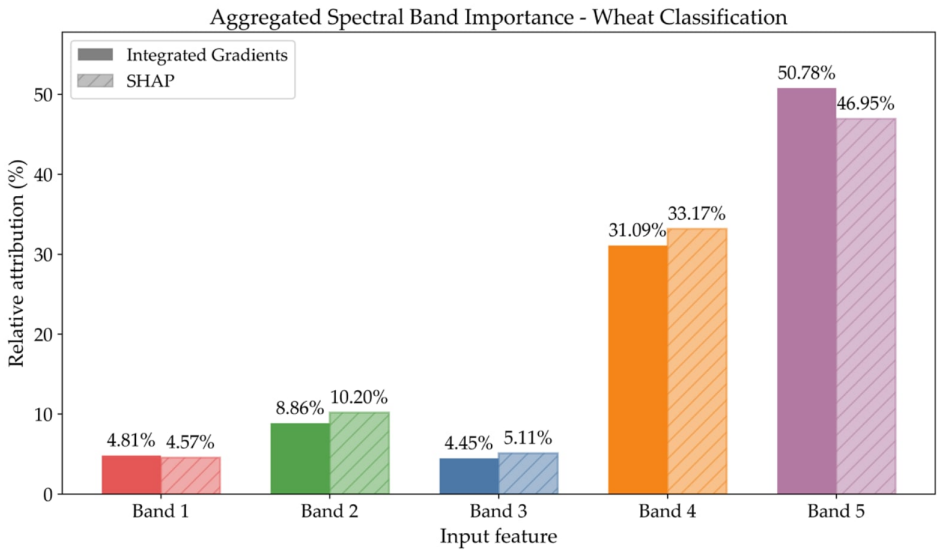


Figure 14. Aggregated spectral band importance for wheat classification.

The temporal attribution analysis (Figure 15) revealed a strong concentration of the model's attribution in the first three Sentinel-2 acquisitions. The 29 April 2025 acquisition had the highest relative attribution, accounting for 37.39% (IG)/36.73% (SHAP) of the total temporal attribution, followed by 4 May 2025, with 30.04% (IG)/33.95% (SHAP) and 8 June 2025, with 21.01% (IG)/15.99% (SHAP). The remaining acquisitions contributed substantially less, each contributing less than 5%. Both IG and SHAP analyses demonstrated strong consistency in ranking spring to early summer as the most relevant period for wheat discrimination.

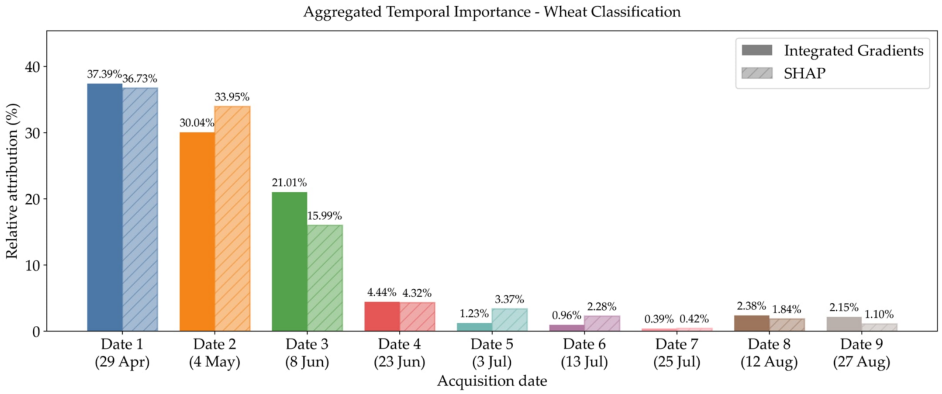


Figure 15. Aggregated temporal importance for wheat classification.

3.8.2. Maize Classification

The aggregated spectral attribution results (Figure 16) showed a notable divergence between the explainability methods, with IG ranking Band 4 (NIR) as the most influential input feature, accounting for 29.89% of the total attribution, followed by NDVI (25.01%) and SHAP ranking Band 5 (NDVI) highest (32.13%), followed by NIR (30.29%). Overall, the results indicate that the model relied most strongly on NIR and NDVI for distinguishing maize parcels from other agricultural crops which accounted in total for 54.90% (IG) and 62.42% (SHAP) of the model's predictions. The Red, Green and Blue bands had a significantly larger contribution to the model's decisions than in the wheat classification, accounting in total for 45.10% (IG) and 37.58% (SHAP).

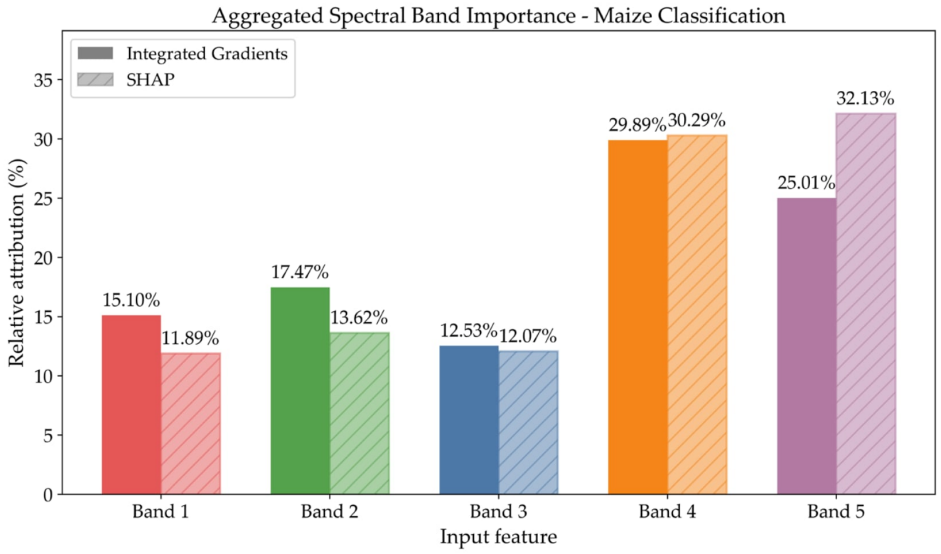


Figure 16. Aggregated spectral band importance for maize classification.

The temporal attribution analysis (Figure 17) revealed a more distributed contribution across the nine Sentinel-2 acquisition dates. The IG analysis revealed that the highest relative attribution was obtained for the 29 April 2025 acquisition (14.10%), followed by 4 May 2025 (13.72%) and 8 June 2025 (13.53%). The 23 June 2025 acquisition accounted for 11.58%, while the subsequent dates contributed 8.24% (3 July), 10.12% (13 July), 7.50% (25 July), 11.14% (12 August) and 10.07% (27 August). The SHAP analysis revealed a substantial contribution of the late summer dates to the models' decisions, with 25.43% for 27 August and 14.73% for 12 August.

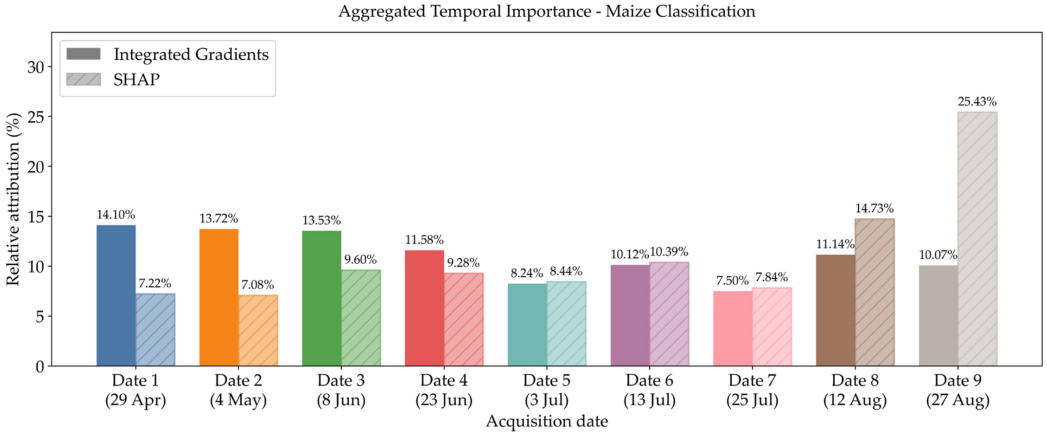


Figure 17. Aggregated temporal importance for maize classification.

3.8.3. Rapeseed Classification

For the rapeseed classification model, the spectral attribution analysis (Figure 18) showed that NDVI was the most influential input feature, followed by the fourth spectral band and the second spectral band. The relative contributions were 39.30% (IG) and 38.11% (SHAP) for NDVI, 21.77% (IG) and 21.92% (SHAP) for Band 4, 21.07% (IG) and 22.03% (SHAP) for Band 2, 12.30% (IG) and 12.42% (SHAP) for Band 3 and 5.56% (IG) and 5.52% (SHAP) for Band 1. The results indicate that the model relied strongly on vegetation-related information and showed very similar band contributions between the two XAI methods.

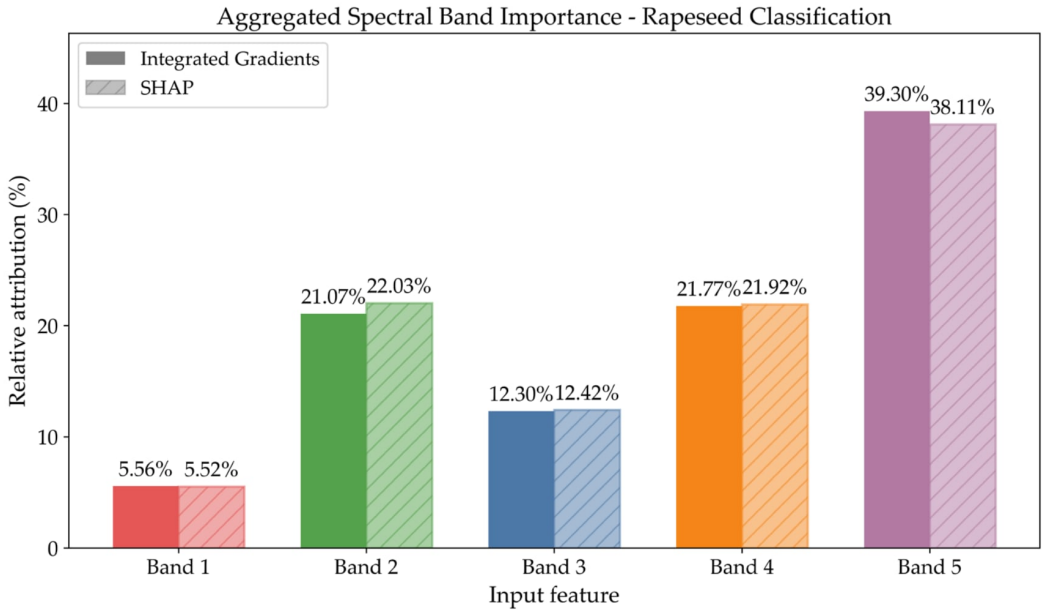


Figure 18. Aggregated spectral band importance for rapeseed classification.

The temporal attribution analysis (Figure 19) showed that the contribution of the nine acquisition dates was not uniform throughout the growing season. The acquisition on 8 June 2025 was the most influential temporal input, accounting for 28.87% (IG) and 37.66% (SHAP) of the total attribution. Both analyses revealed high contributions of the late spring–early summer dates and less contributions from the images acquired at the end of July and August.

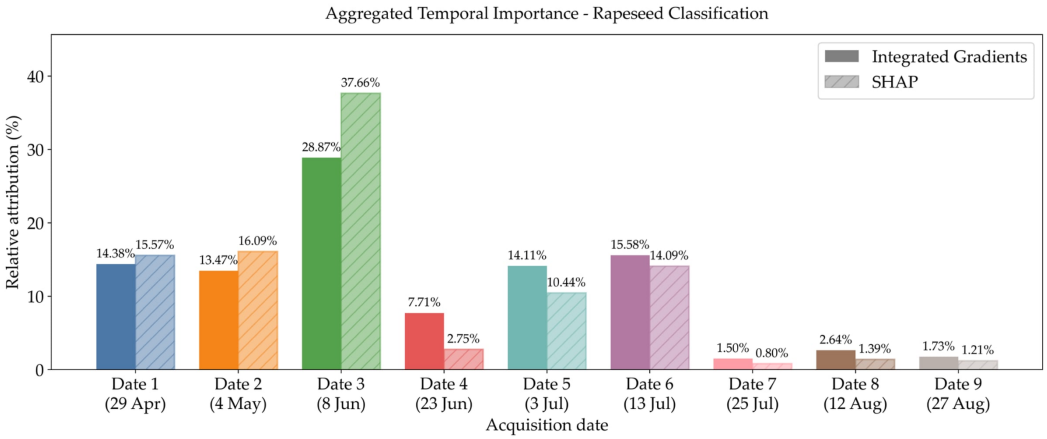


Figure 19. Aggregated temporal importance for rapeseed classification.

3.8.4. Sunflower

For the sunflower classification model, the spectral attribution analysis (Figure 20) showed that Band 4 (NIR) was the most influential input feature, with a relative contribution of 41.00% (IG) and 41.41% (SHAP). The other relative contributions were 21.54% (IG) and 21.63% (SHAP) for NDVI, 14.28% (IG) and 15.29% (SHAP) for Band 2, 12.46% (IG) and 9.81% (SHAP) for Band 3 and 10.73% (IG) and 11.86% (SHAP) for Band 1. The results from both analyses indicate that the model relied strongly on information contained in Band 4, while NDVI represented the second most influential feature.

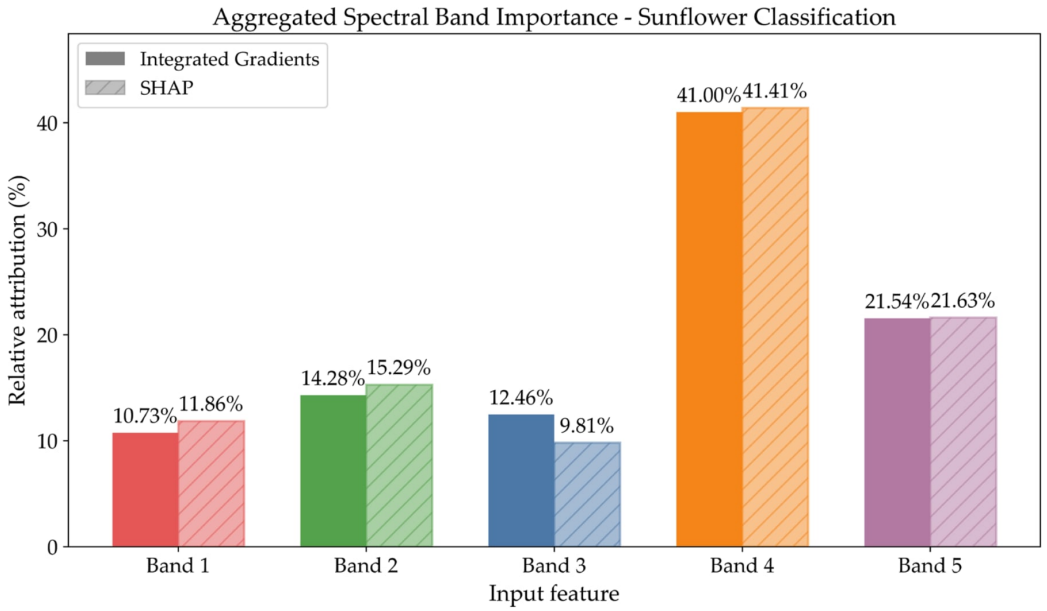


Figure 20. Aggregated spectral band importance for sunflower classification.

The temporal attribution analysis (Figure 21) showed that the contribution of the nine acquisition dates was not uniform throughout the growing season. The sunflower crop

exhibited two distinct peaks of temporal importance—23 June which contributed with 15.09% (IG) and 19.93% (SHAP) and 27 August, which contributed with 21.80% (IG) and 16.79% (SHAP). In contrast, early season acquisition dates contributed minimally.

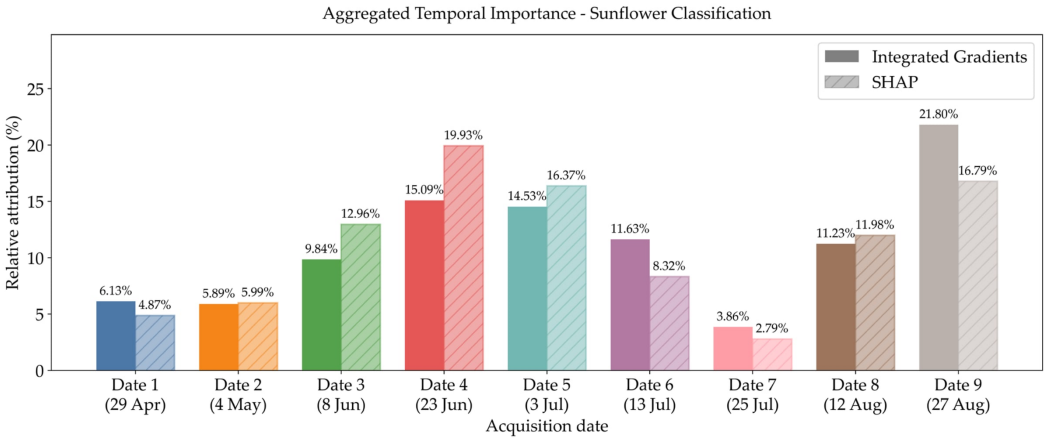


Figure 21. Aggregated temporal importance for sunflower classification.

3.8.5. Watermelon and Melon

For the watermelon and melon classification model, the spectral attribution analysis (Figure 22) showed that NDVI was the most influential input feature, which accounted for 40.01% (IG) and 41.46% (SHAP) of the aggregated attribution, followed by Band 4 (NIR), contributing 23.68% (IG) and 23.22% (SHAP) of the total attribution. The other relative contributions were 7.56% (IG) and 7.02% (SHAP) for Band 1, 8.28% (IG) and 7.77% (SHAP) for Band 2 and 20.47% (IG) and 20.53% (SHAP) for Band 3. The results from both analyses showed very similar values and indicate that the model relied strongly on NDVI.

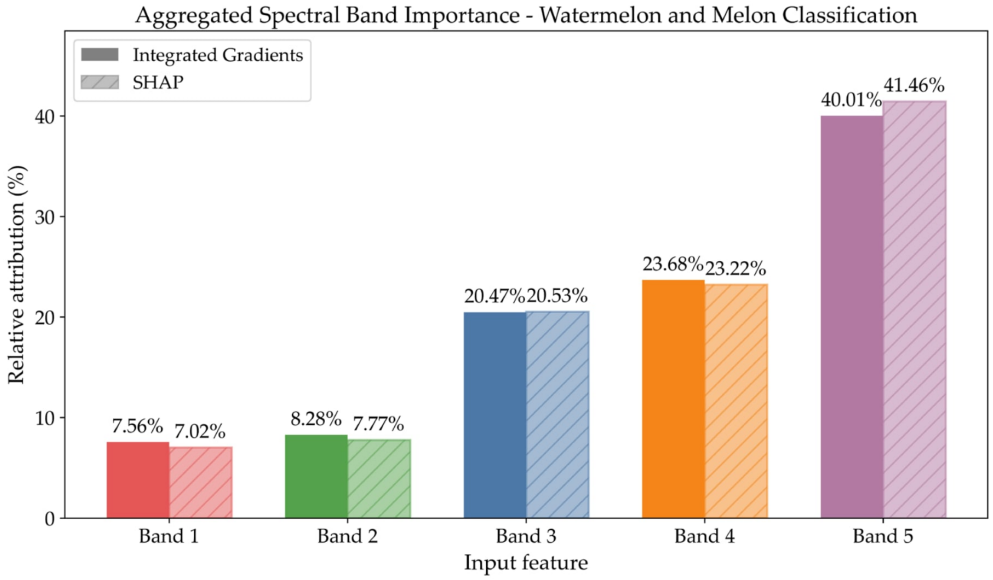


Figure 22. Aggregated spectral band importance for watermelon and melon classification.

The temporal attribution analysis (Figure 23) showed that the contribution of the nine acquisition dates was not uniform throughout the growing season. The acquisition on 25 July 2025 was the most influential temporal input, according to the IG analysis, accounting for 20.17% of the total attribution while the acquisition on 27 August was the second most influential date, contributing 16.99%. The SHAP analyses ranked the acquisition on 27 August the highest, accounting for 22.31% of the total attribution, followed by the

acquisition on 25 July (19.69%). Both analyses provided the same ranking for the rest of the acquisition dates.

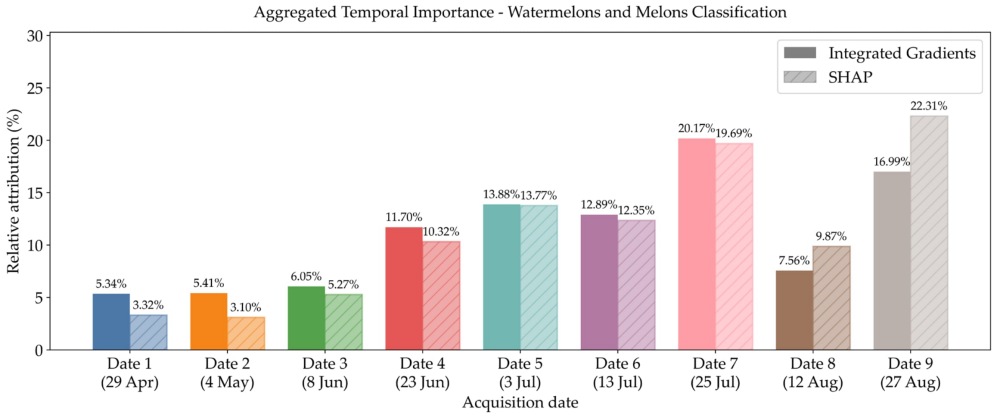


Figure 23. Aggregated temporal importance for watermelon and melon classification.

4. Results

The evaluation of the crop classification models demonstrates high overall effectiveness (Figure 24), with validation accuracies ranging from 94.33% for maize, to 97.48% for rapeseed. The wheat classification model, with the highest number of samples, recorded the highest validation precision (92.50%) while the watermelon and melon classification model, with the lowest number of samples, achieved the lowest validation precision (76.62%). The binary classification model's ability to distinguish between the two classes was measured with the ROC-AUC performance metric, which ranged between 98.43% for rapeseed and 96.97% for maize, demonstrating the models' capacity to accurately make distinctions between the selected crop type and other crops. Given the class imbalance in the training dataset, the F1-score metric was used to evaluate the performance of the models. The wheat classification model achieved the highest F1-score (92.12%), while the watermelon and melon classification model achieved the lowest F1-score (80.82%).

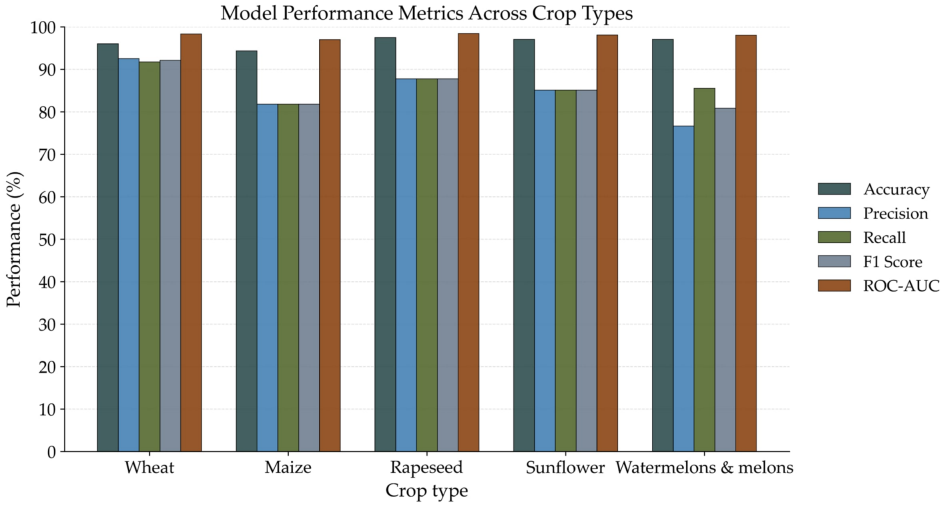


Figure 24. Model performance metrics across crop types.

The observed differences in performance can be partially attributed to the unequal distribution of training samples among crop classes. Wheat, which was represented by the largest number of parcels, achieved the best overall performance, whereas watermelon and melon, the least represented class, exhibited the lowest performance. The consistently

high ROC-AUC values demonstrate that the one-versus-rest strategy effectively separates each target crop from the remaining crop classes, despite the variability among the non-target class. The results indicate that the spatial and temporal information captured by the Sentinel-2 image series provides distinctive phenological signatures that enable reliable discrimination between crop types throughout the growing season.

To provide a more comprehensive evaluation of the five classification models, Balanced Accuracy and Macro F1-score were also calculated. These metrics were included to account for the differences in the representation of the positive (target crop) and negative (non-target crops) classes and are presented in Table 4. Wheat and rapeseed classification models showed high Balanced Accuracy and Macro F1-scores, indicating that the models performed very well for both target crop and non-target crop samples. The sunflower classification model showed similarly strong results, while for the watermelon and melon classification model, the precision/recall balance is less consistent between the target and non-target crop. The maize classification model recorded the lowest values among the five models, although its Balanced Accuracy and Macro F1-score still indicate good overall classification performance.

Table 4. Balanced Accuracy and Macro F1-Score.

Classified Crop	Balanced Accuracy	Macro F1-Score
Wheat	94.60%	94.72%
Maize	89.20%	89.20%
Rapeseed	93.17%	93.17%
Sunflower	91.74%	91.74%
Watermelon and melon	91.73%	89.61%

While the quantitative evaluation confirms the predictive capability of the proposed framework, these metrics do not reveal the features underlying the classification decisions. Therefore, two explainability analyses based on Integrated Gradients and SHAP were conducted to identify the spectral bands and acquisition dates that contributed most to each crop classification.

Based on the explainable GeoAI analysis, important conclusions regarding the influence of each spectral band (Figure 25) and each acquisition date (Figure 26) can be drawn. Spectral bands 4 (NIR) and 5 (NDVI) generally demonstrate the highest overall importance percentages across the majority of evaluated crops, compared to bands 1, 2 and 3. Wheat classification relies most heavily on the NDVI band, which accounts for 50.78% (IG) and 45.95% (SHAP) of the total contributions of spectral bands to the model's prediction while 29 April was identified as the acquisition date with the greatest influence. This suggests that the model exploits the characteristic phenological development of wheat during the peak vegetative growth stage, when the crop exhibits distinct canopy structure. For maize classification, the NIR and NDVI bands were the most influential spectral features. The temporal analysis revealed a relatively balanced contribution from multiple acquisition dates which suggests that the model relied on phenological information captured across several stages of crop development. The higher contribution of the NIR band highlights the importance of canopy structure and biomass in distinguishing maize from other crop types. Although both XAI methods resulted in balanced contributions among the acquisition dates, the resulting temporal importance was showing several discrepancies. This, combined with the reduced classification performance of the maize model, could be attributed to certain limitations of the dataset represented by differences between maize crops caused by different agricultural practices within the same area. For rapeseed, the NDVI band was the most influential spectral feature, accounting for 39.30% (IG) and 38.11% (SHAP) of the

model's predictions, while the 8 June acquisition date contributed to the highest temporal importance (28.87% for IG and 37.66% for SHAP analyses). These results indicate that the model primarily relied on canopy development during the later stages of the growing season, when its phenological development differs significantly from that of other crops. The significant contribution of the green band (Band 2) is consistent with the distinctive yellow flowering stage, which reflects heavily in the green regions of the visible spectrum. Temporal contributions were also distributed across the flowering window (late spring–early summer). For sunflower, the NIR band was the most influential spectral feature, accounting for 41.00% (IG) and 41.41% (SHAP) of the model's predictions. The XAI analysis revealed two distinct peaks of temporal importance (23 June and 27 August). The first major temporal importance peak occurred during active canopy growth and flowering while the second peak occurred at the end of the season. In contrast, early season acquisitions contributed minimally, due to sparse canopy cover. These findings indicate that the model primarily relied on canopy structure and biomass-related information captured by the NIR band, while the phenological characteristics during the last stages of sunflower development provide higher discriminative information. For watermelon and melon, NDVI was the dominant spectral feature, contributing to 40.01% (IG) and 41.46% (SHAP) of the model's predictions, whereas the 25 July and 27 August acquisition dates had the greatest temporal importance. This pattern suggests that the discrimination of watermelons and melons was driven primarily by the dense vegetation cover observed during the middle and end of the growing season, when the crop shows distinctive spectral responses compared to the surrounding agricultural fields. The significant reliance on band 3 (Blue) can be attributed to the distinct structural layouts of these fields and the spreading vines of the crops with wide spaces between them, leaving considerable amounts of bare soil or reflective plastic exposed.

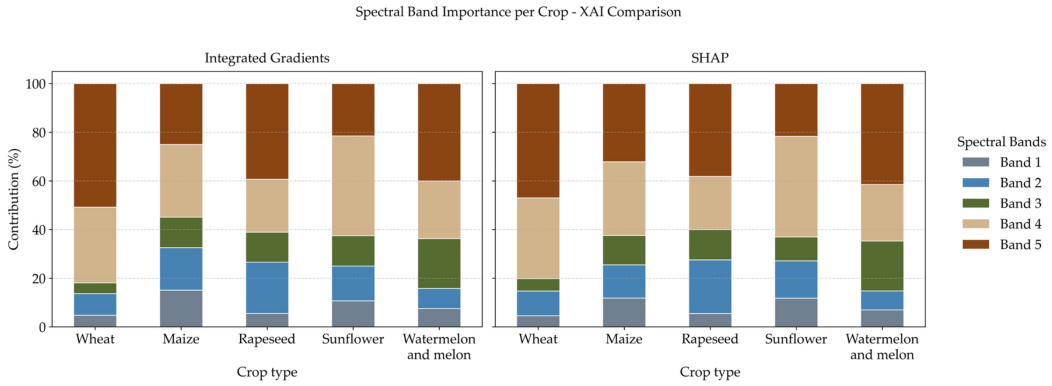


Figure 25. Spectral band importance per crop.

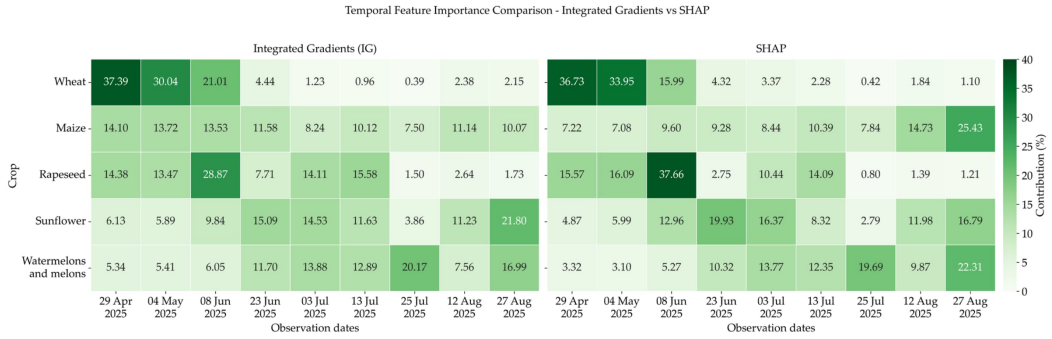


Figure 26. Temporal feature importance heat map.

5. Discussion

The use of nine Level-2A Sentinel-2 observations distributed from April to August provided multi-temporal information describing the evolution of the crops within one growing season. These products provide atmospheric corrections and were selected to ensure minimal cloud coverage over the study area. However, crop spectral response may vary between agricultural seasons as a result of differences in meteorological conditions, including temperature, precipitation and water availability. Such variability can influence the time of crop development, biomass accumulation and yield, leading to changes in the spectral characteristics observed by satellite sensors. The one-versus-rest strategy provides an advantage in the context of this variability. Meteorological conditions may affect different crop types to different degrees, resulting in changes in their spectral characteristics and in such cases, the structure of the proposed approach allows the model associated with a specific crop to be independently retrained using data from the new season.

Monitoring the validation loss proved to be an effective criteria for model selection, while the EarlyStopping callback enabled the identification of the best-performing model before the onset of overfitting. Several models continued to improve on the training dataset after the optimal validation epoch, whereas the validation loss gradually increased. Restoring the model weights corresponding to the minimum validation loss therefore ensured that the final models retained the best generalization capacity. This approach was particularly beneficial considering the limited number of available training samples and the class imbalance present in the agricultural dataset.

The proposed one-versus-rest classification strategy effectively addressed the class imbalance in the ground truth dataset, as each model focused exclusively on distinguishing a single target crop, rather than learning decision boundaries among multiple crop classes. Furthermore, this approach provides greater flexibility, allowing individual models to be independently optimized. This workflow also facilitates the future inclusion of additional crop types without requiring the complete retraining of the classification model.

The use of the Integrated Gradients and SHAP XAI methods, which provided similar results, enhanced the transparency of the proposed CNN-LSTM framework, by identifying the spectral bands and acquisition dates that contributed most to each classification decision. Unlike conventional DL approaches, which are often regarded as "black-box" models, the proposed framework provides quantitative evidence regarding the information used during classification. The explainability analyses demonstrated that the models relied on meaningful information, in accordance with the growing stages of each crop, with NDVI and NIR emerging as the most influential spectral features. These findings increase confidence in the models' predictions by demonstrating that the classification process is based on vegetation dynamics.

6. Conclusions

The proposed framework has considerable potential for supporting smart agriculture and sustainable land management. Accurate and interpretable crop classification provides timely information on crop distribution, enabling efficient monitoring of agricultural practices and crop rotation patterns. The explainability component increases confidence in automated classification results, allowing decision-makers to better understand and validate the models' predictions before integrating them into operational workflows.

The proposed methodology is also consistent with current EU policies promoting digital transformation and sustainable agriculture. By combining multi-temporal Sentinel-2 imagery, ground truth data provided by official authorities (APIA), DL and XAI, the proposed framework contributes to the development of reliable decision-support systems ca-

pable of assisting farmers, public authorities, and paying agencies in agricultural monitoring and policy development and implementation.

7. Limitations and Future Research Directions

Despite the promising results, several limitations of this study should be acknowledged. First, although the data were acquired on nine different dates, this research was conducted using data from a single growing season (2025), which may not fully capture the variability resulting from different growing conditions each year, caused by weather and different agricultural management practices. Further evaluation using multi-year datasets is therefore necessary to assess the generalization and temporal transferability of the proposed models across different growing seasons. Second, the ground truth dataset was obtained from farmers' declarations provided by the Romanian Agency for Payments and Interventions in Agriculture (APIA). Although this data represent the official reference source for agricultural parcel information, they may contain inaccuracies resulting from reporting errors or other discrepancies between declared and actual crop conditions. Such uncertainties may affect the model training process and influence the final classification performance. Another limitation is related to the fact that the selected crop types showed an imbalanced distribution of training samples, which influenced the performance obtained for the less represented classes, despite the adoption of one-versus-rest classification strategy and class weights.

Future research should include the use of images and ground truth data from multiple growing seasons and expand the methodology to larger areas, with larger training datasets. Future investigations should examine the sensitivity of the classification performance to different class-weight values in the one-versus-rest framework and compare the proposed approach with a corresponding multi-class CNN in terms of computational complexity, as well as inference efficiency. Such comparisons could help determine which approach is more suitable for crop classification on imbalanced datasets, considering both predictive performance and computational efficiency. Future studies could also investigate the integration of other remote sensing data, such as UAV-acquired imagery. From the explainability perspective, employing other XAI methods could provide a more comprehensive understanding of model behavior.

Finally, Explainable GeoAI crop classification models could be further integrated into agricultural digital triplets, to support not only the real-time monitoring of crop conditions, but also the simulation of future scenarios and the timely optimization of agricultural practices.

Author Contributions: Conceptualization, V.A.N., G.B., A.C.B., A.P.G., M.V.H. and L.O.D.; methodology, V.A.N. and A.P.G.; validation, V.A.N., G.B., A.P.G., A.C.B., M.V.H. and L.O.D.; formal analysis, V.A.N., G.B., A.C.B. and A.P.G.; investigation, V.A.N. and A.P.G.; resources, V.A.N. and A.P.G.; data curation, V.A.N. and A.P.G.; writing—original draft preparation, V.A.N. and A.P.G.; writing—review and editing, G.B., A.C.B., M.V.H. and L.O.D.; visualization, V.A.N., G.B., A.C.B., A.P.G., M.V.H. and L.O.D.; supervision, G.B. and A.C.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the GNaC2025 ARUT-UTCB-7 Project "Research on the Use of Artificial Intelligence in Remote Sensing for Sustainable Agriculture"—AgriAI. The article processing charge is funded by the Doctoral School of the Technical University of Civil Engineering, Bucharest.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and legal reasons. Restrictions apply to the availability of these data: training data were obtained from the Romanian Agency for Payments and Intervention

in Agriculture (APIA) and are available upon request at https://apia.org.ro/ with the permission of APIA.

Acknowledgments: This study was conducted within the Geodetic Engineering Measurements and Spatial Data Infrastructures Research Centre, Faculty of Geodesy, Technical University of Civil Engineering Bucharest. The research presented in this study contributes to the implementation of GNaC2025 ARUT-UTCB-7 Project "Research on the Use of Artificial Intelligence in Remote Sensing for Sustainable Agriculture"–AgriAI. The ground truth data used in this study were provided by the Romanian Agency for Payments and Intervention in Agriculture (APIA) for the 2025 year. During the preparation of this manuscript, the authors used ChatGPT (GPT-5.6, OpenAI) for English language editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
APIA	Agency for Payments and Intervention in Agriculture
CAP	Common Agricultural Policy
CNN	Convolutional Neural Network
DL	Deep learning
DNN	Deep Neural Networks
EU	European Union
GeoAI	Geospatial Artificial Intelligence
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
MLP	multi-layer perceptron
MODIS	Moderate Resolution iImaging Spectroradiometer
PSE-CNN	Pixel-Set Encoder Convolutional Neural Network
RF	Random Forest
RNN	Recurrent Neural Networks
ROC-AUC	the Area Under the Receiver Operating Characteristic Curve
SHAP	SHapley Additive exPlanations
UAV	Unmanned Aerial Vehicle)
XAI	Explainable Artificial Intelligence

References

1. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.-J. Big Data in Smart Farming—A Review. *Agric. Syst.* **2017**, *153*, 69–80. [CrossRef]
2. Raj, M.; Prahadeeswaran, M. Revolutionizing Agriculture: A Review of Smart Farming Technologies for a Sustainable Future. *Discov. Appl. Sci.* **2025**, *7*, 937. [CrossRef]
3. Donmez, E.; Heckelei, T.; Storm, H. Satellite Remote Sensing-Based Crop Cover Classification over Europe: Accuracy of Different Methodological Approaches. *Int. J. Remote Sens.* **2025**, *46*, 8251–8294. [CrossRef]
4. Yao, J.; Wu, J.; Xiao, C.; Zhang, Z.; Li, J. The Classification Method Study of Crops Remote Sensing with Deep Learning, Machine Learning, and Google Earth Engine. *Remote Sens.* **2022**, *14*, 2758. [CrossRef]
5. Kumhálová, J.; Sedlák, J.; Marčan, J.; Vandírková, V.; Novotný, P.; Kohútek, M.; Kumhála, F. Monitoring of Agricultural Crops by Remote Sensing in Central Europe: A Comprehensive Review. *Remote Sens.* **2026**, *18*, 1075. [CrossRef]
6. Zhang, C.; Kerner, H.; Wang, S.; Hao, P.; Li, Z.; Hunt, K.A.; Abernethy, J.; Zhao, H.; Gao, F.; Di, L.; et al. Remote Sensing for Crop Mapping: A Perspective on Current and Future Crop-Specific Land Cover Data Products. *Remote Sens. Environ.* **2025**, *330*, 114995. [CrossRef]
7. Abou Samra, R.M.; Ali, R.R. Explainable Artificial Intelligence in Smart Agriculture: A Comprehensive Review of Interpretable Remote Sensing for Sustainable Decision-Making. *AgriEngineering* **2026**, *8*, 270. [CrossRef]
8. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. [CrossRef]

9. Rajbongshi, A.; Johora, F.T.; Hossain, A.; Sarker, M.D.S.; Rahman, M.H.; Rahman, M.W.; Alotaibi, F.T.; Moni, M.A. Leveraging Explainable AI for Sustainable Agriculture: A Comprehensive Review of Recent Advances. *Artif. Intell. Rev.* **2026**, *59*, 105. [CrossRef]
10. AI Act | Shaping Europe's Digital Future. Available online: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (accessed on 2 August 2026).
11. Eco-Schemes—Agriculture and Rural Development—European Commission. Available online: https://agriculture.ec.europa.eu/common-agricultural-policy/income-support/eco-schemes_en (accessed on 2 August 2026).
12. Romania—Agriculture and Rural Development—European Commission. Available online: https://agriculture.ec.europa.eu/cap-my-country/cap-strategic-plans/romania_en (accessed on 2 August 2026).
13. România—Agricultură și Dezvoltare Rurală—Comisia Europeană. Available online: https://agriculture.ec.europa.eu/cap-my-country/cap-strategic-plans/romania_ro (accessed on 2 August 2026).
14. Agenţia de Plăţi şi Intervenţie Pentru Agricultură. Available online: https://apia.org.ro/ (accessed on 2 August 2026).
15. December 2025 How Diverse Are EU Crops? Available online: https://storymaps.arcgis.com/stories/b57718885fde42968b1519032565e81e (accessed on 2 August 2026).
16. Zheng, H.; Wang, H.; Dong, H.; Qian, Y. A Survey of Crop Disease Recognition Methods Based on Spectral and RGB Images. *J. Imaging* **2026**, *12*, 66. [CrossRef]
17. Gao, J.; Gujarati, K.; Hegde, M.; Arra, P.; Gupta, S.; Buch, N. Integration of UAV and Remote Sensing Data for Early Diagnosis and Severity Mapping of Diseases in Maize Crop Through Deep Learning and Reinforcement Learning. *Remote Sens.* **2025**, *17*, 3427. [CrossRef]
18. Ma, Z.; Wang, C.; Wang, X.; Chen, X. Advances in Artificial Intelligence-Enabled Crop Pest and Disease Detection: A Systematic Review. *Agriculture* **2026**, *16*, 1262. [CrossRef]
19. Yang, M.-D.; Lee, T.-H.; Tseng, H.-H.; Su, T.-C.; Hsu, Y.-C. Label-Free Deep Learning with Feature Adaptation for Crop Anomaly Detection on Small Datasets. *Agriculture* **2026**, *16*, 854. [CrossRef]
20. Chouriya, A.; Soni, P.; Chandel, A.K.; Patel, A.K. An AI-Enabled System for Automated Plant Detection and Site-Specific Fertilizer Application for Cotton Crops. *Automation* **2025**, *6*, 53. [CrossRef]
21. Adhinata, F.D.; Wahyono; Sumiharto, R. A Comprehensive Survey on Weed and Crop Classification Using Machine Learning and Deep Learning. *Artif. Intell. Agric.* **2024**, *13*, 45–63. [CrossRef]
22. Glykos, I.; Peteinatos, G.G.; Arvanitis, K.G. Indirect Crop Line Detection in Precision Mechanical Weeding Using AI: A Comparative Analysis of Different Approaches. *Eng. Proc.* **2025**, *104*, 32. [CrossRef]
23. Ma, H.; Li, M.; Jin, X.; Jiang, S.; Cui, H.; Li, X.; Yang, C.; Zhang, K.; Lu, J. Integrating Multi-Source and Multi-Temporal Features for Winter Wheat Yield Estimation Using Vegetation Indices and Growth Indicators. *Agronomy* **2026**, *16*, 1419. [CrossRef]
24. Liu, J.; Liu, T.; Huang, L.; Ren, Y.; He, P. Next-Generation Drought Forecasting: Hybrid AI Models for Climate Resilience. *Remote Sens.* **2025**, *17*, 3402. [CrossRef]
25. Zhao, M.; Hu, X.; Han, D.; Guo, P.; Wang, Z.; Ning, W.; Wan, H. Early-Season Classification and Mapping of Winter Wheat from Multi-Source Satellite Imagery and Evaluation of Optimal Observation Windows for Agricultural Production Management. *Sustainability* **2026**, *18*, 7657. [CrossRef]
26. Tapia, A.A.; Bennett, A. A Comparative Analysis of Dynamic Time Warping and Machine Learning Models for Crop Classification: Case Study of Limarí River Basin, Chile. *Earth* **2026**, *7*, 122. [CrossRef]
27. Nan, V.A.; Badea, G.; Badea, A.C.; Grădinaru, A.P. A Systematic Review of AI-Based Classifications Used in Agricultural Monitoring in the Context of Achieving the Sustainable Development Goals. *Sustainability* **2025**, *17*, 8526. [CrossRef]
28. Argenton, E.V.; Gomede, E.; Mendes, L.D.S. Resource-Aware Citrus Crop Mapping from Sentinel-2 Time Series Using a Pixel-Set Encoder Convolutional Neural Network for Sustainable Agricultural Monitoring. *Green* **2026**, *1*, 5. [CrossRef]
29. Yin, F.; Yu, X.; Wang, Y.; Liu, L. Progressive Deep Learning for Accurate Winter Rapeseed Mapping in Complex Terrain: A Case Study of Hanzhong Basin, China. *Remote Sens.* **2026**, *18*, 1706. [CrossRef]
30. da Silva, M.P.; Correa, S.P.L.P.; Schaefer, M.A.R.; Reis, J.C.S.; Nunes, I.M.; dos Santos, J.A.; Oliveira, H.N. Advancing Agricultural Remote Sensing: A Comprehensive Review of Deep Supervised and Self-Supervised Learning for Crop Monitoring. *Comput. Graph.* **2025**, *133*, 104434. [CrossRef]
31. Gallo, I.; La Grassa, R.; Landro, N.; Boschetti, M. Sentinel 2 Time Series Analysis with 3D Feature Pyramid Network and Time Domain Class Activation Intervals for Crop Mapping. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 483. [CrossRef]
32. Zheng, Z.; Yuan, J.; Yao, W.; Yao, H.; Liu, Q.; Guo, L. Crop Classification from Drone Imagery Based on Lightweight Semantic Segmentation Methods. *Remote Sens.* **2024**, *16*, 4099. [CrossRef]
33. Quach, L.-D.; Quoc, K.N.; Quynh, A.N.; Thai-Nghe, N.; Nguyen, T.G. Explainable Deep Learning Models with Gradient-Weighted Class Activation Mapping for Smart Agriculture. *IEEE Access* **2023**, *11*, 83752–83762. [CrossRef]
34. Malashin, I.; Tynchenko, V.; Gantimurov, A.; Nelyub, V.; Borodulin, A.; Tynchenko, Y. Predicting Sustainable Crop Yields: Deep Learning and Explainable AI Tools. *Sustainability* **2024**, *16*, 9437. [CrossRef]

35. Choi, J.W.; Hidayat, M.S.; Cho, S.B.; Hwang, W.-H.; Lee, H.; Cho, B.-K.; Kim, M.S.; Baek, I.; Kim, G. Recent Trends in Machine Learning, Deep Learning, Ensemble Learning, and Explainable Artificial Intelligence Techniques for Evaluating Crop Yields Under Abnormal Climate Conditions. *Plants* **2025**, *14*, 2841. [CrossRef]
36. Ullah, N.; Ruocco, M.; Della Cioppa, A.; De Falco, I.; Sannino, G. An Explainable Deep Learning Framework with Adaptive Feature Selection for Smart Lemon Disease Classification in Agriculture. *Electronics* **2025**, *14*, 3928. [CrossRef]
37. Chihaoui, N.; Waleed, A.; Mufleh, A.S.S.; Alshraah, S.M.; Alajmi, L.H.R.; Altuwayjiri, S. A Comprehensive Review on AI-Based Crop Disease Detection Using Leaf Image Classification and Explainable AI. *Discov. Appl. Sci.* **2026**, *8*, 663. [CrossRef]
38. Lin, T.; Lau, R.Y.K.; Hu, S. A Critical Review of State-of-the-Art Explainable Artificial Intelligence (XAI) Methods and Their Business Applications. *Artif. Intell. Rev.* **2026**, *59*, 161. [CrossRef]
39. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. Available online: https://arxiv.org/html/2305.02012v3 (accessed on 2 August 2026).
40. Li, Q.; Tian, J.; Tian, Q. Deep Learning Application for Crop Classification via Multi-Temporal Remote Sensing Images. *Agriculture* **2023**, *13*, 906. [CrossRef]
41. Zhao, H.; Duan, S.; Liu, J.; Sun, L.; Reymondin, L. Evaluation of Five Deep Learning Models for Crop Type Mapping Using Sentinel-2 Time Series Images with Missing Information. *Remote Sens.* **2021**, *13*, 2790. [CrossRef]
42. Tufail, R.; Tassinari, P.; Torreggiani, D. Deep Learning Applications for Crop Mapping Using Multi-Temporal Sentinel-2 Data and Red-Edge Vegetation Indices: Integrating Convolutional and Recurrent Neural Networks. *Remote Sens.* **2025**, *17*, 3207. [CrossRef]
43. Chan, A.; Schneider, M.; Körner, M. XAI for Early Crop Classification. In *Proceedings of the IGARSS 2023—2023 IEEE International Geoscience and Remote Sensing Symposium*; Curran Associates, Inc.: New York, NY, USA, 2023; pp. 2657–2660.
44. National Commission for Strategy and Prognosis—Report on the Economy of Ialomița County. Available online: https://cnp.ro/ (accessed on 2 August 2026).
45. Scharvogel, D.; Brandmeier, M.; Weis, M. A Deep Learning Approach for Calamity Assessment Using Sentinel-2 Data. *Forests* **2020**, *11*, 1239. [CrossRef]
46. Doi, K.; Iwasaki, A. The Effect of Focal Loss in Semantic Segmentation of High Resolution Aerial Image. In *Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: New York, NY, USA, 2018; pp. 6919–6922.
47. Chen, J.; Yuan, Z.; Peng, J.; Chen, L.; Huang, H.; Zhu, J.; Liu, Y.; Li, H. DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1194–1206. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.