

## Article

# Machine-Learning Crop-Type Mapping Sensitivity to Feature Selection and Hyperparameter Tuning

Mayra Perez-Flores <sup>1,2</sup>, Frédéric Satgé <sup>3,\*</sup>, Jorge Molina-Carpio <sup>2</sup>, Renaud Hostache <sup>3</sup>, Ramiro Pillco-Zolá <sup>2</sup>, Diego Tola <sup>2,3,4,5</sup>, Elvis Uscamayta-Ferrano <sup>3,6</sup>, Lautaro Bustillos <sup>2,3</sup>, Marie-Paule Bonnet <sup>3</sup> and Celine Duwig <sup>1</sup>

<sup>1</sup> IGE, University Grenoble Alpes, CNRS, INRAE, IRD, Grenoble INP, 34800 Grenoble, France; mayra.perez@ird.fr (M.P.-F.); celine.duwig@ird.fr (C.D.)

<sup>2</sup> Instituto de Hidráulica e Hidrología (IHH), Universidad Mayor de San Andrés, La Paz 10077, Bolivia; amolina@umsa.bo (J.M.-C.); rpillco@umsa.edu.bo (R.P.-Z.); 20211517@lamolina.edu.pe (D.T.); lautaro.bustillos-pena@ird.fr (L.B.)

<sup>3</sup> ESPACE-DEV, University Montpellier, IRD, University Antilles, University Guyane, University Réunion, 34093 Montpellier, France; renaud.hostache@ird.fr (R.H.); elvis.uscamayta@ird.fr (E.U.-F.); marie-paule.bonnet@ird.fr (M.-P.B.)

<sup>4</sup> Programa de Doctorado en Recursos Hídricos (PDRH), Universidad Nacional Agraria La Molina, Lima 15024, Peru

<sup>5</sup> Área de Ciencias Agrícolas, Pecuarias y Recursos Naturales (ACAPRN), Universidad Pública de El Alto, La Paz 10077, Bolivia

<sup>6</sup> Graduate Program in Applied Geosciences and Geodynamics, Institute of Geosciences, University of Brasília (UnB), Brasília 70910-900, Brazil

\* Correspondence: frederic.satge@ird.fr

## Highlights

### What are the main findings?

- Crop-type mapping reliability is highly sensitive to features selection and hyperparameter tuning.
- The model and target independent VIF feature selection is not recommended for crop-type mapping

### What are the implications of the main findings?

- Most reliable crop-type mapping is obtained through a proposed three-step process combining wrapped features selection with hyperparameter tuning.
- Based on open-access data and software, the proposed method can be used to support agriculture monitoring in a complex socio-economic context.

## Abstract

To improve crop yields and incomes, farmers consistently adapt their practices to climate and market fluctuations, resulting in highly variable crop field distribution and coverage in space and time. As these dynamics illustrate farmers' challenges, up-to-date crop-type mapping is essential for understanding farmers' needs and supporting their adoption of sustainable practices. With global coverage and frequent temporal observations, remote sensing data are generally integrated into machine learning models to monitor crop dynamics. Unlike physical-based models that rely on straightforward use, implementing machine learning models requires extensive user interaction. In this context, this study assesses how sensitive the models' outputs are to feature selection and hyperparameter tuning, as both processes rely on user judgment. To achieve this, Sentinel-1 (S1) and Sentinel-2 (S2) features are integrated into five distinct models (Random Forest (RF), Support Vector Machine (SVM), Light Gradient Boosting (LGB), Histogram-based Gradient Boosting (HGB), and Extreme Gradient Boosting (XGB)), considering several features selection (Variance



Academic Editor: Xiaoyang Zhang

Received: 17 December 2025

Revised: 3 February 2026

Accepted: 6 February 2026

Published: 11 February 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](https://creativecommons.org/licenses/by/4.0/).

Inflation Factor (VIF) and Sequential Feature Selector (SFS)) and hyperparameter tuning (Grid-Search) setup. Results show that the preprocess modeling feature selection (VIF) discards the features that the wrapped method (SFS) keeps, resulting in less reliable crop-type mapping. Additionally, hyperparameter tuning appears to be sensitive to the input features, and considering it after any feature selection improved the crop-type mapping. In this context a three-step nested modeling setup, including first hyperparameter tuning, followed by a wrapped feature selection (SFS) and additional hyperparameter tuning, leads to the most reliable model outputs. For the study region, LGB and XGB (SVM) are the most (least) suitable models for crop-type mapping, and model reliability improves when integrating S1 and S2 features rather than considering S1 or S2 alone. Finally, crop-type maps are derived across different regions and time periods to highlight the benefits of the proposed method for monitoring crop dynamics in space and time.

**Keywords:** crop-type mapping; sentinel; machine learning; Bolivia; Altiplano

## 1. Introduction

In recent years, remote sensing data have enabled the development of accurate crop-type mapping to monitor agricultural land distribution in space and time [1,2]. As a detection and monitoring tool, remote sensing has many advantages, including cost-effectiveness and the capacity to cover large areas over extended periods [3]. In this way, the lack of data and their limited availability in regions with complex socio-economic systems can be avoided [2,4]. Remote sensing-based crop-type mapping relies on visible-near-infrared (VIS-NIR) region sensitivity to crop features [5], such as total biomass [6], nitrogen [7,8], and chlorophyll content [9,10], allowing for their identification and classification. For this purpose, the Sentinel-2 VIS-NIR images (S2) are generally used [11–21]. As the VIS-NIR satellite images are limited to clear sky conditions, some authors have assessed the reliability of synthetic aperture radar (SAR) satellite images for crop-type mapping to ensure continuous data acquisition even in cloud-covered regions and/or periods [2,22,23]. Freely available at the global scale, the Sentinel-1 C-band SAR images (S1) are generally used [24–29]. Finally, some authors have taken advantage of both VIS-NIR and SAR satellite images to improve crop-type mapping reliability [30–36].

In these studies, machine learning (ML) models allow for addressing the complex patterns linking VIS-NIR and/or SAR signals with the observed crop-type, thereby facilitating automated decision-making processes [37,38]. ML models such as Random Forest (RF) and Support Vector Machine (SVM) are generally used for satellite-based crop-type mapping [39], due to their robustness in data processing, offering a reliable classification when compared to other models [30,31,33,34,36,39–42]. Despite their advantages in various scenarios, these models have limitations in modeling complex, high-order interactions among variables. As a powerful and flexible alternative, the Gradient Boost (GB) algorithm has emerged [43,44]. In recent years, several optimized implementations of GB, including Extreme Gradient Boosting (XGB), Light Gradient Boosting (LGB), and Histogram-based Gradient Boosting (HGB), have been developed, with significant improvements in Land Use Land Cover (LULC) classification applications [45–52]. In some regions, XGB has been found to provide more reliable LULC classification than both RF and SVM [51], while LGB has been found to be more reliable than both RF, SVM, and XGB in other regions [46,48–50]. The examples mentioned above show that no ML model systematically outperforms the others, turning crop-type mapping sensitive to the selected ML models.

Contrary to physical-based models (e.g., hydrological models) that rely on predefined input variables to run specific internal equations, ML models adapt their internal structure to map input variables (predictive variables) to the variable to be predicted (dependent variable). Therefore, ML models can be trained on any input variables, but their performance depends mainly on the relevance of those variables [35,36,53,54]. In this context, many authors use feature selection processes to address this challenge. The Variance Inflation Factor (VIF), which assesses the extent to which the variance of the estimated regression coefficients is inflated by linear dependence among predictor variables, is commonly used [36,54,55]. Despite its effectiveness in reducing multicollinearity among the predictor variables, omitting the dependent variable in VIF-based feature selection results in the retention (discarding) of irrelevant (relevant) predictor variables. To overcome this issue, wrapper features selection methods, such as the Genetic Algorithm (GA) [42,56–59], Recursive Feature Elimination (RFE) [60,61], or the Sequential Feature Selector (SFS) [53,62–65], are generally used. Model-dependent wrapper methods are based on the ML model's performance [66,67]. In wrapped methods, different feature sets are iteratively considered to assess the ML model's predictive capabilities in order to retain the feature subset that produces the most reliable ML model predictions. By considering both the dependent variable and the ML model, wrapped methods improve feature selection and, consequently, the reliability of the final ML output [63,67–70].

The ML model's training process relies on different hyperparameters. While some authors have adopted the default ML hyperparameter values [21,71–73], others have implemented optimization techniques (i.e., Grid-Search) to enhance ML training [30,52,74,75]. Grid-Search optimization consists of running the ML model with all possible hyperparameter combinations to identify the combination leading to the most reliable ML configuration. Although hyperparameter tuning often results in values that differ from the default settings [75], the sensitivity of ML's reliability to hyperparameter tuning has not been extensively studied [76,77].

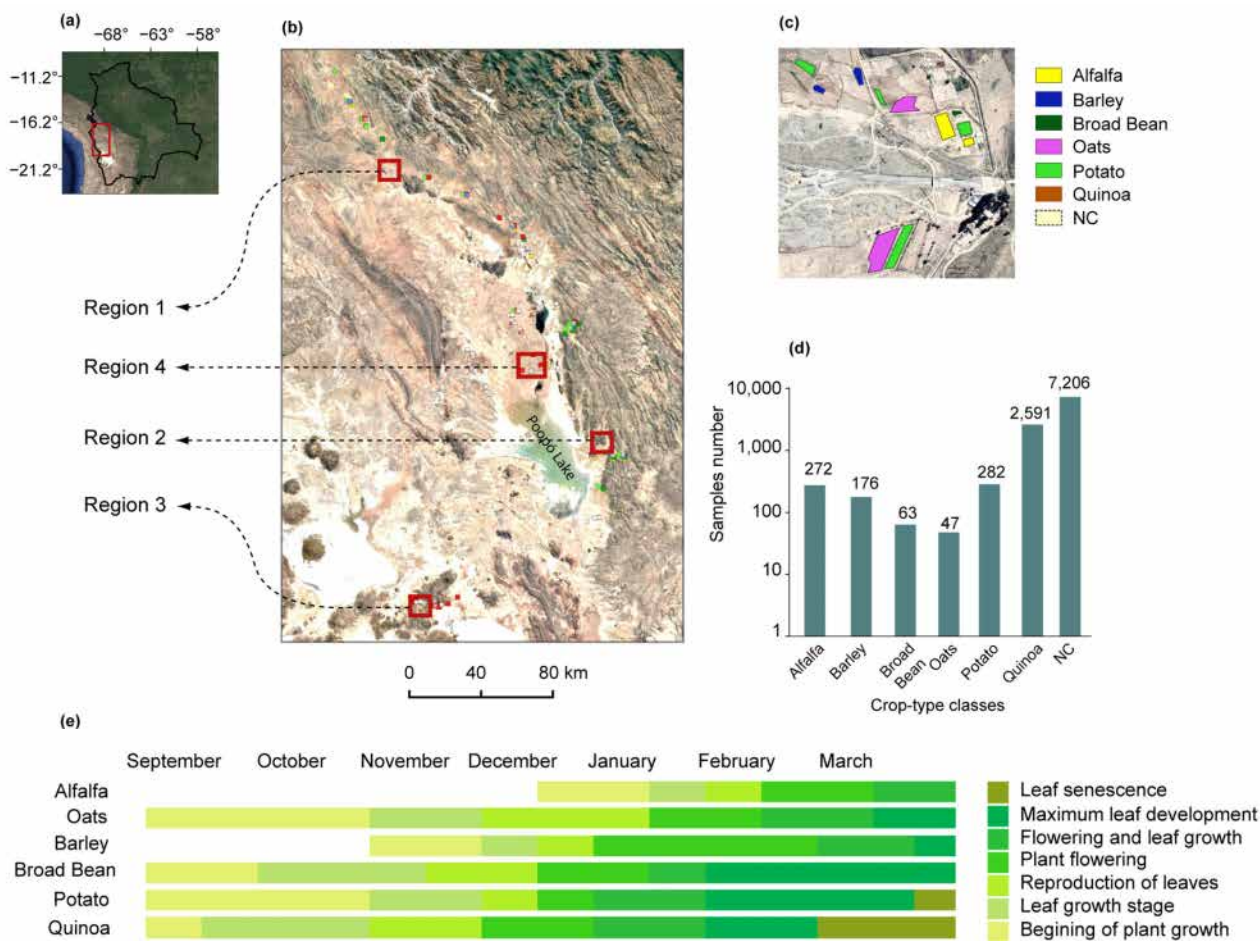
In this context, remote sensing-based crop-type mapping is expected to be sensitive to (i) the choice of the ML model, (ii) the input variables selected, and (iii) the ML model's hyperparameter values. This study assesses this sensitivity by testing five ML models (RF, SVM, XGB, HGB, and LGB) using Sentinel-2 (VIS-NIR) and Sentinel-1 (SAR) features as predictive variables under various combinations of feature selection (VIF and SFS) and hyperparameter tuning (Grid-Search). The goal is to identify the most reliable ML setup for satellite-based crop-type mapping. Although focused on crop-type mapping, this sensitivity analysis can apply to any ML modeling forecast, underscoring the importance of rigorous strategies for ML models and feature selection, along with hyperparameter optimization, all controllable by the ML end-user.

## 2. Materials

### 2.1. Study Area

The study area is located in the central part of the Bolivian Altiplano, around the Poopó basin (16°54'S, 66°20'W–20°01'S, 67°55'W), at a mean elevation of approximately 3700 m.a.s.l., in the plain [78–80] (Figure 1). In this region, agriculture (along with mining) is the main economic activity. The main crops grown in the area are potato, quinoa, and alfalfa, along with other crops such as broad beans and barley [50,54]. Potatoes are a key source of revenue for farmers and are significant for the development of other crops due to their role in crop rotation systems. Seeds such as quinoa are often sown in soils where potatoes have previously been cultivated [81–83]. Historically, quinoa cropping was mainly located in the southern part of the studied region, known as the intersalar (saltpan) zone, between the Poopó Basin and the Salar de Uyuni [84]. By 2020, quinoa

cultivation had expanded significantly along Lake Poopó's eastern coast—encompassing Toledo, Corque, Huari, and Challapata—engaging roughly 6000 families [78,85] Quinoa's exceptional nutritional profile has driven rising global demand, yet its future cultivation in this region is at serious risk of becoming unsustainable, due to water scarcity, as has already been observed in the area [4,86].



**Figure 1.** Study area location (a,b) with agricultural plot delimitations (c), number of observations per crop type (d), and phenological stages of the considered crop type (e). Adapted for sowing and harvesting periods in the region [87–90].

## 2.2. Reference Observations

Field campaigns were conducted between February and March 2022 to delineate crop and non-crop (NC) areas using a GPS navigator (Garmin GPS Map 65s, Schaffhausen, Switzerland). This period corresponds to the stage of maximal leaf development for the studied crops (Figure 1e), which facilitates crop-type differentiation in both the field and the S1 and S2 images. The selections focused on the six main crops in the region (quinoa, potato, broad beans, barley, alfalfa, and oats) (Figure 1c,d). Then, a 5-m buffer was applied to all delineated polygons to ensure that the included Sentinel pixels (S1 and S2) were influenced solely by the target crop type and not by mixed signals from adjacent areas.

## 2.3. Sentinel-1 Images and Preprocessing

Sentinel-1 (S1) is a constellation of two satellites launched with 180° opposing orbits in April 2014 (S1-A) and April 2016 (S1-B, decommissioned in December 2021 following a power supply issue). S1-A has a 12-day revisit time period and acquires radar signals (C-band) according to two polarizations, vertical–vertical and vertical–horizontal (VV and

VH, respectively) in interferometry mode with a central frequency of 5.405 GHz. S1 images are available (i) as a Single Look Complex (SLC) product, which includes both phase and amplitude data, and (ii) as a Ground Range Detected (GRD) product, which includes direct surface characteristics through intensity images. Both SLC and GRD products are available in ascending and descending orbits. To ensure consistency in the S1 observations across space and time, only the GRD products from the ascending orbit were used. To cover the entire area of interest and to minimize potential signal interference, three mosaics (made of two S1 scenes) for three consecutive dates were made. The mosaicked images were then stacked to generate a composite image, from which the mean VV and VH backscatter values were extracted. Prior to the elaboration of the composite images, all required S1 scenes were downloaded from the Canadian open-access platform Alaska Satellite Facility (ASF) and preprocessed according to five successive steps: (i) edge noise removal, (ii) calibration, (iii) terrain correction, (iv) thermal noise removal, and (v) speckle filter pixel smoothing to improve the visualization of the images [2,30,38]. S1 preprocessing was carried out using the Sentinel Application Platform (SNAP) 9.0, free software designed specifically for handling Sentinel images [23].

#### 2.4. Sentinel-2 Images and Preprocessing

Sentinel-2 (S2) images are from a constellation of two satellites launched in June 2015 (S2-A) and March 2017 (S2-B), arranged 180° apart, with a 5-day revisit time. S2 images are composed of 13 spectral bands operating in (i) the visible (B1, B2, B3, and B4), (ii) the near-infrared (B6, B7, B8, B8a, and B9), and (iii) the shortwave infrared (B10, B11, and B12) spectrum, with the following different spatial resolutions: 10 m (B2, B3, B4, and B8), 20 m (B5, B6, B7, B8A, B11, and B12), and 60 m (B1 and B9). In this study, S2-A images are used because they provide orthorectified, atmospherically corrected earth-surface spectral information (i.e., Bottom of Atmosphere—BOA). To cover the study area, a total of eight S2 scenes were downloaded from the European Space Agency (ESA) website. To minimize cloud coverage that would prevent observation of the delimited agricultural plots, images acquired on 29 March 2022 were used, as they had cloud cover of less than 20% and corresponded to the maximum leaf development of the crops (Figure 1e).

#### 2.5. Machine Learning Models

Five classification machine learning (ML) models were selected for this study: (i) Random Forest (RF), (ii) Support Vector Machine (SVM), (iii) Histogram-based Gradient Boosting (HGB), (iv) Extreme Gradient Boosting (XGB), and Light Gradient Boosting (LGB). These models were chosen as they all include a class weight hyperparameter that allows for dealing with unbalanced datasets. This is crucial for addressing the imbalance in the number of observations per class in the available learning database (Figure 1d). It achieves this by adjusting the classification cost for underrepresented classes (those with fewer observations) [91]. This correction prevents the model outputs from becoming biased toward the classes with the highest number of observations in the learning database [59,74].

RF is based on decision trees [92]. This algorithm consists of training multiple trees on random subsets of the observations and predictor variables of the training sample [93]. RF is widely used in statistical classification and in solving non-parametric regression problems [94–96]. Due to its high adaptability to different observed data, RF is increasingly used for LULC classification [2,21,34,51,52].

SVM is a classifier-supervised learning algorithm based on the separation of data by hyperplanes [97], which is a non-parametric technique for discriminating classes based on differences in their characteristics [52,98]. This technique is commonly used in remote sensing due to its high capability to handle non-normal data. Additionally, it allows for

the selection of a kernel function to process distinct data and to fit the optimal model for the intended application. SVM is regarded as one of the top-performing classifiers and has been successfully applied to LULC mapping in various regions [39,54,99,100].

XGB is a tree-based model that employs gradient boosting to optimize a loss function. It sequentially refines weak models to correct prior errors, achieving high accuracy and efficiency. However, its performance may decline in high dimensional spaces [101], requiring proper preprocessing and feature selection. For tasks such as crop-type identification and land-use classification, XGB frequently demonstrates superior performance compared to other machine learning models [42,47,52,102].

LGB is a GB variant designed for faster training and greater efficiency [49,103]. It is particularly valued for handling large datasets with relatively low GPU resource consumption [48]. This algorithm has been used in various land cover mapping studies [41,48–50,61], frequently outperforming other machine learning models [43,48–50].

HGB is an improved version of GB, that reduces model processing time by discretizing continuous input variables into histogram-based sets, making memory usage more efficient when dealing with large datasets. Although HGB was previously used for different tasks, such as accident prediction [104], computer security [105], or in the industry [106,107], its use in environmental science is limited, with only one known LULC classification study [108]. Consequently, this study will assess its suitability for crop-type mapping and benchmark its performance against the other well-established models (RF, SVM, XGB, and LGB).

### 3. Methods

#### 3.1. Machine Learning Database Elaboration

Firstly, S1 polarization (VV and VH) and S2 reflectance (B1, B2, B3, B4, B5, B6, B7, B8, B8a, B9, B11, and B12) data were resampled to a 20-m spatial resolution using the nearest neighbor method [59,109,110]. Geometric alignment between the S1 and S2 images was then performed with the SNAP collocation tool, which leverages image metadata to ensure accurate spatial overlap, particularly for areas not affected by topography [23].

Secondly, to capture specific crop traits, a comprehensive set of predictors was computed. This set included the 12 S2-spectral reflectance bands and 12 vegetation indices (VIs) commonly used for LULC classification (Table 1) [2,31,38,51,111–113]. Regarding S1, in addition to the VV and VH polarizations, 3 polarization indices (PIs) [1,30,38,114] and 16 texture indices (TIs) were derived (Table 2). The TIs have been shown to enhance vegetation classification accuracy [30,115,116] and were calculated using the Gray-Level Co-Occurrence Matrix window size  $7 \times 7$  (GLCM). The S1 local incidence angle (LIA) is also considered, as it significantly influences the energy reflected to the sensor, with variations resulting from different geographical conditions [117].

**Table 1.** S2 features used in the modeling process.

Index/Acronym	Definition	References
Sentinel-2 bands	B1, B2-Blue, B3-Green, B4-Red, B5-Rededge1, B6-Rededge2, B7-Rededge3, B8-NIR, B8a-Rededge4, B11-SWIR1, B12-SWIR2	
BSI	$(\text{SWIR1} + \text{RED} - \text{NIR} - \text{BLUE}) / (\text{SWIR1} + \text{NIR} + \text{RED} + \text{BLUE})$	[54]
EVI	$(2.5 \times (\text{NIR} - \text{RED})) / (\text{NIR} + 6 \times \text{RED} - 7.5 \times \text{BLUE} + 1)$	[54,118]
GNDVI	$(\text{NIR} - \text{GREEN}) / (\text{NIR} + \text{GREEN})$	[119]
MNDWI	$(\text{GREEN} - \text{SWIR2}) / (\text{GREEN} + \text{SWIR2})$	[120]

Table 1. Cont.

Index/Acronym	Definition	References
NDMI	$(\text{NIR} - \text{SWIR1})/(\text{RED} + \text{SWIR1})$	[54]
NDVI	$(\text{NIR} - \text{RED})/(\text{NIR} + \text{RED})$	[54]
NDWI	$(\text{GREEN} - \text{NIR})/(\text{GREEN} + \text{NIR})$	[120]
SAVI	$((\text{NIR} - \text{RED})/(\text{NIR} + \text{RED} + 0.428)) \times 1.428$	[118]
IAF	$(\text{VNIR4} - \text{VNIR1})/(\text{VNIR4} + \text{VNIR1})$	[17]
TC_Brightness	$0.3037 \text{ BLUE} + 0.2793 \text{ GREEN} + 0.4743 \text{ RED} + 0.5585 \text{ NIR} + 0.5082 \text{ SWIR1} + 0.1863 \text{ SWIR2}$	[54]
TC_Greenness	$0.2848 \text{ BLUE} - 0.2435 \text{ GREEN} - 0.5436 \text{ RED} + 0.7243 \text{ NIR} + 0.084 \text{ SWIR1} - 0.18 \text{ SWIR2}$	[54]
TC_Wetness	$0.1509 \text{ BLUE} + 0.1973 \text{ GREEN} + 0.3279 \text{ RED} + 0.3406 \text{ NIR} - 0.7112 \text{ SWIR1} - 0.4572 \text{ SWIR2}$	[54]

Table 2. S1 features used in the modeling process.

Index/Acronym	Definition	References
Sentinel-1 Polarization	VV and VH	[23]
LIA		[121]
Ratio	VV/VH	[1]
RVI	$4 \cdot \text{VH}/(\text{VV} + \text{VH})$	[122]
NDI VV	$(\text{VV} - \text{VH})/(\text{VV} + \text{VH})$	[38,54]
NDI VH	$(\text{VH} - \text{VV})/(\text{VH} + \text{VV})$	[38,54]
VV_GLCM	VV_Contrast, VV_Dissimilarity, VV_Homogeneity, VV_AngularSecondMoment, VV_Energy, VV_Entropy, VV_Correlation, VV_Mean, and VV_Variance	[87]
VH_GLCM	VH_Contrast, VH_Dissimilarity, VH_Homogeneity, VH_AngularSecondMoment, VH_Energy, VH_Entropy, VH_Correlation, VH_Mean, and VH_Variance	[87]

Finally, the values of these features were extracted from the pixels within the delineated agricultural plots. The resulting learning database comprises 10,637 observations, with 2591, 282, 176, 63, 47, 272, and 7206 corresponding to quinoa, potatoes, barley, broad beans, oats, alfalfa, and non-crops, respectively (Figure 1d). The database was then randomly split into training (70%) and validation (30%) datasets.

### 3.2. Feature Selection

Multicollinearity and hyperparameters are two critical aspects to handle to achieve reliable machine learning outputs [39]. Multicollinearity is a well-known problem that can reduce model robustness due to redundancy among the selected independent variables. To address this point, the Variance Inflation Factor (VIF) is commonly employed [41,63,74,78]. The VIF assesses the correlation among independent variables to identify redundancy and thus to eliminate variables. As a model-independent process, the VIF does not consider the model's structural sensitivity to the features taken (i) independently and/or (ii) in combination with the feature sensitivity to the dependent variable. As a result, the VIF may discard (maintain) some features that could have been relevant (irrelevant). In this context, wrapper feature selection methods, such as the Genetic Algorithm (GA), Recursive

Feature Elimination (RFE), or the Sequential Feature Selector (SFS), offer the advantage of accounting for the model's feature sensitivity, enabling model-specific feature subset selection. In fact, the wrapper approach consists of assessing the model's sensitivity to various feature subsets to retain the one leading to the most reliable model prediction. Despite higher computational costs, the wrapper method significantly increases model's reliability rather than the simple VIF method [59]. This study employs sequential feature selection (SFS), a method that has been shown to be effective for LULC classification [63,65]. The SFS operates under an iterative procedure that incrementally constructs an optimal feature subset by adding or removing features based on model performance [60]. It comes in two main variants: sequential forward selection, which starts with no features and adds them one at a time, and sequential backward selection, which starts with the full set of features and removes them one at a time. At each step, sequential forward selection (backward) adds (removes) the feature and selects the one whose modification leads to the best model performance. In this study, the SFS with the sequential forward selection is adopted.

### 3.3. Hyperparameter Tuning

Each machine learning model (i.e., RF, SVM, XGB, HGB, and LGB) comes with its own hyperparameters that governs its learning process (Table 3). Optimizing the hyperparameters is crucial to significantly increase the model's performance [42] and helps to prevent issues of overfitting or underfitting [123,124]. The Grid-Search tool is commonly used to explore all possible hyperparameter combinations using a k-fold validation to evidence the hyperparameter combinations leading to the most reliable model output. Many authors have used this procedure to improve the model's reliability for soil moisture [123], soil salinity [125], and LULC mapping [28,52].

**Table 3.** The models' hyperparameter tuning.

Model	Hyperparameters	Range	Default Values	Optimum Values		
				Scenario 1 (S1)	Scenario 2 (S2)	Scenario 3 (S1 + S2)
RF	n_estimators	50–200 (Step = 10)	100	180	200	180
	max_features	1-N, sqrt, log2	sqrt	3	6	5
	max_depth	10, 20, 25, 30	None	20	30	25
	bootstrap	True, False	True	True	False	True
SVM	C	0.1, 10, 100	1	10	100	10
	gamma	0.001, 0.01, 0.1, 1	auto	0.001	1	0.001
	kernel	Rbf', 'Linear'	rbf	rbf	rbf	Linear
LGB	n_estimators	50–200 (Step = 10)	100	130	200	100
	colsample_bytree	0–1 (Step = 0,1)	1	0.4	0.4	0.8
	max_depth	10, 20, 25, 30	–1	25	10	10
	num_leaves	31, 41, 51, 61	31	31	41	61
XGB	n_estimators	50–200 (Step = 10)	100	200	70	200
	colsample_bytree	0–1 (Step = 0,1)	1	0.8	0.7	0.7
	max_depth	10, 20, 25, 30	6	20	10	10
HGB	max_iter	50–200 (Step = 10)	100	190	170	170
	max_leaf_nodes	None, 30, 60	30	None	None	60
	max_depth	10, 20, 25, 30	None	25	25	30
	max_features	0–1 (Step = 0,1)	1	0.9	0.8	0.4

It is worth mentioning here that other techniques (i.e., Random-Search tool) are available for hyperparameter tuning. While Grid-Search explores the model for sensitivity to all

hyperparameter combinations possible, the Random-Search tool evaluates a fixed number of hyperparameter combinations making it more efficient for high dimensional datasets. As the considered dataset includes more observations than variables (i.e., low dimensional dataset) we selected the Grid-Search tool rather than the Random-Search tool to reduce the randomness in hyperparameter tuning and to ensure the selection of the optimum hyperparameter combinations.

### 3.4. Issues Related to Unbalanced Datasets

An imbalanced training dataset—where certain classes contain significantly more observations than others—tends to increase (decrease) the prediction accuracy of the class with the highest (lowest) number of observations [97]. To mitigate this effect and to enhance overall model efficiency, a class weight hyperparameter was systematically applied during the training of each algorithm (RF, SVM, LGB, XGB, and HGB). A class weight hyperparameter assigns class-specific weights to encourage the model to emphasize minority classes and to mitigate bias toward majority classes. Higher weights increase the penalty for misclassifying minority classes, whereas lower weights decrease the influence of frequent errors on majority classes in the overall loss.

### 3.5. Machine Learning and Dataset Assessment

To assess the sensitivity of crop-type mapping to machine learning models and remote sensing features, three scenarios are considered. The first scenario (Scenario-1) uses S1 polarization along with TIs, PIs, and LIA as independent variables to assess the standalone S1 potential. The second scenario (Scenario-2) employs S2 bands along with VIs as independent variables to determine the standalone S2 potential. The third scenario (Scenario-3) integrates independent variables from Scenario-1 and Scenario-2 to assess the complementary value of combining S1 and S2 data. Each scenario is evaluated for the five machine learning models (RF, SVM, XGB, HGB, and LGB) using the commonly used two-step approach applying (i) the VIF feature selection and (ii) hyperparameter tuning (Grid-Search) on top of the VIF-selected features. In this process, the learning database is separated in two sets: the training set (gathering 70% of the total observations) and the validation set (30% of the total observations). For each model and scenario, the hyperparameters are tuned on the training set via the Grid-Search function with 5-fold cross-validation using overall accuracy (OA, Equation (3)) as the objective function. Following hyperparameter tuning, each model is trained for each scenario (Scenario-1, Scenario-2, and Scenario-3) with the optimum hyperparameter combination using the training dataset and validated using the validation dataset. For the validation step, the model performance is evaluated using a suite of metrics derived from the confusion matrix: Precision, Recall, Overall Accuracy (OA), and the F1 Scores (Equations (1)–(4)). These metrics quantify the predictive performance by measuring the balance between true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), with scores ranging from 0 to 1, where 1 represents the optimal performance.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (2)$$

$$\text{OA} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (3)$$

$$\text{F1 score} = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

### 3.6. Crop-Type Mapping Sensitivity to Feature Selection and Hyperparameter Setup

Using the most reliable model and scenario (i.e., Scenario-3) identified in the previous section, different feature selection and hyperparameter tuning combinations are

assessed. In this process, the commonly used two-step approach (VIF, Grid-Search) is used as a benchmark (hereafter referred to as SU-1) and compared to five feature selection and hyperparameter setups (SU). SU-2 consists of running the SFS with the optimized hyperparameters obtained from the VIF-selected features (SU-1), while SU-3 performs additional hyperparameter tuning on top of the SFS-selected features (SU-2). SU-4 consists of running hyperparameter tuning considering all available features (without the VIF), while SU-5 consists of running the SFS with the optimized hyperparameters obtained in SU-4. Finally, SU-6 consists of running additional hyperparameter tuning on top of the SFS-selected features (SU-5). The SUs' consistency is assessed through the Precision, Recall, OA, and F1 Scores along with their respective confusion matrices, while the input features' relevance is quantified by the Permutation Importance (PI) value. The PI value measures the sensitivity of a trained model to each feature, with values ranging from 0 to 1. It is calculated in two steps. First, the model is evaluated on the validation dataset to obtain its original accuracy (i.e., OA). Then, for each feature, the values are randomly shuffled, and the model is run again using the modified validation dataset. The PI value is determined by subtracting the post-shuffle from the original accuracy. Thus, the PI value reflects the reduction in model accuracy that occurs when the values of each variable are independently disrupted.

### 3.7. Crop-Type Mapping

To provide more insight into the differences among the model–scenario combination outputs, crop-type maps derived from the best model–scenario–SU combination are provided for Regions 1, 2, and 3 (Figure 1b). Then, the best model–scenario–SU combination is used to develop crop-type maps for the 2019–2024 (6-year) period to highlight the benefits of the proposed method for crop monitoring in space and time. For this step, the mapping is limited to Region 4 (Figure 1), a region where cloud-free images were available for each year (2019–2024) and for the maximum leaf development of the studied crops (March, Figure 1e).

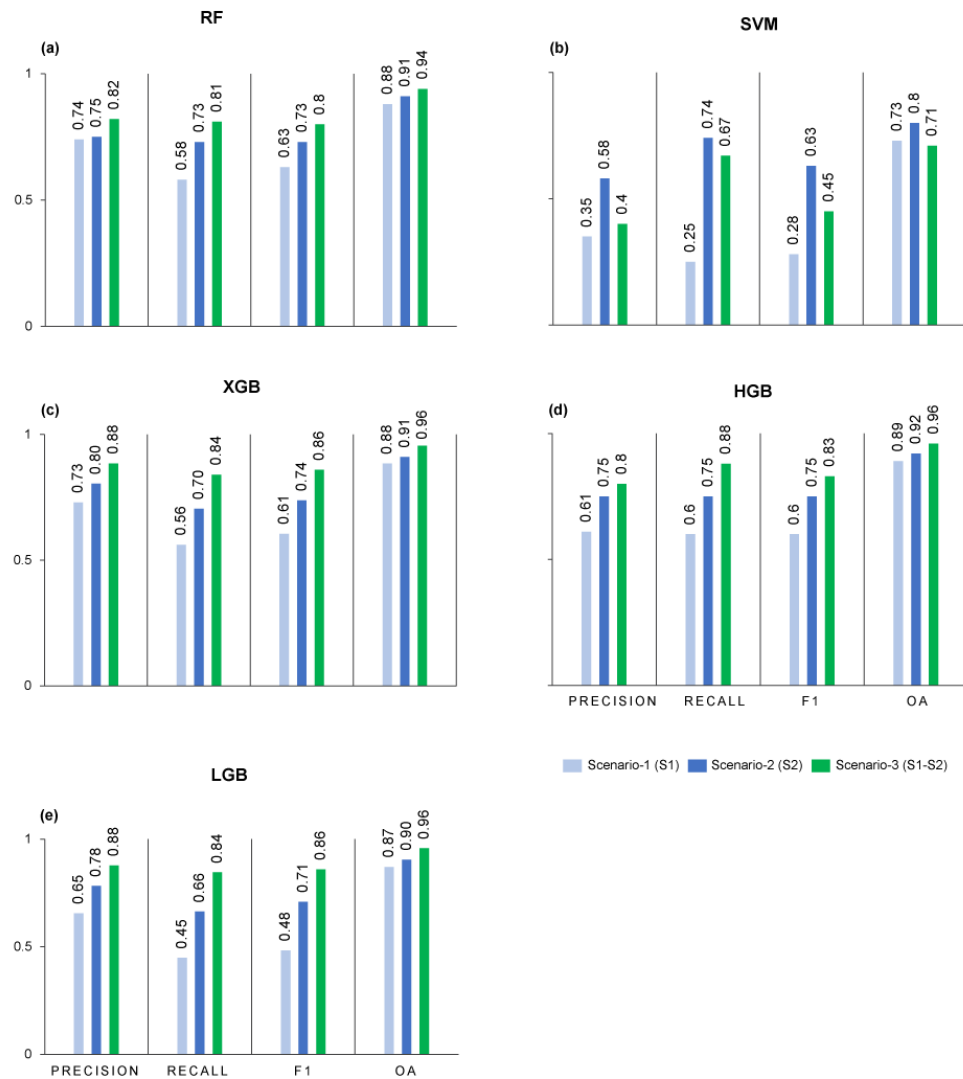
## 4. Results

### 4.1. Crop-Type Mapping Sensitivity to Machine Learning Models and Input Features

Figure 2 shows the models' scores obtained at the validation steps for the scenarios (Scenario-1, Scenario-2, and Scenario-3).

All models provided lower crop-type identification accuracy when using S1 data alone (Scenario-1) compared to using S2 data alone (Scenario-2). An improvement in the OA scores of approximately 3.4%, 9.6%, 3.4%, 3.4%, and 3.4% is observed for RF, SVM, XGB, HGB, and LGB when Scenario-2 is considered in comparison to Scenario-1. The lower statistical scores observed for Scenario-1 are due to the S1 radar sensitivity to surface roughness. In the case of agriculture plots, surface roughness is expected to be larger than the C-band (S1) wavelength (i.e., 5 cm), which leads to the saturation of the signal returned to the satellite. Unlike the C-band, the VIS-NIR spectrum is sensitive to crop composition. Actually, in the VIS-NIR spectral range, the return signal to the satellite is a function of crop nitrogen concentration [6,8], aboveground biomass [6], chlorophyll content [10], fiber content [126], and water content [127]. As these features are specific for each crop type, they allow a better crop discrimination from the S2 features (Scenario 2).

The integration of Sentinel-1 and Sentinel-2 data (Scenario-3) significantly enhanced reliability for most models, achieving overall accuracies (OA) of 0.94 (RF), 0.96 (XGB), 0.96 (HGB), and 0.96 (LGB). This demonstrates the complementary value of the VIS-NIR and radar signals for crop identification, a finding consistent with previous studies carried out in Brazil [33], India [31], Croatia [30], and China [34].



**Figure 2.** OA, F1, Recall, and Precision scores obtained for Scenario-1, Scenario-2, and Scenario-3 using RF (a), SVM (b), XGB (c), HGB (d), and LGB (e).

#### 4.2. Hyperparameter Tuning Sensitivity to Input Features

For every model and scenario, the tuning process yielded values distinct from the default hyperparameter settings (Table 3). Hyperparameter tuning is sensitive to the input feature set, with different hyperparameter combinations observed for each scenario (Scenario-1, Scenario-2, and Scenario-3). The discrepancies observed—(i) between tuned and default hyperparameters, and (ii) across the scenarios—highlight the importance of tuning hyperparameters to establish an optimal model and to ensure stable, reliable predictions. In the RF model, the “max\_depth” value is set to “none” by default, meaning this hyperparameter increases indefinitely. However, the tuning process shows that intermediate values (e.g., 20, 25, and 30) can be adopted to avoid overfitting. In the SVM model, the “kernel” hyperparameter aims to understand the distribution of the data in the hyperplane. Scenario-3 presents a different value (“linear”) than the default (“rbf”) highlighting a tendency toward linearity with an increase of the considered variables. In the XGB model, “col\_sample\_tree”, which shows the proportion of the features used for the model, is set to 1 by default. Its reduction after hyperparameter tuning suggests a possible overfitting, since the randomness of the samples is reduced by removing a small percentage of the predictive variables in each tree. Similarly, the optimization process also tends to reduce “colsample\_bytree” in the LGB model to minimize a potential overfitting

issue. The differences in the “colsample\_bytree” values observed for both models (LGB and XGB) and for all scenarios should be related to their different structure. Finally, in the HGB model, the optimization process used fewer trees as subsets (e.g., “max\_depth”) but allowed for deeper leaf formation (e.g., “max\_leaf\_node”), enabling higher levels of subsets than the default setting. In addition, the “max\_feature” decrease observed for all scenarios (especially Scenario-3) highlight an attempt toward an overfitting reduction.

#### 4.3. Crop-Type Mapping Sensitivity to Feature Selection and Hyperparameter Setup

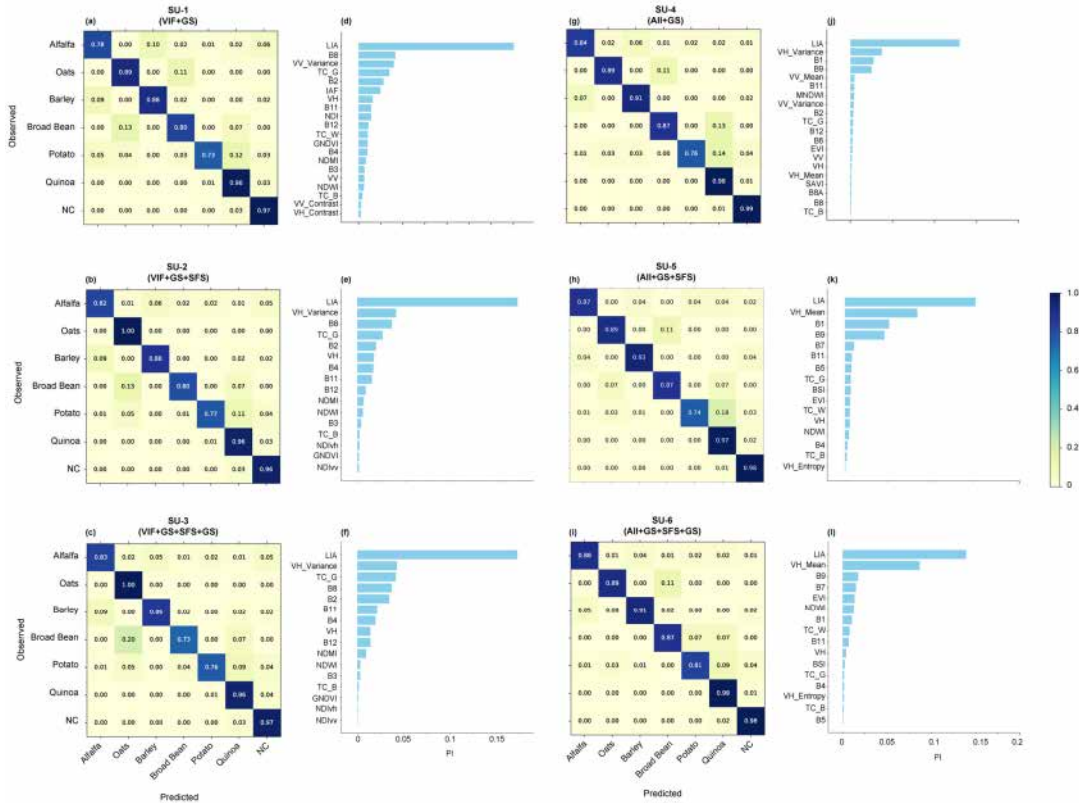
To avoid redundancy, the sensitive analysis is conducted on the XGB and LGB models (Scenario-3 dataset), which are the two most reliable models for the Altiplano crop-type mapping (Figure 2).

Figure 3 shows the confusion matrices obtained and the corresponding features’ PI values for all SUs obtained with the LGB model. A comparison between SU-1 and SU-4 reveals higher Recall scores across all crop classes for SU-4. (Figure 3). This suggests that the VIF may discard (or retain) features that could increase (or decrease) the model’s predictive reliability. Actually, three of the top five variables with the highest PI score in SU-4 (VH\_Variance, B9, and B1) were removed by the VIF filter in SU-1. This apparent adverse effect of the VIF is observed for all SUs where it was applied (SU-1, SU-2, and SU-3). Despite subsequent wrapped feature selection (SFS) and hyperparameter tuning, model reliability remained relatively stable and showed limited improvement across these units and classes. The pattern is further corroborated by results from SU-5, in which four of the top five features (B1, B9, B7, and VH\_Variance) were also removed by the VIF filter in SU-1, even though SU-5 itself utilized the SFS. A comparison between SU-5 and SU-4 shows that overall model reliability remains stable across all classes. This behavior can be explained by the fact that SU-5 uses the tuned hyperparameters obtained with all features as input (SU-4). As indicated in Table 3, hyperparameter tuning is sensitive to the input features. Consequently, the additional hyperparameter tuning performed for SU-6 leads to improved model reliability, most notably for the class with the previously lowest Recall score (potato).

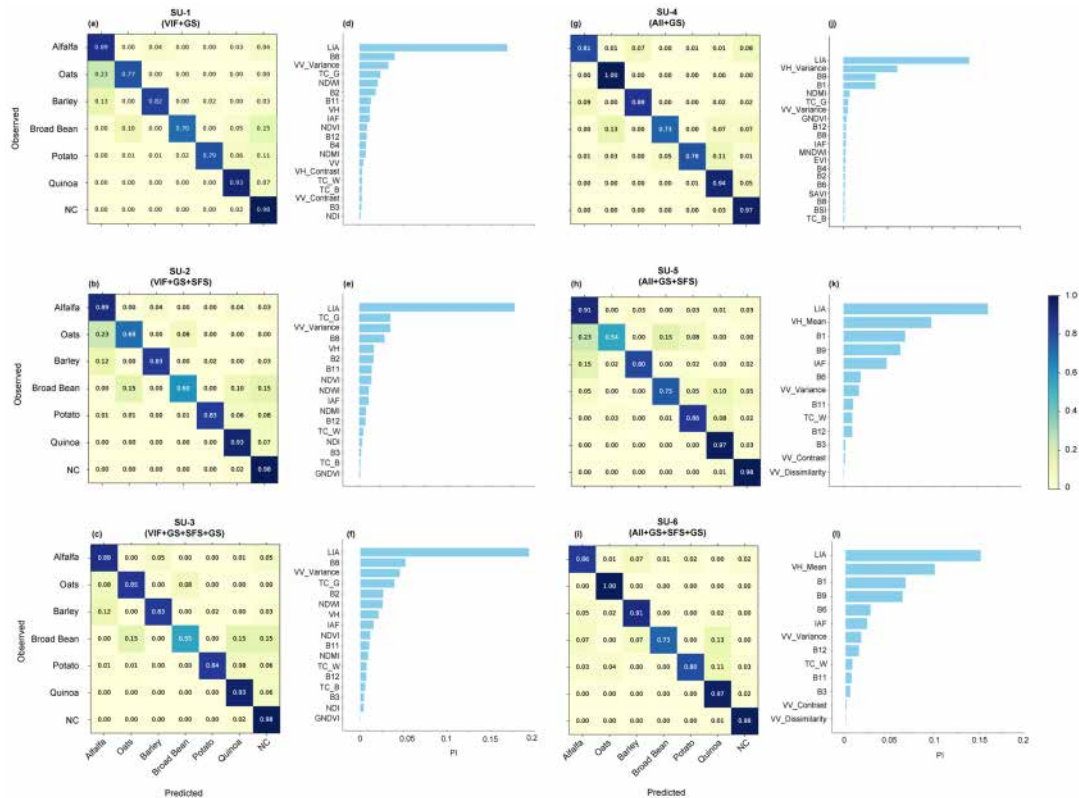
Figure 4 shows the confusion matrices obtained and the corresponding features’ PI values for all SUs obtained with the XGB model.

As observed for the LGB model, the consideration of the VIF discarded some features (B1 and B9) that played a key role in the model’s predictive reliability ( $PI > 0.05$  in SU-4). Consequently, similar higher Recall score values are observed for most of the considered classes in SU-4 than in SU-1 (Figure 4a,g). As observed for the LGB model, additional wrapped feature selection (SU-2) and hyperparameter tuning (SU-3) after the VIF (SU-1) do not improve the XGB model accuracy (Figure 4b,c). This observation confirms the VIF adverse effect on model reliability related to the exclusion (or retention) of features that could increase (or decrease) the model’s predictive reliability. Finally, comparing SU-2 (SU-5) with SU-3 (SU-6), higher Recall score values are obtained for most of the considered classes after hyperparameter tuning (SU-3 and SU-6). This shows that hyperparameter tuning allow the models to fit to the selected features to ensure consistent model outputs.

Despite the benefit of the proposed three-step nested approach, inconsistencies remain in the models prediction with misclassified classes in the confusion matrices. In an agriculture environment with limited material/financial resources (such as the Altiplano region), important crop development heterogeneity is observed in between and inside the agriculture plots. Due to the spectral signature sensitivity to the crop development (water content, nitrogen, fiber, etc.) the spectral signature range associated to the different crop type is large, and similar spectral signatures may have been attributed to the different crop type during model training to contribute to the misclassification of some pixels.



**Figure 3.** Confusion matrix obtained with all considered SUs (a–c,g–i) along with the SU-selected features and corresponding PI values (d–f,j–l) for the LGB model. To improve the graphical representation, only the top 20 features with the highest PI values are shown for SU-1 (d) and SU-4 (j).



**Figure 4.** Confusion matrix obtained with all considered SU (a–c,g–i) along with the SU-selected features and corresponding PI values (d–f,j–l) with the XGB model. To improve the graphical representation, only the top 20 features with the highest PI values are shown for SU-1 (d) and SU-4 (j).

#### 4.4. Computational Costs

Table 4 presents the training process time required for all considered SUs for both the LGB and XGB models on a computing environment with the following features: Operating System—Microsoft Windows 11 Professional (Version 25h2, Build 26200); Processor—Intel Core i7-10610U CPU @ 1.80GHz (2.30 GHz); RAM—32 GB; System Type—64-bit, x64-based architecture.

**Table 4.** Training process time (in hours) to run all SUs for both the LGB and XGB models.

	SU-1 (VIF + GS)	SU-2 (All + GS)	SU-3 (VIF + GS + SFS)	SU-4 (All + GS)	SU-5 (All + GS + SFS)	SU-6 (All + GS + SFS + GS)
LGB	4.1	4.6	8.3	5.7	9.6	10
XGB	0.8	1	1.5	1.3	2.4	2.9

When comparing the LGB and XGB models, the training processing time is more than four times faster for the XGB model and for all considered SUs. The wrapped features selection (SFS) implies significant additional process time than hyperparameter tuning (GS). Actually, an extra processing time of 50% (84.6%) and 80.4% (68.4%) is required to run the SFS on top of SU-3 for the LGB and XGB model, respectively, whereas hyperparameter tuning (on top of SU-5) increases the processing time by only 4.2% and 20.8% for the LGB and XGB model, respectively. Regarding the improvement brought by the three-step nested modeling approaches (SU-6) the additional process time is worth it. Actually, for the LGB (XGB) model, improvements of 12.8% (−3.4%), 0% (Bolfe.3%), 3.4% (10.8%), 8.7% (4.3%), 10.9% (1.3%), 2.1% (4.3%), and 1% (0%) are observed for the alfalfa, oats, barley, broad bean, potato, quinoa, and NC classes when comparing SU-6 with SU-1. It is worth mentioning that the training processing time is only required once before model application. Therefore, despite a longer calibration processing time, the LGB model should be considered for crop-type mapping. Indeed, considering SU-6, the LGB model reached a Recall score superior to 0.8 for all the considered classes and a similar-to-higher Recall score than the XGB model value for six of the seven considered classes (Figures 3i and 4i).

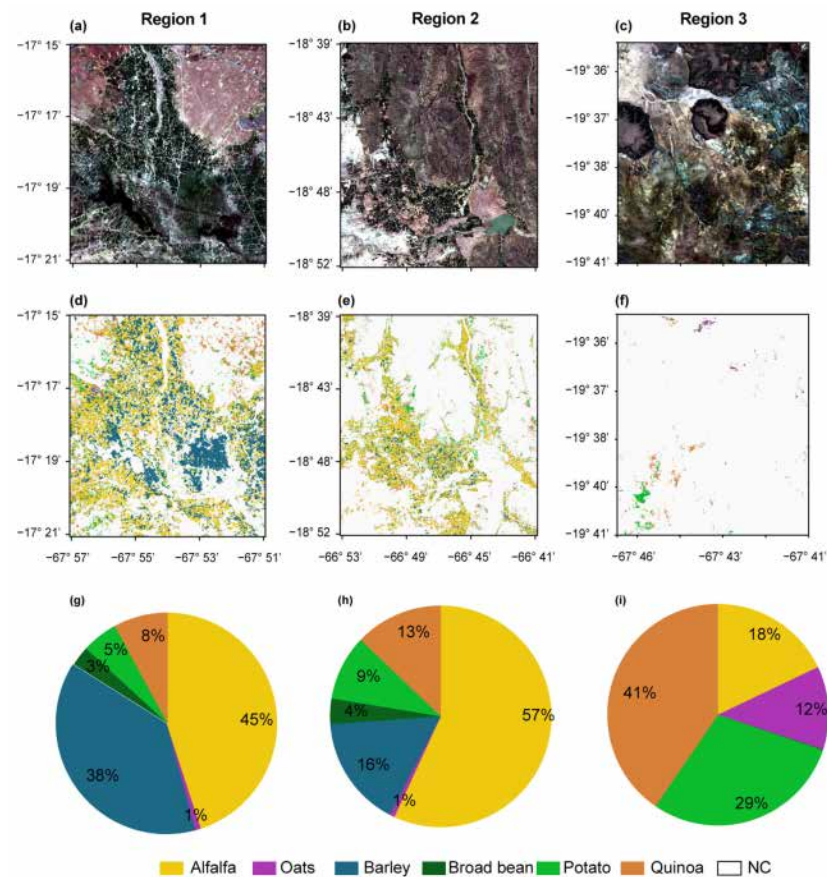
#### 4.5. Crop-Type Mapping and Temporal Analysis

Figure 5 displays the crop-type maps for Region 1, Region 2, and Region 3 (Figure 1) derived from the optimal model–scenario–SU combination (LGB, scenario 3, SU-6).

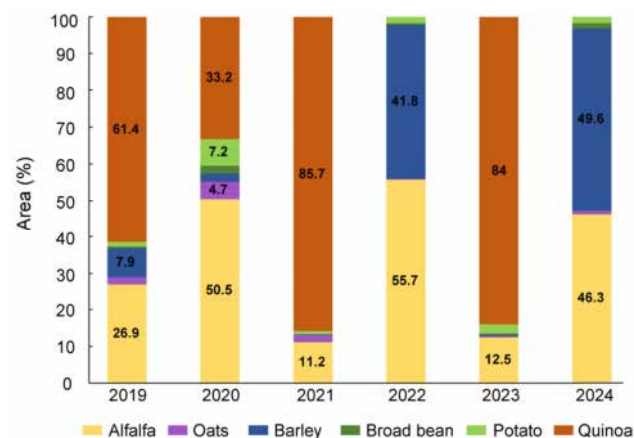
Region 1 is primarily cultivated with alfalfa (45%), barley (38%), and quinoa (8%). While alfalfa is distributed throughout the region, quinoa and barley are primarily grown in the north and south, respectively. Region 2 is also dominated by alfalfa (57%), along with barley (16%) and quinoa (13%). In contrast to Region 1, the crop distribution here is more homogeneous, with no distinct spatial clustering or production hotspots for any single crop-type. Located in the southern Altiplano, Region 3 is a traditional quinoa production hotspot, with quinoa accounting for the largest share of cultivated area (41%). The other primary crops are potato (29%), alfalfa (18%), and oats (12%). Spatially, quinoa cultivation is concentrated in the central part of the region, whereas potatoes and oats are grown primarily in the southwestern and northern areas, respectively. Notably, alfalfa is a significant crop in all three regions, cultivated primarily to support livestock breeding (including cattle, sheep, and llamas).

Figure 6 shows the crop-type surface extent time evolution for the 2019–2024 period in Region 4 (see Figure 1b), as observed by the LGB model with SU-6. In this region, quinoa, alfalfa, and barley account for the main crop-type activities, with their cumulative surface area collectively exceeding 80%. An interesting pattern of crop rotation between quinoa,

alfalfa, and barley emerge, where an increase in quinoa cultivation corresponds to a decrease in alfalfa area, and vice versa. Notably, quinoa was replaced by barley in the 2022 and 2024 growing seasons. Despite the fact that such rotations are a standard agricultural practice in the Altiplano to prevent monoculture and to maintain soil fertility, this observation has to be considered with caution. Actually, the crop-type spectral signature varies along time due to crop development sensitivity to climate conditions (i.e., rainy season onset/offset, heat/cold waves). As the model was calibrated for a specific year (2022) with inherent crop-type development, misclassification among the crop types may occur along the considered time span (2019–2024), introducing uncertainties in the temporal analysis.



**Figure 5.** Regions 1, 2, and 3 (a–c) with corresponding crop-type maps obtained using the LGB model with SU-6 (d–f), along with the surface extent of the detected crop-types (g–i).



**Figure 6.** Crop-type surface extent as detected by the LGB model with Scenario 3 across Region-4 for the 2019–2024 period.

## 5. Discussion

The use of the default hyperparameters for the LULC classification has proven highly effective [21,71,73]. However, the introduction of hyperparameter tuning achieved considerable classification improvements [30,52,74,75]. Similarly, the consideration of feature selection has proven effective to improve model classification [53,59]. In this line, some authors have combined feature selection with hyperparameter tuning to ensure an as reliable as possible modeling output [30,123]. However, these studies did not compare this modeling set-up (hyperparameter tuning and feature selection) with a modeling relying on hyperparameter calibration or features selection alone. In this context, the proposed framework assessed modeling sensitivity to several different hyperparameter tuning and feature selection combinations.

The results reveal a key limitation of the Variance Inflation Factor (VIF) as a filter-based feature selection method. As a model-agnostic process, the VIF assesses multicollinearity without evaluating a model's structural sensitivity to features, either individually or in combination. Consequently, it may discard predictive features or retain irrelevant ones, potentially undermining model performance. In this context, wrapper feature selection methods, such as the SFS, are a better alternative because they iteratively select and evaluate feature subsets during model training, thereby identifying the subset that produces the most accurate predictions. This model-specific approach accounts for feature interactions and the algorithm's unique sensitivities, which explains the significant improvement in model reliability observed when using the wrapped method over the VIF—a finding consistent with prior remote sensing applications [60]. Therefore, despite their higher computational costs, wrapper methods should be recommended to enhance the reliability of crop-type mapping models. To advance this work, a comprehensive comparison of wrapper techniques would provide guidelines to elucidate their specific advantages and limitations (e.g., SFS, GA, and RFE).

Despite S1's radar sensitivity to surface roughness [110], combining S1 with S2 features increases crop-type mapping reliability (Figure 3). Actually, in the most efficient SU (SU-6), the two most influential variables are derived from S1 (i.e., LIA and VH\_Mean). The LIA is sensitive to the topography variation. However, within the relatively flat terrain of the study area, its variation effectively captures differences in aboveground vegetation height. As the studied crops exhibit significant height variation, distinct LIA values can be associated with specific crops. This is in line with a previous study highlighting the correlation between the LIA and the precision of terrestrial object classification [117]. The VH\_Mean derived from the VH-based GLCM (Gray-Level Co-Occurrence Matrix), which quantifies image texture based on the spatial relationships among pixels. Significant differences in vegetation texture are expected among crops due to variations in plant morphology and ground-cover density. Consequently, GLCM indices, such as VH\_Mean, serve as effective discriminators between crop types. In fact, previous studies have shown that the GLCM indices improve image analysis and land mapping [128,129]. Figure 2 shows the benefit of the S1 (i.e., LIA and VH\_Mean) increase in combination with S2. In certain circumstances, a similar S2 spectral signature can be observed for different crops. For these specific situations, aggregating the S1-based crop morphology features (i.e., height and above-crop texture—indirectly approximated by LIA and VH\_Mean) improve the crop-type classification, as the classification is not only based on spectral criteria but on morphologic criteria

Among the crops studied, quinoa achieved the highest Recall score (Figure 3). This can be attributed to its distinct structural and spectral characteristics, which differ markedly from those of other crops. Quinoa grows primarily vertically, with plots forming rows equally spaced (up to 1 m), which leaves much of the soil surface exposed. This unique,

open-row structure of quinoa likely creates a highly contrasting C-band radar backscatter signature compared to the other crops, such as alfalfa, oats, barley, broad beans, and potatoes, which exhibit more balanced vertical and horizontal growth, leading to denser canopy coverage. Spectrally, at maturity, quinoa varieties develop distinctive hues—such as red, orange, or black—that are highly divergent from the senescence colors of other crops. This pronounced spectral difference enhances its identifiability using Sentinel-2's visible-to-near-infrared (VIS-NIR) bands and derived vegetation indices.

It is important to acknowledge that applying crop-type models to areas or periods outside their training conditions—a challenge known as model transferability [130]—introduces uncertainties, primarily due to radiometric variations [16]. In regions with important crop development heterogeneity in between and inside the agriculture plots, similar spectral signatures can be observed for different crops introducing uncertainties in the model learning leading to misclassification occurrences. As these spectral similarities are expected to be spatially limited (i.e., pixel scale), using the average spectral signature derived from the centered pixel and its eight neighboring pixel should decrease crop heterogeneity uncertainties. Crop development heterogeneity is exacerbated along time due to crop development sensitivity to climate variability (i.e., rainy season onset/offset, heat/cold waves). A specific concern for this study is that the learning database was built using data from 2022, a relatively dry year [131]. This likely induced above-average water stress and under-average crop development, meaning the spectral signatures used for training as well as the hyperparameters adapted to the specific domain of the training set may not be fully representative of conditions in wetter years or regions. To mitigate this effect and to enhance model robustness, additional crop plots should be delineated from a broader range of years and geographic areas. This would broaden the spectral and phenological variability in the training data and increase the model's temporal and spatial reliability. Visually interpreting high spatial resolution imagery from UAVs or satellites (e.g., Pleiades) provides a practical method for delineating additional reference fields, thereby improving the spatial distribution of reference observations and increasing the overall number of observations in the learning database.

## 6. Conclusions

This study assesses the sensitivity of machine learning-based crop-type mapping to various input features and modeling setups (feature selection and hyperparameter tuning). The main findings are summarized below:

- All models show that combining Sentinel-1 (S1) and Sentinel-2 (S2) yields more reliable crop-type maps than using either data source alone, highlighting the complementary value of radar and optical data. Hyperparameter tuning (i.e., Grid Search) proved highly valuable, as optimal values consistently differed from model defaults and led to significant performance improvements, underscoring the model's sensitivity to the input feature set.
- In comparison to the VIF selection feature, which does not consider the model sensitivity to the different input features, the wrapped features selection (SFS) consistently improves the model output.
- The most reliable model outputs are achieved using a three-step nested modeling setup, including an initial hyperparameter calibration, followed by a wrapped feature selection (SFS) and a final hyperparameter recalibration on the selected feature subset.
- Among the tested models (RF, SVM, LGB, XGB, HGB), LGB and XGB (SVM) were the most (least) reliable models for crop-type mapping
- From an end-user perspective, the integration of both S1 and S2 in the LGB model with the proposed three-step nested modeling setup enables effective crop monitoring

in space and time, facilitating the identification of dominant crop patterns in specific regions and the analysis of crop rotation practices

While the proposed three-step nested modeling approach achieved high reliability in crop-type mapping, future studies should assess more complex machine learning setups, such as stacking or embedded modeling, to further enhance model reliability and stability. Furthermore, improvements to the learning database is recommended to broaden the range of crop-type spectral variations observed in space and time. Addressing this limitation is a critical prerequisite for reliably deploying such models for operational, regional-scale crop mapping and/or temporal monitoring.

**Author Contributions:** Conceptualization, M.P.-F. and F.S.; methodology, M.P.-F. and F.S.; formal analysis, M.P.-F. and F.S.; investigation, M.P.-F., F.S., J.M.-C., R.H., R.P.-Z., D.T., E.U.-F., L.B., M.-P.B., and C.D.; data curation, M.P.-F., F.S., and R.H.; writing—original draft preparation, M.P.-F. and F.S.; writing—review and editing, M.P.-F., F.S., J.M.-C., R.H., R.P.-Z., D.T., E.U.-F., L.B., M.-P.B., and C.D.; supervision F.S.; project administration, F.S.; funding acquisition, F.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Agropolis Foundation (project number: 2001-032) in the framework of the project WASACA (Wastewater irrigation: a sustainable agriculture adaptation to climate changes over the Bolivian Altiplano?) and by the Centre National d'Études Spatiales (CNES) in the framework of the QUIMONOS project (Quinoa monitoring by satellite) and is a contribution to the ALTIPLANO International Joint Laboratory (LMI ALTIPLANO), supported by a grant from the IRD (Institut de Recherche pour le Développement, France).

**Data Availability Statement:** The original contributions presented in this study are included in the article; further inquiries can be directed to the corresponding author.

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

## References

1. Nizar, E.M.; Wahbi, M.; Ait Kazzi, M.; Yazidi Alaoui, O.; Boulaassal, H.; Maatouk, M.; Zaghoul, M.N.; El Kharki, O. High Resolution Land Cover Mapping and Crop Classification in the Loukkos Watershed (Northern Morocco): An Approach Using SAR Sentinel-1 Time Series. *Rev. Teledetección* **2022**, *60*, 47–69. [[CrossRef](#)]
2. Guo, L.; Zhao, S.; Gao, J.; Zhang, H.; Zou, Y.; Xiao, X. A Novel Workflow for Crop Type Mapping with a Time Series of Synthetic Aperture Radar and Optical Images in the Google Earth Engine. *Remote Sens.* **2022**, *14*, 5458. [[CrossRef](#)]
3. Morell Monzó, S. *Desarrollo de Procedimientos Para La Deteccion Del Abandono de Cultivos de Cítricos Utilizando Técnicas de Teledetección*; Universitat Politècnica de València: Valencia, Spain, 2023.
4. Satgé, F.; Hussain, Y.; Xavier, A.; Zolá, R.P.; Salles, L.; Timouk, F.; Seyler, F.; Garnier, J.; Frappart, F.; Bonnet, M.-P. Unraveling the Impacts of Droughts and Agricultural Intensification on the Altiplano Water Resources. *Agric. For. Meteorol.* **2019**, *279*, 107710. [[CrossRef](#)]
5. Kaur, R.; Tiwari, R.K.; Maini, R.; Singh, S. A Framework for Crop Yield Estimation and Change Detection Using Image Fusion of Microwave and Optical Satellite Dataset. *Quaternary* **2023**, *6*, 28. [[CrossRef](#)]
6. Yin, H.; Li, F.; Yang, H.; Di, Y.; Hu, Y.; Yu, K. Mapping Plant Nitrogen Concentration and Aboveground Biomass of Potato Crops from Sentinel-2 Data Using Ensemble Learning Models. *Remote Sens.* **2024**, *16*, 349. [[CrossRef](#)]
7. De Clerck, E.; D.Kovács, D.; Berger, K.; Schlerf, M.; Verrelst, J. Optimizing Hybrid Models for Canopy Nitrogen Mapping from Sentinel-2 in Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2024**, *218*, 530–545. [[CrossRef](#)]
8. Nino, P.; D'Urso, G.; Vanino, S.; Di Bene, C.; Farina, R.; Falanga Bolognesi, S.; De Michele, C.; Napoli, R. Nitrogen Status of Durum Wheat Derived from Sentinel-2 Satellite Data in Central Italy. *Remote Sens. Appl.* **2024**, *36*, 101323. [[CrossRef](#)]
9. Candiani, G.; Tagliabue, G.; Panigada, C.; Verrelst, J.; Picchi, V.; Caicedo, J.P.R.; Boschetti, M. Evaluation of Hybrid Models to Estimate Chlorophyll and Nitrogen Content of Maize Crops in the Framework of the Future CHIME Mission. *Remote Sens.* **2022**, *14*, 1792. [[CrossRef](#)]
10. Delloye, C.; Weiss, M.; Defourny, P. Retrieval of the Canopy Chlorophyll Content from Sentinel-2 Spectral Bands to Estimate Nitrogen Uptake in Intensive Winter Wheat Cropping Systems. *Remote Sens. Environ.* **2018**, *216*, 245–261. [[CrossRef](#)]
11. Alvarez, H. *Análisis Espacial de La Severidad de Incendios y Su Impacto En El Uso y Cobertura de Suelo En La Provincia de Loja Entre Los Años 2017 y 2020*; Instituto de Altos Estudios Nacionales, Universidad de Posgrado del Estado: Quito, Ecuador, 2021.

12. Ávila-Pérez, I.D.; Ortiz-Malavassi, E.; Soto-Montoya, C.; Vargas-Solano, Y.; Aguilar-Arias, H.; Miller-Granados, C. Evaluación de Cuatro Algoritmos de Clasificación de Imágenes Satelitales Landsat-8 y Sentinel-2 Para La Identificación de Cobertura Boscosa En Paisajes Altamente Fragmentados En Costa Rica. *Rev. Teledetección* **2020**, *57*, 37. [[CrossRef](#)]
13. Quillupangui Nasimba, C.D. *Determinación Del Comportamiento Espectral de Coberturas y Usos de La Tierra de La Subcuenca Del Río San Pedro. Trabajo de titulación previo a la obtención del Título de Ingeniero Ambiental; Carrera de Ingeniería Ambiental; Universidad Central del Ecuador: Quito, Ecuador, 2019.*
14. Ramírez, M.; Martínez, L.; Montilla, M.; Sarmiento, O.; Lasso, J.; Díaz, S. Obtención de Coberturas Del Suelo Agropecuarias En Imágenes Satelitales Sentinel-2 Con La Inyección de Imágenes de Dron Usando Random Forest En Google Earth Engine. *Rev. Teledetección* **2020**, *56*, 49–68. [[CrossRef](#)]
15. Varona, D. Clasificación Supervisada de La Cobertura Terrestre En Fincas Ganaderas. Master's Thesis, Universidad de Córdoba (UCO), Córdoba, Spain, 2022.
16. Wasniewski, A.; Hoscilo, A.; Zagajewski, B.; Moukétou-Tarazewicz, D. Assessment of Sentinel-2 Satellite Images and Random Forest Classifier for Rainforest Mapping in Gabon. *Forests* **2020**, *11*, 941. [[CrossRef](#)]
17. De Peppo, M.; Taramelli, A.; Boschetti, M.; Mantino, A.; Volpi, I.; Filipponi, F.; Tornato, A.; Valentini, E.; Ragagnoli, G. Non-Parametric Statistical Approaches for Leaf Area Index Estimation from Sentinel-2 Data: A Multi-Crop Assessment. *Remote Sens.* **2021**, *13*, 2841. [[CrossRef](#)]
18. Pasqualotto, N.; Delegido, J.; Van Wittenbergh, S.; Rinaldi, M.; Moreno, J. Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). *Sensors* **2019**, *19*, 904. [[CrossRef](#)]
19. Sitokonstantinou, V.; Papoutsis, I.; Kontoes, C.; Arnal, A.L.; Andrés, A.P.A.; Zurbano, J.A.G. Scalable Parcel-Based Crop Identification Scheme Using Sentinel-2 Data Time-Series for the Monitoring of the Common Agricultural Policy. *Remote Sens.* **2018**, *10*, 911. [[CrossRef](#)]
20. Tran, K.H.; Zhang, H.K.; McMaine, J.T.; Zhang, X.; Luo, D. 10 m Crop Type Mapping Using Sentinel-2 Reflectance and 30 m Cropland Data Layer Product. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102692. [[CrossRef](#)]
21. Kumar, R.; Rai, A.; Mishra, V.N.; Diwate, P.; Arya, V.S. Performance Evaluation of Supervised Classifiers for Land Use and Land Cover Mapping Using Sentinel-2 MSI Image. *J. Geosci. Res.* **2021**, *6*, 231–241.
22. Dingle Robertson, L.; McNairn, H.; van der Kooij, M.; Jiao, X.; Ihuoma, S.; Joesse, P. Monitoring Autumn Agriculture Activities Using Synthetic Aperture Radar (SAR) and Coherence Change Detection. *Heliyon* **2023**, *9*, e17322. [[CrossRef](#)]
23. ESA Sentinel Missions. Available online: <https://sentinels.copernicus.eu/web/sentinel/missions> (accessed on 19 June 2022).
24. Dahhani, S.; Raji, M.; Hakdaoui, M.; Lhissou, R. Land Cover Mapping Using Sentinel-1 Time-Series Data and Machine-Learning Classifiers in Agricultural Sub-Saharan Landscape. *Remote Sens.* **2022**, *15*, 65. [[CrossRef](#)]
25. Hütt, C.; Waldhoff, G.; Bareth, G. Fusion of Sentinel-1 with Official Topographic and Cadastral Geodata for Crop-Type Enriched LULC Mapping Using FOSS and Open Data. *ISPRS Int. J. Geoinf.* **2020**, *9*, 120. [[CrossRef](#)]
26. Lima Ramos Barbosa, F.; Fontes Guimarães, R.; Abílio de Carvalho Júnior, O.; Arnaldo Trancoso Gomes, R. Classificação Do Uso e Cobertura Da Terra Utilizando Imagens SAR/Sentinel 1 No Distrito Federal, Brasil. *Soc. Nat.* **2021**, *33*, e55954. [[CrossRef](#)]
27. Nikaein, T.; Iannini, L.; Molijn, R.A.; Lopez-Dekker, P. On the Value of Sentinel-1 InSAR Coherence Time-Series for Vegetation Classification. *Remote Sens.* **2021**, *13*, 3300. [[CrossRef](#)]
28. Pham, L.H.; Pham, L.T.H.; Dang, T.D.; Tran, D.D.; Dinh, T.Q. Application of Sentinel-1 Data in Mapping Land-Use and Land Cover in a Complex Seasonal Landscape: A Case Study in Coastal Area of Vietnamese Mekong Delta. *Geocarto Int.* **2022**, *37*, 3743–3760. [[CrossRef](#)]
29. Prudente, V.H.R.; Sanches, I.D.; Adami, M.; Skakun, S.; Oldoni, L.V.; Xaud, H.A.M.; Xaud, M.R.; Zhang, Y. SAR Data for Land Use Land Cover Classification in a Tropical Region with Frequent Cloud Cover. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Waikoloa, HI, USA, 26 September–2 October 2020; pp. 4100–4103. [[CrossRef](#)]
30. Dobrinčić, D.; Gašparović, M.; Medak, D. Sentinel-1 and 2 Time-Series for Vegetation Mapping Using Random Forest Classification: A Case Study of Northern Croatia. *Remote Sens.* **2021**, *13*, 2321. [[CrossRef](#)]
31. Prasad, P.; Loveson, V.J.; Chandra, P.; Kotha, M. Evaluation and Comparison of the Earth Observing Sensors in Land Cover/Land Use Studies Using Machine Learning Algorithms. *Ecol. Inform.* **2022**, *68*, 101522. [[CrossRef](#)]
32. Mercier, A.; Betbeder, J.; Rumiano, F.; Baudry, J.; Gond, V.; Blanc, L.; Bourgoin, C.; Cornu, G.; Ciudad, C.; Marchamalo, M.; et al. Evaluation of Sentinel-1 and 2 Time Series for Land Cover Classification of Forest–Agriculture Mosaics in Temperate and Tropical Landscapes. *Remote Sens.* **2019**, *11*, 979. [[CrossRef](#)]
33. Tavares, P.A.; Beltrão, N.E.S.; Guimarães, U.S.; Teodoro, A.C. Integration of Sentinel-1 and Sentinel-2 for Classification and LULC Mapping in the Urban Area of Belém, Eastern Brazilian Amazon. *Sensors* **2019**, *19*, 1140. [[CrossRef](#)]
34. Chen, Y.; Hou, J.; Huang, C.; Zhang, Y.; Li, X. Mapping Maize Area in Heterogeneous Agricultural Landscape with Multi-Temporal Sentinel-1 and Sentinel-2 Images Based on Random Forest. *Remote Sens.* **2021**, *13*, 2988. [[CrossRef](#)]
35. Steinhausen, M.J.; Wagner, P.D.; Narasimhan, B.; Waske, B. Combining Sentinel-1 and Sentinel-2 Data for Improved Land Use and Land Cover Mapping of Monsoon Regions. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 595–604. [[CrossRef](#)]

36. Maskell, G.; Chemura, A.; Nguyen, H.; Gornott, C.; Mondal, P. Integration of Sentinel Optical and Radar Data for Mapping Smallholder Coffee Production Systems in Vietnam. *Remote Sens. Environ.* **2021**, *266*, 112709. [[CrossRef](#)]
37. Ge, G.; Shi, Z.; Zhu, Y.; Yang, X.; Hao, Y. Land Use/Cover Classification in an Arid Desert-Oasis Mosaic Landscape of China Using Remote Sensed Imagery: Performance Assessment of Four Machine Learning Algorithms. *Glob. Ecol. Conserv.* **2020**, *22*, e00971. [[CrossRef](#)]
38. Schulz, D.; Yin, H.; Tischbein, B.; Verleysdonk, S.; Adamou, R.; Kumar, N. Land Use Mapping Using Sentinel-1 and Sentinel-2 Time Series in a Heterogeneous Landscape in Niger, Sahel. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 97–111. [[CrossRef](#)]
39. Talukdar, S.; Singha, P.; Mahato, S.; Pal, S.; Liou, Y.-A.; Rahman, A. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sens.* **2020**, *12*, 1135. [[CrossRef](#)]
40. Khuc, T.D.; Luong, N.D.; Dang, D.H.; Tran, V.A. Comparison of Random Forest and Extreme Gradient Boosting Algorithms in Land Cover Classification in Van Yen District, Yen Bai Province, Vietnam. *J. Hydrometeorol.* **2025**, *6*, 50–59. [[CrossRef](#)]
41. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Beier, C.M.; Johnson, L.; Phoenix, D.B. A Comparison of Random Forest and Light Gradient Boosting Machine for Forest Above-Ground Biomass Estimation Using a Combination of Landsat, Alos Palsar, and Airborne LiDAR Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.-ISPRS Arch.* **2021**, *44*, 163–168. [[CrossRef](#)]
42. Huang, Z.; Wu, W.; Liu, H.; Zhang, W.; Hu, J. Identifying Dynamic Changes in Water Surface Using Sentinel-1 Data Based on Genetic Algorithm and Machine Learning Techniques. *Remote Sens.* **2021**, *13*, 3745. [[CrossRef](#)]
43. Sibindi, R.; Mwangi, R.W.; Waititu, A.G. A Boosting Ensemble Learning Based Hybrid Light Gradient Boosting Machine and Extreme Gradient Boosting Model for Predicting House Prices. *Eng. Rep.* **2023**, *5*, e12599. [[CrossRef](#)]
44. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [[CrossRef](#)]
45. Mhangara, P.; Gidey, E.; Mayise, B.S. Using Extreme Gradient Boosting for Predictive Urban Expansion Analysis in Rustenburg, South Africa from 2000 to 2030. *Sci. Rep.* **2025**, *15*, 19050. [[CrossRef](#)]
46. Magidi, J.; Bangira, T.; Kelepile, M.; Shoko, M. Land Use and Land Cover Changes in Notwane Watershed, Botswana, Using Extreme Gradient Boost (XGBoost) Machine Learning Algorithm. *Afr. Geogr. Rev.* **2025**, *44*, 497–517. [[CrossRef](#)]
47. Matyukira, C.; Mhangara, P. Land Cover and Landscape Structural Changes Using Extreme Gradient Boosting Random Forest and Fragmentation Analysis. *Remote Sens.* **2023**, *15*, 5520. [[CrossRef](#)]
48. McCarty, D.A.; Kim, H.W.; Lee, H.K. Evaluation of Light Gradient Boosted Machine Learning Technique in Large Scale Land Use and Land Cover Classification. *Environments* **2020**, *7*, 84. [[CrossRef](#)]
49. Candido, C.; Blanco, A.C.; Medina, J.; Gubatanga, E.; Santos, A.; Ana, R.S.; Reyes, R.B. Improving the Consistency of Multi-Temporal Land Cover Mapping of Laguna Lake Watershed Using Light Gradient Boosting Machine (LightGBM) Approach, Change Detection Analysis, and Markov Chain. *Remote Sens. Appl.* **2021**, *23*, 100565. [[CrossRef](#)]
50. Li, R.; Gao, X.; Shi, F. A Framework for Subregion Ensemble Learning Mapping of Land Use/Land Cover at the Watershed Scale. *Remote Sens.* **2024**, *16*, 3855. [[CrossRef](#)]
51. Bolfe, É.L.; Parreiras, T.C.; Silva, L.A.P.D.; Sano, E.E.; Bettiol, G.M.; Victoria, D.D.C.; Sanches, I.D.; Vicente, L.E. Mapping Agricultural Intensification in the Brazilian Savanna: A Machine Learning Approach Using Harmonized Data from Landsat Sentinel-2. *ISPRS Int. J. Geoinf.* **2023**, *12*, 263. [[CrossRef](#)]
52. Zhang, W.; Liu, H.; Wu, W.; Zhan, L.; Wei, J. Mapping Rice Paddy Based on Machine Learning with Sentinel-2 Multi-Temporal Data: Model Comparison and Transferability. *Remote Sens.* **2020**, *12*, 1620. [[CrossRef](#)]
53. Aggrawal, R.; Pal, S. Sequential Feature Selection and Machine Learning Algorithm-Based Patient's Death Events Prediction and Diagnosis in Heart Disease. *SN Comput. Sci.* **2020**, *1*, 344. [[CrossRef](#)]
54. Mudereri, B.T.; Abdel-Rahman, E.M.; Ndlela, S.; Makumbe, L.D.M.; Nyanga, C.C.; Tonnang, H.E.Z.; Mohamed, S.A. Integrating the Strength of Multi-Date Sentinel-1 and-2 Datasets for Detecting Mango (*Mangifera Indica* L.) Orchards in a Semi-Arid Environment in Zimbabwe. *Sustainability* **2022**, *14*, 5741. [[CrossRef](#)]
55. Guo, X.; Bian, Z.; Wang, S.; Wang, Q.; Zhang, Y.; Zhou, J.; Lin, L. Prediction of the Spatial Distribution of Soil Arthropods Using a Random Forest Model: A Case Study in Changtu County, Northeast China. *Agric. Ecosyst. Environ.* **2020**, *292*, 106818. [[CrossRef](#)]
56. Alibrahim, H.; Ludwig, S.A. Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization. In Proceedings of the 2021 IEEE Congress on Evolutionary Computation (CEC), Kraków, Poland, 28 June–1 July 2021; pp. 1551–1559.
57. Azedou, A.; Amine, A.; Kisekka, I.; Lahssini, S. Genetic Algorithm Optimization of Ensemble Learning Approach for Improved Land Cover and Land Use Mapping: Application to Talassemtane National Park. *Ecol. Indic.* **2025**, *177*, 113776. [[CrossRef](#)]
58. Haji Mohammadi, M.; Nazari Samani, A.; Keshtkar, H.; Zare Garizi, A.; Arabkhedri, M.; Shafaie, V.; Movahedi Rad, M. Incorporating Climate and Land Use Projections with Spatial Optimization of Best Management Practices for Soil Erosion and Sediment Control in a Semi-Arid Mountainous Watershed. *Sci. Total Environ.* **2025**, *1008*, 180993. [[CrossRef](#)] [[PubMed](#)]

59. Sirpa-Poma, J.W.; Satgé, F.; Resongles, E.; Pillco-Zolá, R.; Molina-Carpio, J.; Flores Colque, M.G.; Ormachea, M.; Pacheco Mollinedo, P.; Bonnet, M.-P. Towards the Improvement of Soil Salinity Mapping in a Data-Scarce Context Using Sentinel-2 Images in Machine-Learning Models. *Sensors* **2023**, *23*, 9328. [[CrossRef](#)] [[PubMed](#)]
60. Wang, Y.; Li, Y. Mapping the Ratoon Rice Suitability Region in China Using Random Forest and Recursive Feature Elimination Modeling. *Field Crops Res.* **2023**, *301*, 109016. [[CrossRef](#)]
61. Ma, Y.; Ma, Y.; Zheng, Q.; Chen, Q. Dynamic Co-Optimization of Features and Hyperparameters in Object-Oriented Ensemble Methods for Wetland Mapping Using Sentinel-1/2 Data. *Water* **2025**, *17*, 2877. [[CrossRef](#)]
62. Bilgili, A.; Arda, T.; Kilic, B.; Uzar, M. A Machine Learning-Driven Approach for Automated Landfill Site Selection: An Experimental Study on Marmara Region, Türkiye. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *XLVIII-M-6*, 73–78. [[CrossRef](#)]
63. Duan, X.; Wu, X.; Ge, J.; Deng, L.; Shen, L.; Xu, J.; Xu, X.; He, Q.; Chen, Y.; Gao, X.; et al. A Novel Hierarchical Clustering Sequential Forward Feature Selection Method for Paddy Rice Agriculture Mapping Based on Time-Series Images. *Agriculture* **2024**, *14*, 1468. [[CrossRef](#)]
64. Shafiee, S.; Lied, L.M.; Burud, I.; Dieseth, J.A.; Alsheikh, M.; Lillemo, M. Sequential Forward Selection and Support Vector Regression in Comparison to LASSO Regression for Spring Wheat Yield Prediction Based on UAV Imagery. *Comput. Electron. Agric.* **2021**, *183*, 106036. [[CrossRef](#)]
65. Hamid, A.; Miloš, R. Ensemble Machine Learning Models for Monitoring Riparian Vegetation Dynamics Using Historical Aerial Orthophotos. *Remote Sens. Appl.* **2025**, *38*, 101545. [[CrossRef](#)]
66. Patel, D.; Saxena, A.; Wang, J. A Machine Learning-Based Wrapper Method for Feature Selection. *Int. J. Data Warehous. Min.* **2024**, *20*, 1–33. [[CrossRef](#)]
67. Nnamoko, N.; Arshad, F.; England, D.; Vora, J.; Norman, J. Evaluation of Filter and Wrapper Methods for Feature Selection in Supervised Machine Learning. In Proceedings of the 15th Annual Postgraduate Symposium on the convergence of Telecommunication, Networking and Broadcasting, Liverpool, UK, 23–24 June 2014; pp. 63–67.
68. Liyew, C.M.; Ferraris, S.; Di Nardo, E.; Meo, R. A Review of Feature Selection Methods for Actual Evapotranspiration Prediction. *Artif. Intell. Rev.* **2025**, *58*, 292. [[CrossRef](#)]
69. Canero, F.M.; Rodriguez-Galiano, V.; Aragones, D. Machine Learning and Feature Selection for Soil Spectroscopy. An Evaluation of Random Forest Wrappers to Predict Soil Organic Matter, Clay, and Carbonates. *Heliyon* **2024**, *10*, e30228. [[CrossRef](#)] [[PubMed](#)]
70. Chandrashekar, G.; Sahin, F. A Survey on Feature Selection Methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [[CrossRef](#)]
71. Gyamfi-Ampadu, E.; Gebreslasie, M.; Mendoza-Ponce, A. Mapping Natural Forest Cover Using Satellite Imagery of Nkandla Forest Reserve, KwaZulu-Natal, South Africa. *Remote Sens. Appl.* **2020**, *18*, 100302. [[CrossRef](#)]
72. Tufail, R.; Ahmad, A.; Javed, M.A.; Ahmad, S.R. A Machine Learning Approach for Accurate Crop Type Mapping Using Combined SAR and Optical Time Series Data. *Adv. Space Res.* **2022**, *69*, 331–346. [[CrossRef](#)]
73. Yang, Q.; Wang, L.; Huang, J.; Lu, L.; Li, Y.; Du, Y.; Ling, F. Mapping Plant Diversity Based on Combined SENTINEL-1/2 Data—Opportunities for Subtropical Mountainous Forests. *Remote Sens.* **2022**, *14*, 492. [[CrossRef](#)]
74. Yingisani, C.; Elhadi, A.; Khalid Adem, A. Exploring the Effect of Balanced and Imbalanced Multi-Class Distribution Data and Sampling Techniques on Fruit-Tree Crop Classification Using Different Machine Learning Classifiers. *Geomatics* **2023**, *3*, 70–92. [[CrossRef](#)]
75. Uscamayta-Ferrano, E.; Satgé, F.; Pillco-Zolá, R.; Roig, H.; Tola-Aguilar, D.; Perez-Flores, M.; Bustillos, L.; Rakotomandrindra, F.P.M.; Rabefitia, Z.; Carrière, S.D. CHIRTS Gridded Air Temperature Downscaling Integrating MODIS Land Surface Temperature Estimates in Machine-Learning Models. *Atmosphere* **2025**, *16*, 1188. [[CrossRef](#)]
76. Schratz, P.; Muenchow, J.; Iturrutxa, E.; Richter, J.; Brenning, A. Hyperparameter Tuning and Performance Assessment of Statistical and Machine-Learning Algorithms Using Spatial Data. *Ecol. Modell.* **2019**, *406*, 109–120. [[CrossRef](#)]
77. Hossain, M.R.; Timmer, D. Douglas Timmer Machine Learning Model Optimization with Hyper Parameter Tuning Approach. *Glob. J. Comput. Sci. Technol.* **2021**, *21*, 31.
78. Quino, I.; Quintanilla, J. Índice de Calidad Del Agua En La Cuenca Del Lago Poopó -Uru Uru Aplicando Herramientas Sig. *Rev. Boliv. Química* **2013**, *30*, 91–101.
79. Zubieta, R.; Molina-Carpio, J.; Laqui, W.; Sulca, J.; Ilbay, M. Comparative Analysis of Climate Change Impacts on Meteorological, Hydrological, and Agricultural Droughts in the Lake Titicaca Basin. *Water* **2021**, *13*, 175. [[CrossRef](#)]
80. MDPyE. Oruro Atlas de Potencialidades Productivas Del Estado Plurinacional de Bolivia 2009. In *Oruro Atlas de Potencialidades*; MDPyE, Ed.; GIZ, Cooperacion Alemana: La Paz, Bolivia, 2009; p. 43.
81. Quezada, C. *Adaptación a Los Impactos Del Cambio Climático de Sistemas Agrícolas Basados En Papa Del Altiplano Boliviano*; Proyecto de grado, Universidad Católica Boliviana: Cochabamba, Bolivia, 2021.
82. Quispe Quispe, M.; Quispe, J.; Mena Herrera, C.; Chipana Rivera, R.; Chipana Mendoza, G.J. Caracterización Socioeconómica de La Producción Agrícola de Las Familias Que Habitan La Microcuenca Mamaniri, Altiplano Boliviano. *Rev. Investig. Innovación Agropecu. Recur. Nat.* **2018**, *5*, 125–132.

83. Tapia, N. *Agroecología y Agricultura Campesina Sostenible En Los Andes Bolivianos*; Plural Editores: Maputo, Mozambique, 2006; Volume 4, ISBN 9990564620.
84. Del Barco-Gamarra, M.T.; Foladori, G.; Soto-Esquivel, R. Insustentabilidad de La Producción de Quinoa En Bolivia. *Estud. Sociales. Rev. Aliment. Contemp. Desarro. Reg.* **2019**, *29*, 54. [CrossRef]
85. DAPRO. Informe Productivo Del Departamento de Oruro; Oruro, 2021. Available online: <https://siip.produccion.gob.bo/noticias/files/2021-80cb0-Oruro.pdf> (accessed on 1 January 2023).
86. Satgé, F.; Espinoza, R.; Zolá, R.; Roig, H.; Timouk, F.; Molina, J.; Garnier, J.; Calmant, S.; Seyler, F.; Bonnet, M.-P. Role of Climate Variability and Human Activity on Poopó Lake Droughts between 1990 and 2015 Assessed Using Remote Sensing Data. *Remote Sens.* **2017**, *9*, 218. [CrossRef]
87. Caballero, G.R.; Platzeck, G.; Pezzola, A.; Casella, A.; Winschel, C.; Silva, S.S.; Ludueña, E.; Pasqualotto, N.; Delegido, J. Assessment of Multi-Date Sentinel-1 Polarizations and GLCM Texture Features Capacity for Onion and Sunflower Classification in an Irrigated Valley: An Object Level Approach. *Agronomy* **2020**, *10*, 845. [CrossRef]
88. Hans, R.P. Las Actividades Agrícolas y Sus Posibilidades. *Rev. Cienc. Cult.* **2005**, *21*, 21–37.
89. Jesus, J. Fecha de Siembra de La Alfalfa. Available online: <https://esseed.com/blog/fecha-siembra-alfalfa/> (accessed on 30 September 2025).
90. Padilla, J.E.R. Composición Nutricional. *Univ. Técnica Ambato Fac. Cienc. Ing. Aliment. Biotecnol.* **2022**, *33*, 1–12.
91. Pereira, J.; Saraiva, F. A Comparative Analysis of Unbalanced Data Handling Techniques for Machine Learning Algorithms to Electricity Theft Detection. 2020 IEEE Congress on Evolutionary Computation, CEC 2020—Conference Proceedings, Glasgow, UK, 19–24 July 2020. [CrossRef]
92. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
93. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 3600. [CrossRef]
94. Gacto, M.J.; Soto-Hidalgo, J.M.; Alcalá-Fdez, J.; Alcalá, R. Experimental Study on 164 Algorithms Available in Software Tools for Solving Standard Non-Linear Regression Problems. *IEEE Access* **2019**, *7*, 108916–108939. [CrossRef]
95. Thanh Noi, P.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* **2017**, *18*, 18. [CrossRef]
96. Ishwaran, H.; Lu, M. Standard Errors and Confidence Intervals for Variable Importance in Random Forest Regression, Classification, and Survival. *Stat. Med.* **2019**, *38*, 558–582. [CrossRef] [PubMed]
97. Ramezan, C.A.; Warner, T.A.; Maxwell, A.E.; Price, B.S. Effects of Training Set Size on Supervised Machine-Learning Land-Cover Classification of Large-Area High-Resolution Remotely Sensed Data. *Remote Sens.* **2021**, *13*, 368. [CrossRef]
98. Mushtaq, F.; Mahmood, K.; Chaudhry, M.H.; Tufail, R. A Comparative Study of Support Vector Machine and Maximum Likelihood Classification to Extract Land Cover of Lahore District, Punjab, Pakistan. *Pak. J. Sci. Ind. Res. Ser. A Phys. Sci.* **2021**, *64*, 265–274. [CrossRef]
99. Loukika, K.N.; Keesara, V.R.; Sridhar, V. Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India. *Sustainability* **2021**, *13*, 13758. [CrossRef]
100. Rao, P.; Zhou, W.; Bhattarai, N.; Srivastava, A.K.; Singh, B.; Poonia, S.; Lobell, D.B.; Jain, M. Using Sentinel-1, Sentinel-2, and Planet Imagery to Map Crop Type of Smallholder Farms. *Remote Sens.* **2021**, *13*, 1870. [CrossRef]
101. Bischl, B.; Binder, M.; Lang, M.; Pielok, T.; Richter, J.; Coors, S.; Thomas, J.; Ullmann, T.; Becker, M.; Boulesteix, A.L.; et al. Hyperparameter Optimization: Foundations, Algorithms, Best Practices, and Open Challenges. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2023**, *13*, e1484. [CrossRef]
102. Şimşek, F.F. Comparison of Agricultural Crop Type Classifications with Different Machine Learning Algorithms by Generating Ground Truth Data from Farmer Declaration Parcels. *Int. J. Eng. Geosci.* **2025**, *10*, 207–220. [CrossRef]
103. Tian, Z.; Wei, J.; Li, Z. How Important Is Satellite-Retrieved Aerosol Optical Depth in Deriving Surface PM<sub>2.5</sub> Using Machine Learning? *Remote Sens.* **2023**, *15*, 3780. [CrossRef]
104. Tamim Kashifi, M.; Ahmad, I. Efficient Histogram-Based Gradient Boosting Approach for Accident Severity Prediction With Multisource Data. *Transp. Res. Rec. J. Transp. Res. Board* **2022**, *2676*, 236–258. [CrossRef]
105. Maftoun, M.; Shadkam, N.; Somayeh, S.; Komamardakhi, S.; Mansor, Z.; Hassannataj Joloudari, J. Malicious URL Detection Using Optimized Hist Gradient Boosting Classifier Based on Grid Search Method. *arXiv* **2024**, arXiv:2406.10286.
106. Aljamaan, H.; Alazba, A. Software Defect Prediction Using Tree-Based Ensembles. In *PROMISE 2020: Proceedings of the 16th ACM International Conference on Predictive Models and Data Analytics in Software Engineering; Co-located with ESEC/FSE 2020*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–10. [CrossRef]
107. Xia, L.; Zheng, P.; Li, J.; Huang, X.; Gao, R.X. Histogram-Based Gradient Boosting Tree: A Federated Learning Approach for Collaborative Fault Diagnosis. *IEEE/ASME Trans. Mechatron.* **2024**, *29*, 2637–2648. [CrossRef]
108. Gudmann, A.; Mucsí, L. Pixel and Object-Based Land Cover Mapping and Change Detection from 1986 to 2020 for Hungary Using Histogram-Based Gradient Boosting Classification Tree Classifier. *Geogr. Pannonica* **2022**, *26*, 165–175. [CrossRef]

109. Díaz-Pacheco, J.; van Delden, H.; Hewitt, R. *The Importance of Scale in Land Use Models: Experiments in Data Conversion, Data Resampling, Resolution and Neighborhood Extent*; Springer International Publishing: Cham, Switzerland, 2018; pp. 163–186.
110. Tola, D.; Satgé, F.; Pillco Zolá, R.; Sainz, H.; Condori, B.; Miranda, R.; Yujra, E.; Molina-Carpio, J.; Hostache, R.; Espinoza-Villar, R. Soil Salinity Mapping of Plowed Agriculture Lands Combining Radar Sentinel-1 and Optical Sentinel-2 with Topographic Data in Machine Learning Models. *Remote Sens.* **2024**, *16*, 3456. [CrossRef]
111. SNIA EVI. Available online: [http://dlibrary.snia.gub.uy/maproom/Monitoreo\\_Agroclimatico/INDICES\\_VEGETACION/EVI/index.html](http://dlibrary.snia.gub.uy/maproom/Monitoreo_Agroclimatico/INDICES_VEGETACION/EVI/index.html) (accessed on 1 January 2023).
112. Castelo-Cabay, M.; Piedra-Fernandez, J.A.; Ayala, R. Deep Learning for Land Use and Land Cover Classification from the Ecuadorian Paramo. *Int. J. Digit. Earth* **2022**, *15*, 1001–1017. [CrossRef]
113. Solís-Silvan, R.; Sanchez-Gutiérrez, F.; Islas-Jesús, R.E.; Gerónimo-Torres, J.D.C.; Pozo-Santiago, C.O.; Sanchez-Díaz, B. Estimation of the Leaf Area Index from Sentinel Images in Eucalyptus Grandis W. Hill Plantations. *Rev. Tecnol. Marcha* **2022**, *35*, 39–47. [CrossRef]
114. Holtgrave, A.-K.; Röder, N.; Ackermann, A.; Erasmi, S.; Kleinschmit, B. Comparing Sentinel-1 and -2 Data and Indices for Agricultural Land Use Monitoring. *Remote Sens.* **2020**, *12*, 2919. [CrossRef]
115. Ghasemi, M.; Karimzadeh, S.; Feizizadeh, B. Urban Classification Using Preserved Information of High Dimensional Textural Features of Sentinel-1 Images in Tabriz, Iran. *Earth Sci. Inform.* **2021**, *14*, 1745–1762. [CrossRef]
116. Mastrorosa, S.; Crespi, M.; Congedo, L.; Munafò, M. Land Consumption Classification Using Sentinel 1 Data: A Systematic Review. *Land* **2023**, *12*, 932. [CrossRef]
117. O’Grady, D.; Leblanc, M.; Gillieson, D. Relationship of Local Incidence Angle with Satellite Radar Backscatter for Different Surface Conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *24*, 42–53. [CrossRef]
118. Zhou, S.; Xu, L.; Chen, N. Rice Yield Prediction in Hubei Province Based on Deep Learning and the Effect of Spatial Heterogeneity. *Remote Sens.* **2023**, *15*, 1361. [CrossRef]
119. Auravant Qué Es El Índice GNDVI. Available online: [https://www.auravant.com/ayuda-es/imagenes-indices-y-capas/3636624-que-es-el-indice-gndvi/#:text=El Índice GNDVI \(Vegetación de, en el dosel del cultivo](https://www.auravant.com/ayuda-es/imagenes-indices-y-capas/3636624-que-es-el-indice-gndvi/#:text=El Índice GNDVI (Vegetación de, en el dosel del cultivo) (accessed on 19 February 2023).
120. Du, Y.; Zhang, Y.; Ling, F.; Wang, Q.; Li, W.; Li, X. Water Bodies’ Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the Swir Band. *Remote Sens.* **2016**, *8*, 354. [CrossRef]
121. Mladenova, I.E.; Jackson, T.J.; Bindlish, R.; Hensley, S. Incidence Angle Normalization of Radar Backscatter Data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1791–1804. [CrossRef]
122. Tsyganskaya, V.; Martinis, S.; Marzahn, P.; Ludwig, R. Detection of Temporary Flooded Vegetation Using Sentinel-1 Time Series Data. *Remote Sens.* **2018**, *10*, 1286. [CrossRef]
123. Bandak, S.; Movahedi Naeini, S.A.R.; Komaki, C.B.; Verrelst, J.; Kakooei, M.; Mahmoodi, M.A. Satellite-Based Estimation of Soil Moisture Content in Croplands: A Case Study in Golestan Province, North of Iran. *Remote Sens.* **2023**, *15*, 2155. [CrossRef]
124. Statistical, M.; Learning, M. Over Fitting, Model Tuning, and Evaluation of Prediction Performance. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer: Cham, Switzerland, 2022; ISBN 9783030890100.
125. Taghadosi, M.M.; Hasanlou, M.; Eftekhari, K. Retrieval of Soil Salinity from Sentinel-2 Multispectral Imagery. *Eur. J. Remote Sens.* **2019**, *52*, 138–154. [CrossRef]
126. Fernandes, M.H.M.D.R.; Fernandes Junior, J.D.S.; Adams, J.M.; Lee, M.; Reis, R.A.; Tedeschi, L.O. Using Sentinel-2 Satellite Images and Machine Learning Algorithms to Predict Tropical Pasture Forage Mass, Crude Protein, and Fiber Content. *Sci. Rep.* **2024**, *14*, 8704. [CrossRef]
127. Boren, E.J.; Boschetti, L. Landsat-8 and Sentinel-2 Canopy Water Content Estimation in Croplands through Radiative Transfer Model Inversion. *Remote Sens.* **2020**, *12*, 2803. [CrossRef]
128. Tanase, M.A.; Mihai, M.C.; Miguel, S.; Cantero, A.; Tijerin, J.; Ruiz-Benito, P.; Domingo, D.; Garcia-Martin, A.; Aponte, C.; Lamelas, M.T. Long-Term Annual Estimation of Forest above Ground Biomass, Canopy Cover, and Height from Airborne and Spaceborne Sensors Synergies in the Iberian Peninsula. *Environ. Res.* **2024**, *259*, 119432. [CrossRef]
129. Balling, J.; Herold, M.; Reiche, J. How Textural Features Can Improve SAR-Based Tropical Forest Disturbance Mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103492. [CrossRef]
130. Barman, P.; Mustak, S.; Kuffer, M.; Singh, S.K. Transfer-Ensemble Learning: A Novel Approach for Mapping Urban Land Use/Cover of the Indian Metropolitans. *Sustainability* **2023**, *15*, 16593. [CrossRef]
131. Gutierrez-Villarreal, R.A.; Espinoza, J.C.; Lavado-Casimiro, W.; Junquas, C.; Molina-Carpio, J.; Condom, T.; Marengo, J.A. The 2022-23 Drought in the South American Altiplano: ENSO Effects on Moisture Flux in the Western Amazon during the Pre-Wet Season. *Weather Clim. Extrem.* **2024**, *45*, 100710. [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.