



Article Comparative Evaluation of Remote Sensing Platforms for Almond Yield Prediction

Nathalie Guimarães ^{1,2,3}, Helder Fraga ^{1,2,*}, Joaquim J. Sousa ^{3,4}, Luís Pádua ^{1,2,3}, Albino Bento ⁵ and Pedro Couto ^{1,2,3}

- ¹ Centre for the Research and Technology of Agro-Environmental and Biological Sciences (CITAB), University of Trás-os-Montes e Alto Douro (UTAD), 5000-801 Vila Real, Portugal; nsguimaraes@utad.pt (N.G.); pcouto@utad.pt (P.C.)
- ² Institute for Innovation, Capacity Building and Sustainability of Agri-Food Production, University of Trás-os-Montes e Alto Douro (UTAD), 5000-801 Vila Real, Portugal
- ³ Engineering Department, University of Trás-os-Montes e Alto Douro (UTAD), 5000-801 Vila Real, Portugal; jjsousa@utad.pt
- ⁴ Centre for Robotics in Industry and Intelligent Systems (CRIIS), INESC-TEC, 4200-465 Porto, Portugal
- ⁵ Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal; bento@ipb.pt
- * Correspondence: hfraga@utad.pt

Abstract: Almonds are becoming a central element in the gastronomic and food industry worldwide. Over the last few years, almond production has increased globally. Portugal has become the third most important producer in Europe, where this increasing trend is particularly evident. However, the susceptibility of almond trees to changing climatic conditions presents substantial risks, encompassing yield reduction and quality deterioration. Hence, yield forecasts become crucial for mitigating potential losses and aiding decisionmakers within the agri-food sector. Recent technological advancements and new data analysis techniques have led to the development of more suitable methods to model crop yields. Herein, an innovative approach to predict almond yields in the Trás-os-Montes region of Portugal was developed, by using machine learning regression models (i.e., the random forest regressor, XGBRegressor, gradient boosting regressor, bagging regressor, and AdaBoost regressor), coupled with remote sensing data obtained from different satellite platforms. Satellite data from both proprietary and free platforms at different spatial resolutions were used as features in the study (i.e., the GSMP: 11.13 km, Terra: 1 km, Landsat 8: 30 m, Sentinel-2: 10 m, and PlanetScope: 3 m). The best possible combination of features was analyzed and hyperparameter tuning was applied to enhance the prediction accuracy. Our results suggest that high-resolution data (PlanetScope) combined with irrigation information, vegetation indices, and climate data significantly improves almond yield prediction. The XGBRegressor model performed best when using PlanetScope data, reaching a coefficient of determination (R^2) of 0.80. However, alternative options using freely available data with lower spatial resolution, such as GSMaP and Terra MODIS LST, also showed satisfactory performance $(R^2 = 0.68)$. This study highlights the potential of integrating machine learning models and remote sensing data for accurate crop yield prediction, providing valuable insights for informed decision support in the almond sector, contributing to the resilience and sustainability of this crop in the face of evolving climate dynamics.

Keywords: *Prunus dulcis*; machine learning; regression models; multispectral data; vegetation indices; remote sensing

1. Introduction

The almond tree, *Prunus dulcis* (var. *dulcis* (*Rosaceae*)), is a globally important nut tree [1]. Originating from the Middle East and South Asia, almonds are extremely important for the human diet due to their high protein content, good fats, and essential micronutrients,



Citation: Guimarães, N.; Fraga, H.; Sousa, J.J.; Pádua, L.; Bento, A.; Couto, P. Comparative Evaluation of Remote Sensing Platforms for Almond Yield Prediction. *AgriEngineering* **2024**, *6*, 240–258. https://doi.org/10.3390/ agriengineering6010015

Academic Editor: Murali Krishna Gumma

Received: 18 December 2023 Revised: 13 January 2024 Accepted: 18 January 2024 Published: 22 January 2024



Copyright: © 2024 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/). including vitamin E and magnesium. Moreover, almond consumption provides diverse health benefits, from improving cholesterol levels and cardiovascular health, to potentially reducing cancer risks [2]. Almond production is very important for the economy in many regions of the world [3]. Almonds are mainly produced in the United States of America (1,858,010 tons), Australia (360,328 tons), and Spain (245,990 tons) [4].

Climate change poses significant risks to crop yields, due to changing weather patterns and more regular risky weather events, such as floods, droughts, and heat waves [5]. These events can have a negative impact on almond productivity and quality [6], compromising stock supplies and promoting price fluctuations. Given these circumstances, precise crop yield prediction has become indispensable, as it equips policymakers and market participants with essential tools to effectively mitigate these risks. Through the analysis of historical data and current agricultural conditions, it becomes possible to develop models for estimating seasonal yield forecasts and evaluating potential supply shortages or surpluses [7]. Furthermore, this information may assist governmental agencies in making informed decisions regarding trade policies, food aid, and agricultural investments [8].

Crop yield prediction is a difficult undertaking that requires integrating several factors, including weather, soil properties, pest and disease incidence, and management practices [9]. More precise modelling techniques for forecasting agricultural yield have recently been developed, due to developments in technology and data analysis. Simple statistical models (i.e., linear regressions) remain the most popular approach for predicting agricultural yield, providing helpful information for decisionmakers [10]. However, machine learning (ML) algorithms have become a promising approach, as they can increase prediction accuracy by finding patterns and relationships in the data [11]. Due to these facts, machine learning (ML) is currently one of the most important subfields of artificial intelligence (AI) [12]. Remote sensing (RS) is another promising field of research that may potentially benefit crop yield prediction. Advances in RS technologies have made it possible to monitor crop development and health in real time from aerial viewpoints, which has enhanced crop production forecasting. RS technologies provide detailed information on the crop conditions, including plant biomass, water content, and nutrient status, which can be utilized to make more precise predictions about future yields [13].

Numerous studies emphasize the significance of ML and RS in predicting crop yield. Klompenburg et al. [14] developed a systematic literature review to detect prevalent models, features, and evaluation parameters in crop yield prediction. The authors observed that linear regression (LR) and neural networks (NN) are frequently applied models, along with random forest (RF) and support vector machines (SVM). Moreover, rainfall, temperature, and soil type are the main features implemented, along with vegetation indices (VIs), such as the normalized difference vegetation index (NDVI) [15] and the enhanced vegetation index (EVI) [16]. Ali et al. [17] highlighted the application of various RS technologies alongside multi- and hyperspectral data, radar, and LiDAR data in crop monitoring and yield prediction. They identified the NDVI, the EVI, and the soil-adjusted vegetation index (SAVI) [18] as commonly used VIs. Similarly, Escolà et al. [19] evaluated the application of Sentinel-2 derived VIs, such as the NDVI, the wide dynamic range vegetation index (WDRVI), the green-red vegetation index (GRVI), and the green normalized difference vegetation index (GNDVI), for estimating barley production. Regarding almond yield prediction using RS data, two studies have emerged. Zhang et al. [20] applied ML models to satellite (Landsat 8) and aerial imagery to forecast almond yield from orchards in California. They achieved a coefficient of determination (R^2) of 0.71 for early and midseason predictions using stochastic gradient boosting (SGB). Tang et al. [21], also in the context of California, explored the use of deep learning (DL) methods, using unmanned aerial vehicle (UAV) data, and developed a convolutional neural network (CNN). Their model obtained an R^2 of 0.96 and a low error of 6.6% for tree-level almond yield estimation, emphasizing the significant potential of DL for precise tree-level yield prediction.

Although the abovementioned studies exhibit strong results in forecasting almond yield, they are tailored to California [20,21] and, as far as we know, there is a notable absence of research specifically dedicated to Portuguese almond yield forecasting. Despite the growing importance of almond cultivation in this country, this absence represents a critical gap in our understanding of the factors that influence almond yields in this region. The current research seeks to fill this research gap by developing a method for predicting almond yields in the TM region of Portugal, using ML regression models, and identifying the key factors that significantly influence almond yields. Furthermore, an improvement on previous studies may be the analysis of RS data from a diverse range of platforms, including freely available medium-resolution data and proprietary higher resolution data. This strategic combination may be used to investigate the effectiveness of medium-resolution RS platforms, compared to their higher resolution counterparts, for predicting almond yields. This information could potentially be used by sector stakeholders to enhance the decision-making process, enabling more informed and strategic choices for optimizing cultivation practices, resource allocation, and overall productivity.

Considering the research gaps identified, the purpose of this study is 4-fold: (1) to use state-of-the-art ML regression models to accurately simulate the yield from several orchards in the TM region; (2) to integrate RS data from different platforms at different spatial resolutions, including from both open and proprietary platforms; (3) to identify the key features that significantly influence these predictions; and (4) to discuss potential applications of these findings in the sector.

2. Materials and Methods

2.1. Study Area

In this study multiple almond orchards (AO) are included, from four distinct almond growers (AGs) within the TM region of northern Portugal.

Regarding AO1, AO2, and AO3, these are located in the Torre de Moncorvo municipality. Regarding AO4, this is located between the Vila Flor and Alfândega da Fé municipalities (Figure 1a). These orchards present different characteristics, namely AO1 has 5.7 hectares with 1387 almond trees; AO2 has 2.9 hectares with 765 almond trees; AO3 has 3.0 hectares with 756 almond trees; and AO4, the largest area, has 12.3 hectares with 3198 almond trees.

This region of TM is mountainous and presents warm and dry summers and moderately cold and wet winters [22]. These characteristics are typical of the Mediterranean climate, which makes the region suitable for almond cultivation.

Considering the yield levels recorded from 2017 to 2021, AO4 and AO3 had the highest productivity, averaging 1041 kg/ha and 785 kg/ha, respectively, while AO1 and AO2 had lower productivity with an average of 462 kg/ha and 372 kg/ha, respectively, in the same period (Figure 1b).

2.2. Data Collection and Processing

The data processing workflow consists of four sequential steps (Figure 2). In the first step, the data acquired from various sources is collected, including the agronomic parameters, vegetation indices, and climate data (identified in the following subsections). The second step involves the integration of various features into a dataset comprising 171 features. The third step includes the application of ML regression models, which includes the feature selection process, the selection of ML regression models, and hyperparameter optimization. In the fourth step, the model evaluation is conducted. The details of the four-step approach are provided in the subsequent subsections.



Figure 1. Overview of the location of the almond orchards, for each almond grower (AG) (**a**), and (**b**) the yield values, from 2017 to 2021, presented in kg/ha, for each AG.

2.2.1. Agronomic Data

The agronomic data contains several parameters collected from each site. In addition to yield data (the target feature), yearly irrigation information was also acquired from each grower, recorded as binary values (0 for no irrigation and 1 for irrigation). Irrigation is recognized as a vital factor that significantly influences the optimal growth and development of trees, consequently affecting crop productivity [23]. The availability and efficient distribution of water directly affects physiological processes, such as transpiration and nutrient uptake, which are critical for trees to reach their full yield potential. Moreover, appropriate irrigation practices can help alleviate the adverse effects of environmental stressors, such as droughts or heat waves, which are becoming increasingly prevalent due to climate change [24]. Furthermore, data regarding the average tree age (plantation date) were incorporated as a feature of the dataset. The age of almond trees is of paramount importance for productivity, as older trees tend to have more extensive root systems, established canopies, and enhanced nutrient storage, leading to increased almond production and increased overall orchard yield [20]. The data were pre-processed for each orchard separately to filter the outliers, based on distribution analysis.

2.2.2. Remote Sensing Data

Several RS data from various platforms with different spatial resolutions were considered (Table 1). The Global Satellite Mapping of Precipitation (GSMaP) by the Japan Aerospace Exploration Agency (JAXA) was used, which provides global precipitation data using a combination of sensors at ~11 km [25]. It was developed in Japan specifically for the GPM mission [26]. The Land Surface Temperature (LST) from the Moderate Resolution Imaging Spectroradiometer (MODIS), operated by the National Aeronautics and Space Administration (NASA), was obtained at a resolution of 1 km [27]. It should be noted that the thermal sensor in the MODIS only offers 1 km resolution. Landsat 8, operated by NASA and the United States Geological Survey (USGS), offers multispectral data at a resolution of 30 m [28]. Land cover classification and analysis of vegetation are potential applications of these data. Sentinel-2 was developed by the European Space Agency (ESA) and provides multispectral imagery at a resolution of 10 m. This is the best resolution imagery available for free today [29]. PlanetScope, which is a proprietary data source, is formed of several small satellites (constellation), operated by Planet Labs Inc., designed for high-frequency global imaging of Earth. The satellites acquire imagery in the visible and near-infrared spectra and provide a spatial resolution of 3 m. The proprietary high-resolution data provided by PlanetScope enables detailed mapping and monitoring of various features, including urban areas, vegetation dynamics, and environmental changes [30].



Figure 2. Data processing workflow stages: (1) data collection; (2) dataset creation; (3) application of machine learning (ML) models; and (4) model evaluation.

Platform/ Satellite	Sensor	Product	Spatial Resolution (m)	Revisiting Time	
GSMaP	Multi-Band Passive Microwave and Infrared Radiometers	Hourly Precipitation Rate	11,000	3 h	
Terra	Moderate- Resolution Imaging Spectroradiometer (MODIS)	Daytime and Nighttime Land Surface Temperature (LST)	1000	1 day	
Landsat 8	Operational Land Imager (OLI)	RGB and NIR bands	30	16 days	
Sentinel-2	Multispectral Instrument (MSI)	RGB and NIR bands	10	5 days	
PlanetScope	DOVE-R	RGB and NIR bands	3	1 day	

Table 1. Remote sensing platform overview: sensors, bands, spatial resolutions, and revisiting time.

Monthly composites from GSMaP, MODIS Terra LST, Landsat 8, and Sentinel-2 (from 2017 to 2021) were computed using the Google Earth Engine (GEE). The GEE is an online infrastructure that archives satellite imagery and geospatial data, offering powerful analytics tools, leveraging cloud-based infrastructure. These benefits make it an invaluable tool for exploring Earth's dynamics and supporting fact-based decision making [31]. On the other hand, PlanetScope's monthly composites were acquired using the Planet Explorer platform, which is a fully automated, cloud-based imaging and analysis platform that grants users access to comprehensive, daily data from the PlanetScope and SkySat constellations.

2.2.3. Vegetation Indices Computation

VIs were also included in the feature dataset, namely the enhanced vegetation index 2 (EVI2), the GRVI, the NDVI, and the SAVI (Table 2). The use of the EVI2 in crop yield prediction models is justified by its many advantages, and has been shown to achieve higher prediction accuracy compared to other VIs, such as the NDVI [32]. It also offers higher sensitivity, especially in areas with high biomass, and provides valuable information on crop conditions and yields [33,34]. Regarding the GRVI, it is often used as a phenological indicator, detecting changes in canopy vegetation [35]. Moreover, in a study by Sanches et al. [36], the GRVI showed a high correlation with sugarcane yields. Concerning the NDVI, it can be implemented to monitor crop growth, detect plant stress, and make decisions regarding irrigation, fertilization, pesticide application, and has also been employed in numerous studies to accurately predict crop yields [37,38]. The SAVI, in turn, is also a suitable VI for use in yield prediction models, since it attempts to minimize the effects of soil brightness using a correction factor [39]. It is similar to the NDVI, but accounts for variations in soils, making it useful in arid and semi-arid regions, where vegetation cover is low and soil brightness can significantly affect vegetation detection [40]. Furthermore, in a study by da Silva et al. [41], the SAVI had the highest correlation with soybean grain yield, possibly due to the use of the soil effect correction, demonstrating its ability to predict crop yields.

As previously mentioned, three different platforms, each with varying spatial resolutions, were used to obtain the VIs (Section 2.2.2). The data from each platform were processed, and atmospheric corrections were implemented before it became available.

Name of Index	Equation	Reference
Enhanced vegetation index 2	$EVI2 = \frac{2.5 \times (N-R)}{(N+2.4 \times R+1)}$	[16]
Green-red vegetation index	$GRVI = \frac{G-R}{G+R}$	[42]
Normalized difference vegetation index	$NDVI = \frac{N-R}{N+R}$	[15]
Soil-adjusted vegetation index	$SAVI = \frac{N-R}{N+R+L} \times (1+L)$	[18]

Table 2. Vegetation indices used in almond yield prediction and their respective equations. G: green; L = 0.5; N: near infrared; R: red.

The data for each orchard were obtained using the geospatial data abstraction library (GDAL) in Python to calculate the mean value for each grower. Figure 3 illustrates the above-described procedure, by displaying an example of the NDVI computed for the three VI platforms, in March 2019. Figure 3a–c depicts the NDVI images, with the spatial resolution associated with each platform. They were then subjected to a mean calculation, yielding a singular value that was then used in the creation of the final dataset. These three VI datasets were produced to compare the performance of the Landsat 8, Sentinel-2, and PlanetScope data.



Figure 3. Example of the procedure for creating the dataset. The values correspond to the mean NDVI calculated for (**a**) Landsat 8, (**b**) Sentinel-2, and (**c**) PlanetScope (e.g., March 2019), corresponding to almond grower 3.

The final dataset is created using three different groups: climate data, VIs, and agronomic data. As mentioned in Section 2.2.3, the three datasets were used in parallel with the VIs customized according to the RS platform (Landsat 8, Sentinel-2, and PlanetScope) to compare the performance of each. This information, extracted for each orchard, was then used as potential inputs into the ML regression models. Each of the datasets contained 171 features, corresponding to the average tree age, irrigation, monthly mean daytime temperature (2017–2021), monthly mean nighttime temperature (2017–2021), monthly accumulated precipitation (2017–2021), monthly EVI2 (2017–2021), monthly GRVI (2017–2021), monthly NDVI (2017–2021), and monthly SAVI (2017–2021). To consider the potential effect of alternate bearing (biannual cyclic production patterns), features from the previous year were also included to assess their potential impact on the following year's production. In addition to features, the datasets also included the yield (mean kg per ha) as a target.

2.4. Application of Machine Learning Regression Models

The application of ML regression models followed a three-step process: feature selection, model implementation, and hyperparameter tuning (detailed in the following subsections). In the feature selection phase, relevant features were carefully chosen to enhance the model's accuracy and performance. Subsequently, the regression models were implemented using a cross-validation methodology, establishing a connection between the input features and the target feature. The final step involved optimizing the hyperparameters to fine tune the models to improve the predictive accuracy and generalization capability.

2.4.1. Feature Selection Process

Feature selection plays a key role in the ML pipeline. Selecting the most suitable features improves the model's performance, reduces computational efforts, and assists in the interpretation of the results. By selecting the most relevant features, more accurate, efficient, and interpretable models can be achieved, facilitating better decision making and a better understanding of the underlying data patterns. Herein, for the selection of the best possible features, the *bestFeatures* script [43] was used. This is a tool for identifying the best possible combination of features for fitting a ML model. This method uses cross-validation (CV) to evaluate different feature subsets and their corresponding performance scores (R^2). The CV method partitions the dataset into training and testing subsets, multiple times (folds). For each fold, part of the dataset (testing) is always unseen by the algorithm. It then computes the cross-validated score for each combination of features and tracks the maximum score achieved, effectively controlling the problem of overfitting. The method also ensures a low level of correlation related to the features. The output of the method includes the R^2 score and error metrics, corresponding to the best feature combination. For the current analysis, this method was used considering a 5-fold CV, and a combination of 1 to 8 features, determined by the script's best performing approach. The analysis was performed separately for each dataset (agronomic features, climate features, and VIs, from the three different platforms), as well as for mixed datasets (e.g., climate features + VIs), depending on the vegetation data acquisition platform. Table 3 lists the selected features, according to the type of features considered.

Type of Features	Selected Features			
Irrig.	Irrigation			
CD	DTJan.; DTAug.; DTNov. (LY); NTMar.			
VI—PS	EVI2Feb.; GRVIMar.; GRVIJul.; SAVIDec. (LY)			
VI—S2	EVI2Aug.; GRVIJan.; NDVIAug.; SAVIAug.			
VI—L8	EVI2May.; NDVIMay.; NDVIAug.; SAVISep. (LY)			
Irrigation and CD	Irrigation; DTMar.; DTAug.; NTMar.			
Irrigation and VI—PS	Irrigation; GRVIMay.; NDVIJan.; SAVIMay.			
Irrigation and VI—S2	Irrigation; EVI2Apr.; EVI2Jul.; GRVIJan.			
Irrigation and VI—L8	Irrigation; SAVIMay; SAVIAug.; SAVISep. (LY)			
CD and VI—PS	DTAug.; EVI2Mar.; NDVIMar.; SAVIDec. (LY)			
CD and VI—S2	DTJan.; DTOct. (LY); EVI2Dec. (LY); NDVIAug.			
CD and VI—L8	DTMar.; DTAug.; EVI2Jun.; SAVISep. (LY)			
Irrigation, CD and VI—PS	Irrigation; NDVIJan.; SAVIMay.; DTMar.			
Irrigation, CD and VI—S2	Irrigation; GRVIMay; DTMar.; DTAug.			
Irrigation, CD and VI–L8	Irrigation; NDVIMay; DTMar.; NTFeb.			

Table 3. Selected features using *bestFeatures* script, categorized by feature type. CD: climate data; DT: daytime temperature; Irrig.: irrigation; L8: Landsat 8; LY: last year; NT: nighttime temperature; PS: PlanetScope; S2: Sentinel-2; VI: vegetation indice.

2.4.2. Machine Learning Regression Model Selection

Several ML regression models were applied to predict almond yield. Among these models, the random forest regressor (RFR) stands out as a prominent option due to its effectiveness in supervised learning [44], and it is used in many fields of study. The RFR algorithm generates an ensemble of decision trees, collectively called an RF. Each decision tree independently learns patterns and relationships within the data, contributing equally to the final prediction, improving the performance and efficacy of the model and dealing with potential overfitting problems [44]. The XGBRegressor (XGBR) was also implemented in this study and is a supervised learning algorithm that belongs to the gradient boosting family. It employs a boosting technique that sequentially improves decision trees, to create a powerful ensemble model [45]. XGBR optimizes the training objective through gradient descent, allowing it to effectively identify complex patterns and dependencies in the data. The model has shown remarkable performance in several areas, making it a valuable tool for almond yield prediction [45]. Regarding the gradient boosting regressor (GBR) algorithm, it is also a gradient boosting-based regression model. It iteratively builds an ensemble of decision trees, with each successive tree better than the previous ones, creating a strong predictive model. The GBR model is widely used in predictive analytics and has demonstrated its efficiency and potential for predicting almond yield [46]. The bagging regressor (BR) algorithm, on the other hand, uses a bagging technique similar to the RFR model. It generates an ensemble of decision trees by resampling the training data and fitting each sample to a separate tree, which are combined to form the final output. Its ability to handle high-dimensional data and complex relationships makes it a suitable candidate for almond yield prediction [47]. Lastly, the AdaBoost regressor (ABR) algorithm is a boosting-based regression model that iteratively adjusts the weights assigned to the training instances, placing more emphasis on the samples that are difficult to predict accurately. ABR is known for its adaptability to different data types and its ability to handle noisy or incomplete datasets [48]. The ML regression models were implemented using the Python library Scikit-learn [49].

2.4.3. Hyperparameter Tuning

Following the selection of the best feature combination, the ML models were applied using hyperparameter tuning. This method may significantly improve the performance of ML regression models [50] by adjusting each model's internal parameters, such as the learning rate and tree depth. In this study, a systematic approach for hyperparameter tuning was employed, utilizing the GridSearchCV method with a 5-fold CV. Again, 5-fold cross-validation further ensures robustness by splitting the dataset into five subdivisions, using four for training and one for validation in a rotating fashion. This iterative process allows the model to be trained and validated multiple times, always withholding the testing data from the algorithm, providing a robust assessment of its performance across various hyperparameter settings. The method examines several hypotheses and identifies the optimal hyperparameters, based on the performance of the models, considering the R^2 results. This approach aims to maximize the predictive accuracy of the ML regression models. The specific hyperparameters applied in the implementation of the regression models are identified in Table 4.

Table 4. Main hyperparameters considered during regression models implementation. Tested hyperparameter values: *n_estimators* (NE): 100, 200, 300; max_depth (MD): 3, 5, 7, 9, 1; max_samples (MS): 0.5, 0.75, 1.0; learning_rate (LR): 0.1, 1.0, 10. CD: climate data; Irrig.: irrigation; L8: Landsat 8; PS: PlanetScope; S2: Sentinel-2; VI: vegetation indice.

Type of	RFR		XGBR		GBR		BR		ABR	
Features	NE	MD	NE	MD	NE	MD	NE	MS	NE	LR
Irrig. CD	100 100	3 3	100 100	3 3	100 300	3 3	100 100	0.5 1.0	100 100	1.0 0.1
VI—PS VI—S2	100 200	5 7	100 100	5 7	100 100	3 3	100 200	1.0 0.75	300 100	0.1 1.0
VI—L8	100	5	200	7	100	3	100	1.0	300	0.1
Irrig. and CD	100	7	100	7	100	3	100	1.0	200	0.1
Irrig. and VI—PS	200	3	100	5	300	3	200	1.0	100	0.1
Irrig. and VI—S2	300	5	100	5	300	3	300	1.0	300	1.0
Irrig. and VI—L8	200	5	200	5	200	3	200	1.0	300	0.1
CD and VI—PS	200	7	200	5	300	3	200	1.0	100	1.0
CD and VI—S2	200	9	200	3	300	3	300	1.0	300	1.0
CD and VI—L8	100	5	100	5	300	3	100	1.0	200	0.1
Irrig., CD and VI—PS	100	5	100	3	100	3	300	1.0	100	1.0
Irrig., CD and VI—S2	100	5	100	3	200	3	100	1.0	300	1.0
Irrig., CD and VI—L8	100	3	100	3	200	3	100	1.0	200	1.0

2.5. Model Evaluation

Several well-known metrics to evaluate regression model performance were used, including the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE). R^2 measures the percentage of variance in the dependent variable that can be described by the independent variables in a regression model [51]. The RMSE accounts for positive and negative deviations between predicted and observed values. Regarding the MAE, it is calculated as the average of the absolute differences between the predicted and observed values of the dependent variable. The average magnitude of errors performed by the model is measured in the units of the same order as that of the dependent variable and denoted as the MAE. While the RMSE takes the squared differences into account, the MAE does not which makes it robust in regard to outliers and insensitive to extreme errors. By considering the mean absolute difference between the observed and simulated values, the MAE provides the overall accuracy of the model, regardless of the direction of the errors [52].

3. Results

3.1. Comparative Analysis of Regression Models for Almond Yield Prediction

This study intends to investigate and compare the performance of several ML regression models in predicting almond tree yield, based on different features extracted from RS platforms and different agronomic parameters. The features considered were irrigation, temperature, precipitation, and VIs. The regression models evaluated included the RFR, XGBR, GBR, BR, and ABR models. Figure 4 shows the performance of the regression models, using different types of features, considering the VIs computed by PlanetScope (Figure 4a), Sentinel-2 (Figure 4b), and Landsat 8 (Figure 4c). When focusing on the irrigation feature alone, none of the models demonstrated exceptional performance as a standalone feature across all three platforms. The R^2 scores for all the models ranged from 0.32 to 0.38, suggesting limited predictive capabilities. When considering only the climatic features, the GBR model and BR model consistently exhibited higher performance, with R^2 scores of around 0.40. For the VIs only, the ABR model outperformed the others, using PlanetScope data ($R^2 = 0.59$). On the other hand, considering the VIs calculated using Sentinel-2 data, the BR model performed better ($R^2 = 0.49$). For the VIs from Landsat 8, the best performance was achieved by the RFR model ($R^2 = 0.61$). When combining features, particularly irrigation and climate data, the RFR model consistently demonstrated strong performance across all three platforms ($R^2 = 0.68$), indicating its superior predictive capabilities for this feature combination. Considering the combination of irrigation and VIs, the XGBR model obtained the best performance using PlanetScope ($R^2 = 0.76$) and Sentinel-2 ($R^2 = 0.66$) data. In contrast, the GBR model produced the best result ($R^2 = 0.73$) when the Landsat 8 data were used. Combining the climate data and VIs, the ABR model performed well using PlanetScope data ($R^2 = 0.61$). On the other hand, the XGBR model performed better using the Sentinel-2 ($R^2 = 0.62$) and Landsat 8 data ($R^2 = 0.69$). Finally, considering all three groups of features combined (irrigation, climate data, and VIs), the XGBR model showed the best performance using PlanetScope data ($R^2 = 0.80$), while the RFR model showed the best performance using Sentinel-2 data ($R^2 = 0.67$) and the ABR model using Landsat 8 data ($R^2 = 0.72$). The results presented showed that the XGBR and RFR models proved to be the most appropriate models for predicting the target feature using the PlanetScope data across different feature combinations. However, when using Landsat 8 data, the ABR model also provided remarkable results.

The performance of several regression models was also evaluated based on *MAE* and *RMSE* metrics, considering the different combinations of feature types (Table 5). Notably, the best combination of features that yielded optimal results varied across the models. When focusing on the irrigation feature, all the models (RFR, XGBR, GBR, BR, and ABR) achieved similar performance, with *MAE* values ranging from 158 to 161 kg/ha and *RMSE* values ranging from 206 to 212 kg/ha. For the climate data features, the XGBR (*MAE* = 210 kg/ha; *RMSE* = 255 kg/ha) and BR (*MAE* = 184 kg/ha; *RMSE* = 238 kg/ha)

models exhibited slightly higher values compared to the other models. Among the VIs, when the PlanetScope data were used, the best performance was achieved with the ABR model (MAE = 143 kg/ha; RMSE = 186 kg/ha). For the Sentinel-2 data, the lowest MAE (165 kg/ha) was obtained with the BR model and the lowest RMSE (214 kg/ha) was obtained with the RFR model. Considering the VIs from Landsat 8, the lowest MAE (133 kg/ha) was obtained with the GBR model and the lowest RMSE (189 kg/ha) with the RFR model. Combining irrigation with climate data generally improved the model's performance, compared to using either feature individually, resulting in lower MAE and RMSE values, ranging from 116 to 137 kg/ha and 158 to 177 kg/ha, respectively. Similarly, incorporating all feature types tended to improve the model's predictive performance, especially when using irrigation, climate data, and VIs from PlanetScope, which achieved the best performance with the XGBR model (MAE: 95 kg/ha; RMSE: 119 kg/ha).

Table 5. Performance of regression models by feature type and remote sensing platform, assessed using mean absolute error (*MAE*) and root mean square error (*RMSE*). Units in kg/ha. CD: climate data; Irrig: irrigation; L8: Landsat 8; PS: PlanetScope; S2: Sentinel-2; VI: vegetation indice.

Type of	RFR		XGBR		GBR		BR		ABR	
Features	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Irrig.	159	211	161	211	161	211	160	212	158	206
CD	184	237	210	255	178	211	184	238	179	223
VI—PS	154	200	173	220	148	194	152	201	143	186
VI—S2	168	214	177	224	181	232	165	215	184	229
VI—L8	144	189	158	250	133	193	145	191	142	202
Irrig. and CD	134	170	125	177	116	158	137	172	133	174
Irrig. and VI—PS	132	158	107	138	117	138	131	159	118	153
Irrig. and VI—S2	147	190	130	168	130	168	149	192	120	163
Irrig. and VI—L8	145	176	141	199	111	147	144	178	142	180
CD and VI—PS	157	193	151	189	152	210	160	193	143	177
CD and VI—S2	156	211	139	175	196	220	154	211	160	209
CD and VI—L8	139	174	127	159	123	162	133	171	114	154
Irrig., CD and VI—PS	126	153	95	119	95	125	129	153	113	151
Irrig., CD and VI—S2	135	173	126	175	124	171	135	174	123	173
Irrig., CD and VI—L8	126	163	137	178	154	194	123	163	108	152

3.2. Selected Features and Their Contribution to Almond Yield Prediction

Considering the information from the previous subsection, optimal performance was achieved with the XGBR model using the irrigation feature, climate data (specifically, the daytime temperature in March), and VIs (the NDVI in January and the SAVI in May) calculated using PlanetScope data. This subsection is intended to present the features that were important to almond yield prediction. Figure 5a shows that in the almond orchards where irrigation was applied (AG3 and AG4), higher yield values were recorded, about 913 kg/ha, while in the almond orchards where irrigation was not applied (AG1 and AG2), lower yield values were recorded, about 417 kg/ha. Comparing the NDVI in January and the yield (Figure 5b), it is possible to observe higher NDVI values related to a higher yield, although it is not a clear linear relationship. In the almond orchards of AG4 (Figure 1b), higher NDVI values were recorded in January from 2019 to 2021, which may have made a

positive contribution to achieving higher production values compared to other growers. Similarly, regarding the SAVI in May (Figure 5c), it does not seem to be linearly associated with the yield, highlighting the importance of ML models, as non-linear patterns can be identified by these models. On the other hand, regarding the daytime temperature in March (Figure 5d), it is evident that lower values were associated with lower yield values. The lowest daytime temperature value in March (12.2 °C) was recorded in 2018 in the almond orchards of AG2, which coincided with the lowest yield value (82 kg/ha) compared to the other growers (Figure 1b). As for the highest values recorded for the daytime temperature in March, the maximum value reached was 22.5 °C in 2019 by AG4, who obtained higher yield values (1203 kg/ha).



Irrigation, Climate Data and Vegetation Indices

Figure 4. Regression model performance based on feature type and remote sensing platform, using the coefficient of determination (R^2). In (**a**) vegetation indices were computed using PlanetScope data; in (**b**) vegetation indices were computed using Sentinel-2 data; and in (**c**) vegetation indices were computed using Landsat 8 data.



Figure 5. Correlation charts on the yield with: (**a**) irrigation; (**b**) NDVI in January; (**c**) SAVI in May; and (**d**) daytime temperature in March, Celsius degrees (°C).

4. Discussion

The present study aimed to develop ML models to simulate almond yields in the TM region, applying open and proprietary RS data. The comparison was made among various free RS platforms (MODIS Terra LST, GSMaP, Landsat 8, and Sentinel-2) and a paid one (PlanetScope). Several ML regression models (RFR, XGBR, GBR, BR, ABR) were applied, and the optimum feature combination was selected to achieve the best performance. The combination of irrigation data, daytime temperature in March, the NDVI in January, and the SAVI in May (from the PlanetScope platform) showed the best performance ($R^2 = 0.80$), using the XGBR model. Indeed, the use of VIs with a higher resolution (3 m) from the PlanetScope data had a positive influence on the almond yield prediction, as the results obtained with Sentinel-2 ($R^2 = 0.67$ —RFR) and Landsat 8 ($R^2 = 0.72$ —ABR) data were lower. However, it is worth noting that free data with lower resolution could also be a viable alternative to PlanetScope, particularly RS platforms providing climatic data, such as MODIS Terra LST and GSMaP, achieving an R^2 of 0.68 when using the RFR model with

irrigation and climate data. Regarding the XGBR model, it was observed that the best results were reached when using three groups of features (irrigation, climate data, and VIs) ($R^2 = 0.80$). On the other hand, the XGBR model obtained inferior performance when using only irrigation information (R^2 : 0.32), only climate data (R^2 : 0.19), or only VIs (PlanetScope— $R^2 = 0.44$). In this case, the RFR model achieved a higher level of performance than the XGBR model. This situation might be due to the XGBR model being more capable of handling complex relationships between features [45], while the RFR model is known to perform better with simpler features. Considering other studies related to almond yield prediction, Zhang et al. tested several ML models, obtaining the best performance with the SGB model, with an R^2 of 0.71, which is also a boosting model [20].

Considering the most important features, irrigation and daytime temperature in March stood out, highlighting the role of water availability and suitable temperatures for almond yields [53]. In fact, AG3 and AG4 show higher yield values, due to the available irrigation. Furthermore, March is considered a crucial period for almond trees, as flowering occurs at this stage [54]. Almond trees are highly sensitive to climatic conditions during the flowering period, and adequate temperatures are essential for successful pollination and fruit development. According to Tamimi [54], the ideal temperature for almond tree flowering during the day is between 15 $^{\circ}$ C and 30 $^{\circ}$ C, and temperatures outside this range can lead to problems, resulting in reduced fruit production. In effect, daytime temperatures in March in the agricultural fields of AG2 were recorded, with a minimum temperature of 12.2 °C recorded in 2018, which is below the considered ideal temperature for almond flowering, which can explain the low production in that year (Supplementary Figure S1). On the other hand, in the agricultural fields of AG4, a maximum temperature of 22.5 $^{\circ}$ C was recorded in 2019, falling within the range of ideal temperatures for flowering, resulting in increased production that year. Similar studies have also highlighted temperaturerelated features. Zhang et al. [20] emphasized the importance of the feature "long-term mean maximum April-June temperature" in predicting almond production. According to the authors, this factor significantly affects the blooming period of almond trees. Almonds are sensitive to temperature fluctuations during this critical stage, and optimal temperatures promote successful pollination and higher yield. However, exceeding the temperature threshold can negatively impact pollination and reduce fruit set, leading to lower almond yield. Therefore, monitoring and considering the long-term mean maximum April-June temperature is essential for accurately predicting almond yield. Other studies, such as the study by Tombesi et al. [55], have also considered that warm springs accelerate fruit development.

The intricate relationships unveiled throughout this analysis underscore the necessity of employing sophisticated ML models for understanding the dynamics influencing almond production. The interdependence of variables like irrigation, climate indicators, and vegetation indices highlights the need for advanced analytical tools, and the application of sophisticated machine learning (ML) models becomes imperative. Unlike simpler models, such as linear regression, which assume linear relationships between the variables, the complexities of almond production necessitate more sophisticated approaches. The utilization of advanced ML models, like the XGBR model applied in this study, allows for the exploration of intricate, non-linear relationships among various contributing factors. In the realm of almond production, where variables often exhibit non-linear dependencies, these models excel in discerning patterns that may elude simpler methodologies.

Some limitations must be acknowledged. Optimal results were achieved through the utilization of a proprietary/paid platform, potentially limiting accessibility for certain users. Furthermore, the lower resolution of data provided by open platforms may impede the identification of smaller orchard areas. Nevertheless, our study underscores the viability of utilizing freely available remote sensing data. While the data were sourced from multiple farmers, expanding the dataset could enhance the robustness of our findings. Despite the abovementioned limitations, the current study methodology holds promise for adaptation and implementation in various agricultural settings worldwide. Another important point

is the careful analysis of features to increase the overall performance of the ML regression models. In this way, this study not only allows for the prediction of almond yields, but also enables the identification of the key factors that significantly influence these predictions. Furthermore, the models developed allow the implementation of early prediction of seasonal almond yields, with the potential integration of climate data and extreme weather events. The comparison between open and proprietary RS data shows that these models can be implemented using these two types of datasets. As such, these results provide valuable insights for farmers and other sector stakeholders, in the decision-making process, which can enhance the sustainability of the almond sector in Portugal.

5. Conclusions

This study investigates the potential of RS data and ML models for predicting almond yield. Various RS platforms were evaluated, including both freely available platforms, including MODIS Terra LST, GSMaP, Landsat 8, and Sentinel-2, as well as a paid platform, PlanetScope. In addition, the performance of several ML regressors, including RFR, XGBR, GBR, BR, and ABR, were evaluated. The inclusion of high-resolution VIs from the PlanetScope platform significantly increased the accuracy of almond yield prediction. The XGBR model trained with a feature set comprising irrigation data, the daily temperature in March, the NDVI in January, and the SAVI in May from the PlanetScope platform showed the highest predictive performance, achieving an R^2 value of 0.80. This indicates that the model could effectively explain 80% of the variation in the almond yield. However, freely available RS platforms, such as MODIS Terra LST and GSMaP, can also serve as viable alternatives to PlanetScope data. Despite the lower spatial resolution, the data from these platforms demonstrated that it still provides valuable insights for predicting almond yield. It is worth noting, however, that the choice of ML model was found to be a critical factor in the prediction accuracy. While the XGBR model consistently outperformed the other models, it proved more prone to noise and outliers when only one or two types of features were used. Therefore, the selection of the most suitable ML algorithm should be based on the dataset and features to be considered. Irrigation and the daytime temperature in March were among the most important features for predicting almond yield, highlighting the pivotal role of water and temperature in crop growth and development. Future research may be aimed at the continuous improvement of the dataset implemented in this study, by increasing the number of almond orchards by considering broader geographical areas and including established climatic and temporal relationships with the yield in the evaluated orchards. This will improve the generalization ability of the models. It would also be useful to consider very-high resolution UAV multispectral data to provide tree-level almond yield.

Supplementary Materials: The following supporting information can be downloaded at: https://www. mdpi.com/article/10.3390/agriengineering6010015/s1; Figure S1. Monthly average land surface temperature (left panels) and monthly precipitation sum, for 2017 to 2021, in the studied almond orchards: (a) almond grower 1, (b) almond grower 2, (c) almond grower 3, (d) almond grower 4.

Author Contributions: Conceptualization, N.G. and J.J.S.; Methodology, N.G. and A.B.; Software, N.G. and H.F.; Validation, N.G. and J.J.S.; Formal analysis, N.G., H.F. and P.C.; Investigation, N.G., H.F. and A.B.; Resources, N.G., H.F. and L.P.; Data curation, N.G., H.F. and L.P.; Writing—original draft, N.G.; Writing—review and editing, N.G., H.F., J.J.S., L.P., A.B. and P.C.; Visualization, N.G., H.F. and L.P.; Supervision, J.J.S., A.B. and P.C.; Project administration, J.J.S.; Funding acquisition, J.J.S. and P.C. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support was provided by national funds through FCT—Portuguese Foundation for Science and Technology (UI/BD/150727/2020), under the Doctoral Programme "Agricultural Production Chains—from fork to farm" (PD/00122/2012), and from European Social Funds and the Regional Operational Programme Norte 2020. This study was also supported by CITAB research unit (UIDB/04033/2020; https://doi.org/10.54499/UIDB/04033/2020 (accessed on 17 December 2023)), Inov4Agro (LA/P/0126/2020; https://doi.org/10.54499/LA/P/0126/2020 (accessed on 17 December 2023)) and by CIMO (UIDB/00690/2020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The remote sensing data presented in this study were collected from Google Earth Engine Catalog (https://developers.google.com/earth-engine/datasets/catalog (accessed on 17 December 2023)) and Planet Labs (http://www.planet.com (accessed on 17 December 2023)). The almond yield data is available from individual farmers at Quinta do Barração da Vilariça and Amendoacoop.

Acknowledgments: We would like to express our sincere gratitude to Quinta do Barracão da Vilariça and Amendoacoop for their invaluable contribution in providing the data for this study. HF thanks the FCT for 2022.02317.CEECIND (https://doi.org/10.54499/2022.02317.CEECIND/CP1749/CT0002 (accessed on 17 December 2023)). We also gratefully acknowledge the Education and Research Program of Planet Labs.

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

References

- 1. Franklin, L.M.; Mitchell, A.E. Review of the Sensory and Chemical Characteristics of Almond (*Prunus dulcis*) Flavor. J. Agric. Food Chem. 2019, 67, 2743–2753. [CrossRef] [PubMed]
- Chavas, J.-P.; Rivieccio, G.; Di Falco, S.; De Luca, G.; Capitanio, F. Agricultural Diversification, Productivity, and Food Security across Time and Space. *Agric. Econ.* 2022, 53, 41–58. [CrossRef]
- Yada, S.; Lapsley, K.; Huang, G. A Review of Composition Studies of Cultivated Almonds: Macronutrients and Micronutrients. J. Food Compos. Anal. 2011, 24, 469–480. [CrossRef]
- FAOSTAT. FAOSTAT—Crops and Livestock Products. Available online: https://www.fao.org/faostat/en/#data/QCL (accessed on 20 November 2023).
- 5. IPCC Climate Change 2022—Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st ed.; Cambridge University Press: Cambridge, UK, 2023; ISBN 978-1-00-932584-4.
- 6. Freitas, T.R.; Santos, J.A.; Silva, A.P.; Fraga, H. Reviewing the Adverse Climate Change Impacts and Adaptation Measures on Almond Trees (*Prunus dulcis*). *Agriculture* **2023**, *13*, 1423. [CrossRef]
- Jones, J.W.; Antle, J.M.; Basso, B.; Boote, K.J.; Conant, R.T.; Foster, I.; Godfray, H.C.J.; Herrero, M.; Howitt, R.E.; Janssen, S.; et al. Brief History of Agricultural Systems Modeling. *Agric. Syst.* 2017, 155, 240–254. [CrossRef]
- Sloat, L.L.; Lin, M.; Butler, E.E.; Johnson, D.; Holbrook, N.M.; Huybers, P.J.; Lee, J.-E.; Mueller, N.D. Evaluating the Benefits of Chlorophyll Fluorescence for In-Season Crop Productivity Forecasting. *Remote Sens. Environ.* 2021, 260, 112478. [CrossRef]
- 9. Cedric, L.S.; Adoni, W.Y.H.; Aworka, R.; Zoueu, J.T.; Mutombo, F.K.; Krichen, M.; Kimpolo, C.L.M. Crops Yield Prediction Based on Machine Learning Models: Case of West African Countries. *Smart Agric. Technol.* **2022**, *2*, 100049. [CrossRef]
- 10. Burdett, H.; Wellen, C. Statistical and Machine Learning Methods for Crop Yield Prediction in the Context of Precision Agriculture. *Precis. Agric.* 2022, 23, 1553–1574. [CrossRef]
- 11. Iniyan, S.; Akhil Varma, V.; Teja Naidu, C. Crop Yield Prediction Using Machine Learning Techniques. *Adv. Eng. Softw.* **2023**, 175, 103326. [CrossRef]
- 12. Nguyen, G.; Dlugolinsky, S.; Bobák, M.; Tran, V.; López García, Á.; Heredia, I.; Malík, P.; Hluchý, L. Machine Learning and Deep Learning Frameworks and Libraries for Large-Scale Data Mining: A Survey. *Artif. Intell. Rev.* **2019**, *52*, 77–124. [CrossRef]
- 13. Muruganantham, P.; Wibowo, S.; Grandhi, S.; Samrat, N.H.; Islam, N. A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. *Remote Sens.* **2022**, *14*, 1990. [CrossRef]
- 14. van Klompenburg, T.; Kassahun, A.; Catal, C. Crop Yield Prediction Using Machine Learning: A Systematic Literature Review. *Comput. Electron. Agric.* 2020, 177, 105709. [CrossRef]
- 15. Rouse, J.W., Jr.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with Erts. *NASA Spec. Publ.* **1974**, *351*, 309.
- 16. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
- Ali, A.M.; Abouelghar, M.; Belal, A.A.; Saleh, N.; Yones, M.; Selim, A.I.; Amin, M.E.S.; Elwesemy, A.; Kucher, D.E.; Maginan, S.; et al. Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *Egypt. J. Remote Sens. Space Sci.* 2022, 25, 711–716. [CrossRef]
- 18. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). Remote Sens. Environ. 1988, 25, 295–309. [CrossRef]
- 19. Escolà, A.; Badia, N.; Arnó, J.; Martínez-Casasnovas, J.A. Using Sentinel-2 Images to Implement Precision Agriculture Techniques in Large Arable Fields: First Results of a Case Study. *Adv. Anim. Biosci.* **2017**, *8*, 377–382. [CrossRef]
- 20. Zhang, Z.; Jin, Y.; Chen, B.; Brown, P. California Almond Yield Prediction at the Orchard Level with a Machine Learning Approach. *Front. Plant Sci.* **2019**, *10*, 809. [CrossRef]

- Tang, M.; Sadowski, D.L.; Peng, C.; Vougioukas, S.G.; Klever, B.; Khalsa, S.D.S.; Brown, P.H.; Jin, Y. Tree-Level Almond Yield Estimation from High Resolution Aerial Imagery with Convolutional Neural Network. *Front. Plant Sci.* 2023, 14, 1070699. [CrossRef]
- 22. Cordeiro, V.; Monteiro, A. Almond growing in Trás-os-Montes Region (Portugal). Acta Hortic. 2002, 2002, 5. [CrossRef]
- Mirás-Avalos, J.M.; Gonzalez-Dugo, V.; García-Tejero, I.F.; López-Urrea, R.; Intrigliolo, D.S.; Egea, G. Quantitative Analysis of Almond Yield Response to Irrigation Regimes in Mediterranean Spain. *Agric. Water Manag.* 2023, 279, 108208. [CrossRef]
- 24. Esparza, G.; DeJong, T.M.; Weinbaum, S.A.; Klein, I. Effects of Irrigation Deprivation during the Harvest Period on Yield Determinants in Mature Almond Trees. *Tree Physiol.* **2001**, *21*, 1073–1079. [CrossRef]
- Kubota, T.; Shige, S.; Hashizume, H.; Aonashi, K.; Takahashi, N.; Seto, S.; Hirose, M.; Takayabu, Y.N.; Ushio, T.; Nakagawa, K.; et al. Global Precipitation Map Using Satellite-Borne Microwave Radiometers by the GSMaP Project: Production and Validation. *IEEE Trans. Geosci. Remote Sens.* 2007, 45, 2259–2275. [CrossRef]
- 26. Kubota, T.; Aonashi, K.; Ushio, T.; Shige, S.; Takayabu, Y.N.; Kachi, M.; Arai, Y.; Tashima, T.; Masaki, T.; Kawamoto, N.; et al. Global Satellite Mapping of Precipitation (GSMaP) Products in the GPM Era. In *Satellite Precipitation Measurement*; Levizzani, V., Kidd, C., Kirschbaum, D.B., Kummerow, C.D., Nakamura, K., Turk, F.J., Eds.; Advances in Global Change Research; Springer International Publishing: Cham, Switzerland, 2020; Volume 1, pp. 355–373, ISBN 978-3-030-24568-9.
- 27. Phan, T.N.; Kappas, M. Application of MODIS Land Surface Temperature Data: A Systematic Literature Review and Analysis. *J. Appl. Remote Sens.* **2018**, *12*, 041501. [CrossRef]
- 28. Acharya, T.; Yang, I. Exploring Landsat 8. Int. J. IT Eng. Appl. Sci. Res. 2015, 4, 4–10.
- 29. Phiri, D.; Simwanda, M.; Salekin, S.; Nyirenda, V.R.; Murayama, Y.; Ranagalage, M. Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sens.* 2020, 12, 2291. [CrossRef]
- Frazier, A.E.; Hemingway, B.L. A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery. *Remote Sens.* 2021, 13, 3930. [CrossRef]
- 31. Zhao, Q.; Yu, L.; Li, X.; Peng, D.; Zhang, Y.; Gong, P. Progress and Trends in the Application of Google Earth and Google Earth Engine. *Remote Sens.* **2021**, *13*, 3778. [CrossRef]
- 32. Ji, Z.; Pan, Y.; Zhu, X.; Zhang, D.; Wang, J. A Generalized Model to Predict Large-Scale Crop Yields Integrating Satellite-Based Vegetation Index Time Series and Phenology Metrics. *Ecol. Indic.* **2022**, *137*, 108759. [CrossRef]
- 33. Ma, C.; Liu, M.; Ding, F.; Li, C.; Cui, Y.; Chen, W.; Wang, Y. Wheat Growth Monitoring and Yield Estimation Based on Remote Sensing Data Assimilation into the SAFY Crop Growth Model. *Sci. Rep.* **2022**, *12*, 5473. [CrossRef] [PubMed]
- 34. Shammi, S.A.; Meng, Q. Use Time Series NDVI and EVI to Develop Dynamic Crop Growth Metrics for Yield Modeling. *Ecol. Indic.* 2021, 121, 107124. [CrossRef]
- Motohka, T.; Nasahara, K.N.; Oguma, H.; Tsuchida, S. Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology. *Remote Sens.* 2010, 2, 2369–2387. [CrossRef]
- Sanches, G.M.; Duft, D.G.; Kölln, O.T.; Luciano, A.C.d.S.; De Castro, S.G.Q.; Okuno, F.M.; Franco, H.C.J. The Potential for RGB Images Obtained Using Unmanned Aerial Vehicle to Assess and Predict Yield in Sugarcane Fields. *Int. J. Remote Sens.* 2018, 39, 5402–5414. [CrossRef]
- 37. Ji, Z.; Pan, Y.; Zhu, X.; Wang, J.; Li, Q. Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. *Sensors* 2021, 21, 1406. [CrossRef]
- 38. Panek, E.; Gozdowski, D. Analysis of Relationship between Cereal Yield and NDVI for Selected Regions of Central Europe Based on MODIS Satellite Data. *Remote Sens. Appl. Soc. Environ.* **2020**, *17*, 100286. [CrossRef]
- 39. Panda, S.S.; Ames, D.P.; Panigrahi, S. Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. *Remote Sens.* **2010**, *2*, 673–696. [CrossRef]
- 40. Satir, O.; Berberoglu, S. Crop Yield Prediction under Soil Salinity Using Satellite Derived Vegetation Indices. *Field Crops Res.* **2016**, 192, 134–143. [CrossRef]
- da Silva, E.E.; Rojo Baio, F.H.; Ribeiro Teodoro, L.P.; da Silva Junior, C.A.; Borges, R.S.; Teodoro, P.E. UAV-Multispectral and Vegetation Indices in Soybean Grain Yield Prediction Based on in Situ Observation. *Remote Sens. Appl. Soc. Environ.* 2020, 18, 100318. [CrossRef]
- 42. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
- 43. Fraga, H.; Guimarães, N.; Santos, J. Vintage Port prediction and climate change scenarios. OENO One 2023, 57. [CrossRef]
- Shen, M.; Luo, J.; Cao, Z.; Xue, K.; Qi, T.; Ma, J.; Liu, D.; Song, K.; Feng, L.; Duan, H. Random Forest: An Optimal Chlorophyll-a Algorithm for Optically Complex Inland Water Suffering Atmospheric Correction Uncertainties. *J. Hydrol.* 2022, 615, 128685. [CrossRef]
- Fazakis, N.; Kostopoulos, G.; Karlos, S.; Kotsiantis, S.; Sgarbas, K. Self-Trained eXtreme Gradient Boosting Trees. In Proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), Patras, Greece, 15–17 July 2019; pp. 1–6.
- 46. Singh, U.; Rizwan, M.; Alaraj, M.; Alsaidan, I. A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments. *Energies* **2021**, *14*, 5196. [CrossRef]
- 47. Yu, N.; Haskins, T. Bagging Machine Learning Algorithms: A Generic Computing Framework Based on Machine-Learning Methods for Regional Rainfall Forecasting in Upstate New York. *Informatics* **2021**, *8*, 47. [CrossRef]

- 48. Priestly, S.E.; Raimond, K.; Cohen, Y.; Brema, J.; Hemanth, D.J. Evaluation of a Novel Hybrid Lion Swarm Optimization— AdaBoostRegressor Model for Forecasting Monthly Precipitation. *Sustain. Comput. Inform. Syst.* **2023**, *39*, 100884. [CrossRef]
- 49. Kramer, O. Scikit-Learn. In *Machine Learning for Evolution Strategies*; Kramer, O., Ed.; Studies in Big Data; Springer International Publishing: Cham, Switzerland, 2016; pp. 45–53, ISBN 978-3-319-33383-0.
- 50. Isabona, J.; Imoize, A.L.; Kim, Y. Machine Learning-Based Boosted Regression Ensemble Combined with Hyperparameter Tuning for Optimal Adaptive Learning. *Sensors* 2022, 22, 3776. [CrossRef]
- 51. Prairie, Y.T. Evaluating the Predictive Power of Regression Models. Can. J. Fish. Aquat. Sci. 1996, 53, 490–492. [CrossRef]
- 52. Chai, T.; Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against Avoiding RMSE in the Literature. *Geosci. Model Dev.* 2014, 7, 1247–1250. [CrossRef]
- Nikolaou, G.; Neocleous, D.; Christou, A.; Kitta, E.; Katsoulas, N. Implementing Sustainable Irrigation in Water-Scarce Regions under the Impact of Climate Change. Agronomy 2020, 10, 1120. [CrossRef]
- 54. Tamimi, J.Z.A. Effects of Almond Milk on Body Measurements and Blood Pressure. Food Nutr. Sci. 2016, 7, 466–471. [CrossRef]
- 55. Tombesi, S.; Scalia, R.; Connell, J.; Lampinen, B.; Dejong, T.M. Fruit Development in Almond Is Influenced by Early Spring Temperatures in California. *J. Hortic. Sci. Biotechnol.* **2010**, *85*, 317–322. [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.