


## Article

# Early Estimation of Tomato Yield by Decision Tree Ensembles

Mario Lillo-Saavedra <sup>1,\*</sup>, Alberto Espinoza-Salgado <sup>1</sup>, Angel García-Pedrero <sup>2</sup>, Camilo Souto <sup>1,3</sup>, Eduardo Holzapfel <sup>1</sup>, Consuelo Gonzalo-Martín <sup>2</sup>, Marcelo Somos-Valenzuela <sup>4</sup> and Diego Rivera <sup>5</sup>

- <sup>1</sup> Facultad de Ingeniería Agrícola, Universidad de Concepción, Chillán 3812120, Chile  
<sup>2</sup> Department of Computer Architecture and Technology, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Spain  
<sup>3</sup> Department of Horticulture, Oregon State University, 4017 Agricultural and Life Sciences Building, Corvallis, OR 97331, USA  
<sup>4</sup> Department of Forest Sciences, Faculty of Agriculture and Environmental Sciences, Universidad de La Frontera, Av. Francisco Salazar 01145, Temuco 4780000, Chile  
<sup>5</sup> Centro de Sustentabilidad y Gestión Estratégica de Recursos (CiSGER), Facultad de Ingeniería, Universidad del Desarrollo, Las Condes, Santiago 7610658, Chile  
\* Correspondence: malillo@udec.cl

**Abstract:** Crop yield forecasting allows farmers to make decisions in advance to improve farm management and logistics during and after harvest. In this sense, crop yield potential maps are an asset for farmers making decisions about farm management and planning. Although scientific efforts have been made to determine crop yields from in situ information and through remote sensing, most studies are limited to evaluating data from a single date just before harvest. This has a direct negative impact on the quality and predictability of these estimates, especially for logistics. This study proposes a methodology for the early prediction of tomato yield using decision tree ensembles, vegetation spectral indices, and shape factors from images captured by multispectral sensors on board an unmanned aerial vehicle (UAV) during different phenological stages of crop development. With the predictive model developed and based on the collection of training characteristics for 6 weeks before harvest, the tomato yield was estimated for a 0.4 ha plot, obtaining an error rate of 9.28%.

**Keywords:** decision tree ensemble; crop yield; UAVs



**Citation:** Lillo-Saavedra, M.; Espinoza-Salgado, A.; García-Pedrero, A.; Souto, C.; Holzapfel, E.; Gonzalo-Martín, C.; Rivera, D. Early Estimation of Tomato Yield by Decision Tree Ensembles. *Agriculture* **2022**, *12*, 1655. <https://doi.org/10.3390/agriculture12101655>

Academic Editor: Nen Fu Huang and Ho-Hsien Chen

Received: 11 September 2022

Accepted: 29 September 2022

Published: 10 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

## 1. Introduction

Tomato (*Solanum lycopersicum*) is the most produced vegetable worldwide, with a volume of 187 MTons in 2020 [1]. The varieties differ in their exterior appearance (shape and color) and internal traits (flavor, texture, and hardness). There are also varieties intended for fresh consumption and others for agroindustrial processing. In particular, processed tomatoes reached annual volumes close to 40 MTons during 2021 [2].

With the systematic increase in the area devoted to tomato cultivation, there is growing research to test new varieties and related agronomic management practices to improve quality and yield [3,4]. Yield estimation is critical for crop management [5]. Timely and reliable estimates enable better management decisions, such as irrigation, fertilizers, and pesticides. In addition, it supports the logistic planning of field operations (stock scheduling, labor) and post-harvest (storage, handling, packing, and anticipated product sales) [6]. Therefore, the generation of potential yield maps would allow farmers to make evidence-based decisions for crop management and planning [3,7,8].

In this way, precision agriculture plays an important role. For example, unmanned aerial vehicles (UAVs) monitor crops in time and space [9–11]. UAVs can fly at low altitudes and often enough to provide information throughout the phenological development period of a crop [12–15]. However, their main disadvantage is that they require qualified operators to perform flights and data post-processing [10,16,17].

Although in recent years some work has used sensors onboard UAVs as a data source to estimate tomato yield, the literature on this subject remains sparse. However, it is important to highlight various focus methods used by some researchers. For example, Senthilnath et al. [18] detected the number of fruits in tomato plants in an advanced stage of development, where plants lose their leaves, using a 5-megapixel (5 Mpx) RGB camera mounted on a UAV that flew at a height of 20 m. Johansen et al. [19] estimated the production of individual tomato plants using a 12.4 Mpx RGB camera on board a UAV flying at a height of 13 m, covering a time record during the last eight weeks before harvest (WBH). Subsequently, the authors developed a machine-learning-based predictive model, obtaining the best results for dates near harvest, with *RMSE* of 168.9 g and 175.7 g in one and four WBH, respectively. However, the results for 8, 7, and 6 WBH significantly underestimated the yield. Similar results were obtained by Johansen et al. [20]. Ashapure et al. [3] estimated the production of tomato plant groups using 14 Mpx RGB cameras on board a UAV during the 2016 season and 20 Mpx during the 2017 and 2018 growing seasons to perform a multitemporal plant development characterization. On flight days, it also collected meteorological information: relative humidity, precipitation, air temperature, solar radiation, and crop evapotranspiration. Based on the data captured in the flights and meteorological information, they developed a predictive model based on a radial basis function neural network. Initially, the model was trained independently for the growing seasons 2016, 2017, and 2018, achieving a performance estimation model with a goodness of fit ( $R^2$ ) between 0.78 and 0.89. They also trained a model using data from the 2016 growing season that successfully predicted performance for data sets from the 2017 and 2018 growing seasons with  $R^2 \geq 0.70$ . However, in their results and conclusions, there is no measure of error to quantify the magnitude of difference between real and estimated production. Enciso et al. [13] validated measurements on high grounds of plants, canopy cover, and normalized difference vegetation index (NDVI), using an RGB camera with a resolution of 20 Mpx and a multispectral camera with 1.2 Mpx on board a UAV, flying at 30 m over the field and compared the results to field data (NDVI was measured using a Green Seeker Handheld Crop Sensor, Trimble). The results showed that there are no significant differences ( $p \geq 0.05$ ) between field measurements and UAV data for the variable height of the plant and the canopy cover. However, significant differences were found when comparing the NDVI from field measurements with the data obtained by the UAV. The authors attributed this low correlation to wavelength differences between the Green Seeker instrument and the multispectral camera. Tatsumi et al. [21], using multiple machine learning algorithms, predicted the masses of the fresh tomato shoots, the weight of the fruits, and the number of fruits. The authors collected 10 RGB and multispectral images by a UAV on two dates 3 months and 3 days before harvest during the 2020 tomato growing season. From these images, first- and second-order statistics were extracted for each plant. The prediction accuracy of shoot mass, fruit weights, and number of fruits by models constructed from all variables (*RMSE* = 8.8–28.1%) was better than that from first-order statistics (*RMSE* = 10.0–50.1%).

Flight heights below 20 m improve the spatial resolution of the images [3,19,20]. Low-altitude flights are an important drawback in developing an operational methodology to predict crop yield, as they involve more overflights to cover larger areas. To overcome this limitation, fixed-wing UAVs have greater autonomy and can cover large areas in a single flight [22].

It should be noted that the aforementioned studies evaluated the crop on certain dates without granting any importance to the effect of the cumulative behavior of the variables measured throughout the phenological development of the plant (temporal behavior), which could provide valuable information for estimating production. However, to our knowledge, the question of when and how many flights or field campaigns suffice for accurate estimations has not been answered.

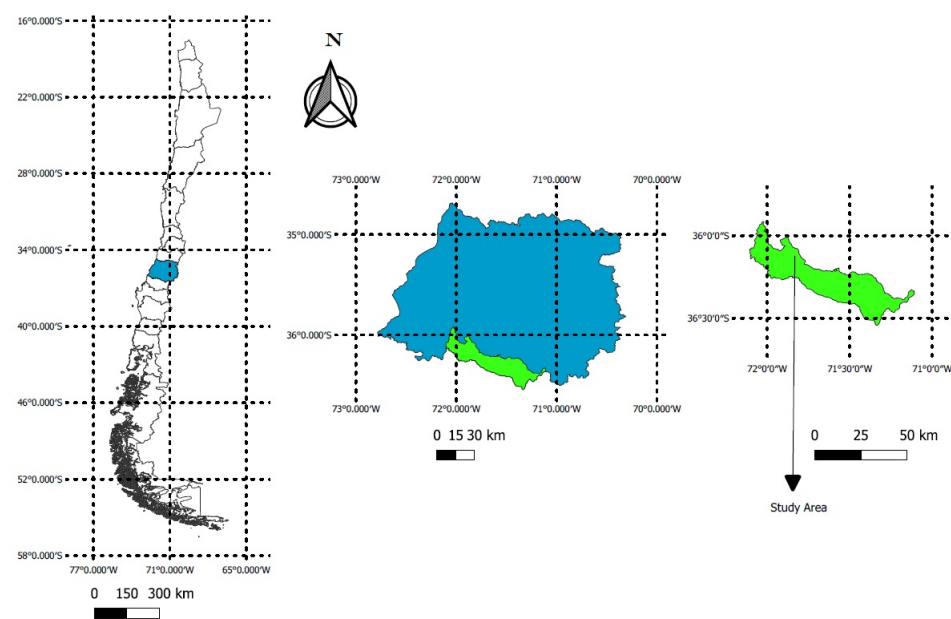
Taking into account the advantages and limitations exposed in the current state of the art, the objective of this study is to propose a methodology for the early estimation of

tomato production that combines decision trees ensembles (DTE) [23], spectral vegetation indices, and form factors (FF) [24] captured at various dates to maximize accuracy while minimizing monitoring costs.

## 2. Materials and Methods

### 2.1. Field Trials

The study site is located in Parral, Maule Region, Chile ( $36^{\circ}06' \text{ S}$ ,  $71^{\circ}50' \text{ W}$ ) (Figure 1), at 137 m above sea level (a.s.l.). The study site belongs to the warm-temperate suprathreshold agroclimatic zone, with a semiarid humidity regimen. Its annual median precipitation is 720 mm, and during vegetative periods its average maximum and minimum air temperatures are  $29.7$  and  $7.6^{\circ} \text{ C}$ , respectively [25]. The soil texture is clay loam. The root depth reached 0.3 m, choosing this depth for effective water management.



**Figure 1.** The geographic location of the study site. Parral, Maule Region, Chile (lat.  $71^{\circ}49'53.29'' \text{ W}$ , long.  $36^{\circ}5'56.21'' \text{ S}$ ).

We carried out a varietal improvement experiment on a plot 0.4 ha containing 88 parcels (22 varieties of tomatoes processed with four repetitions) considering a random parcel distribution (Figure 2). Each parcel was made up of 3 rows of 6 m, with a planting frame over and between rows of 0.25 m and 1.5 m, respectively. To avoid border effects at the boundaries of the east and west rows, a filler row was added on both sides of the experiment [13]. The tomato plants were transplanted on 8 October 2019 and harvested on 20 February 2020. The irrigation system used was drip irrigation with 0.2 m between the emitters on the lateral and 1.5 m between the irrigation laterals (one lateral per row of tomatoes). The flow per emitter was  $1.1 \text{ L} \cdot \text{h}^{-1}$ . The irrigation frequency was daily, and the irrigation time varied according to the water demand of the crop monitored by a reference meteorological station near the study site.

### 2.2. Images Acquisition from UAV

For image acquisition, the UAV Ebee SQ platform was used, equipped with a multi-spectral Parrot Sequoia camera [26]. The Parrot camera records spectral wavelengths in green (530–570 nm), red (640–680 nm), red edge (730–740 nm), and near-infrared (Nir) (770–810 nm) and has a resolution of 1.2 Mpx. Before each flight, the Parrot Sequoia camera was radiometrically calibrated using the AIRINOV calibration panel [27].



**Figure 2.** Experimental design that consider 22 varietales, and 4 repetitions, using a planting frame of 3 rows of 6 m, with an above and between rows of 0.25 m and 1.5 m, respectively, on a 0.4 ha farm.

Table 1 shows the flight planning for the Ebee SQ parameters in the EmotionAg software. Data collection dates are presented in Table 2. Five ground control points (GCP) were established at the study site (Figure 3), four at the corners and one at the center of the plot. The GCPs were kept fixed throughout the study period [4,13] for the orthomosaic generation process in each of the flights performed during the study period. Orthomosaics were generated by Pix4Dmapper software following the developer's recommendations (<https://support.pix4d.com/hc/en-us/articles/202557359-Getting-Started-Index>, accessed on 2 August 2021).

**Table 1.** Parameters for flight planning.

Parameter	Ebee SQ
Camera	Parrot Sequoia
Flight Height	42.5 m
Lateral Overlap	80%
Vertical Overlap	80%
Number of Images Per Flight	41 per band
Spatial Resolution	4 cm

**Table 2.** Data collection dates.

Flight	Date	Day after Transplant	Weeks before Harvest (WBH)
1	21 November, 2019	44	12
2	30 November 2019	53	11
3	11 December 2019	64	10
4	11 January 2020	95	5
5	25 January 2020	109	3
6	29 January 2020	113	3
7	5 February 2020	120	2

### 2.3. Agronomic Measurements

Agronomic measurements were obtained on harvest day (20 February 2020). The criterion for the day of harvest is when 80% of the fruits reached the color of commercial ripeness and the average sugar content of an aqueous solution of  $^{\circ}\text{Brix} \geq 5$ . The harvest measurements were made on 2 m length in the middle row of each parcel, as shown in Figure 4. The data collected were: (i) yields, total weight of all fruits with caliber



greater than 30 mm; and (ii) counting by hand of red, turning, and green fruits, ruling out low-caliber fruits (<30 mm) [20].



**Figure 3.** Location of the ground control points (GCP).



**Figure 4.** Harvest strategy: 2 m on a line in the middle row of each of the parcels.

#### 2.4. Images Segmentation

To separate the zones with tomato plants (tomato plant class, *TP*) from bare soil (no tomato plant class, *NTP*), the study area was segmented for different phenological stages. Since image segmentation is a time-consuming task when performed manually and also depends on the operator performing the segmentation, the process was automated. In this work, the proposed segmentation process used the normalized vegetation index (*NDVI*) [28] from images captured by the multispectral Parrot Sequoia camera.

$$NDVI = \frac{(Nir - Red)}{(Nir + Red)} \quad (1)$$

Each of the *NDVI* maps was oversegmented with the superpixel (SP) technique known as SLIC (simple linear iterative clustering) [28,29]. To adjust the segmentation, we generated 10% of SP from the total image pixels and set the compactness to 2, that is, the most irregular form that SP can have. Then, multilevel thresholds using the Otsu thresholding method were applied to SP histograms [30]. The thresholds were selected using the maximum variation method between classes [31] to separate the *TP* and *NTP* classes. Finally, a small object elimination process was performed to avoid noise in class assignments via an area threshold. In [4,19], the authors reported that the minimum area used to define a tomato plant was <150 cm<sup>2</sup>. This value becomes inadequate in close flights at the transplant stage since small tomato plants are not detected or considered. This may have a negative impact

on the crop yield forecast. Therefore, in our work, all tomato plants with areas  $<40 \text{ cm}^2$  were eliminated.

To assess the ability of the implemented algorithm, the segmented images were compared with manually segmented images obtained for the same dates using *Dice* coefficient (Equation (2)) [32]. The *Dice* similarity coefficient is used as a statistical validation of both the reproducibility of manual segmentation and the spatial overlap accuracy of automated segmentation methods. The values of *Dice* range from 0, indicating no spatial overlap between two sets of binary segmentation results, to 1, indicating complete overlap.

$$Dice = \frac{2 \times |X \cap Y|}{X + Y} \quad (2)$$

where  $X$  is the predicted set of pixels, and  $Y$  is the ground truth.

### 2.5. Data Sets

The data sets contain information on the production measured in the harvest zones (Figure 4): vegetation indices obtained for each segment on each flight date (Table 2), mean reflectance values (obtained from the Green, Red, Red Edge, and Nir bands of the Parrot Sequoia camera), and form factors ( $FF$ ).

Considering the results obtained in Ramos et al. [33], in this work  $NDVI$  (Equation (1)), and normalized difference red edge index ( $NDRE$ ) (Equation (3)) were used [34,35].

$$NDRE = \frac{(Nir - Red\ Edge)}{(Nir + Red\ Edge)} \quad (3)$$

and the form factors ( $FF$ ): canopy cover ( $FFC$ ) (Equation (4)) [9], size ( $FFS$ ) (Equation (5)), and density ( $FFD$ ) (Equation (6)) [24]. These form factors are positively correlated with tomato production [36].

$$FFC = \frac{A_{TP}}{HA} \quad (4)$$

$$FFS = \frac{A_{TP}}{P_{TP}} \quad (5)$$

$$FFD = \frac{P_{TP}^2}{A_{TP}} \quad (6)$$

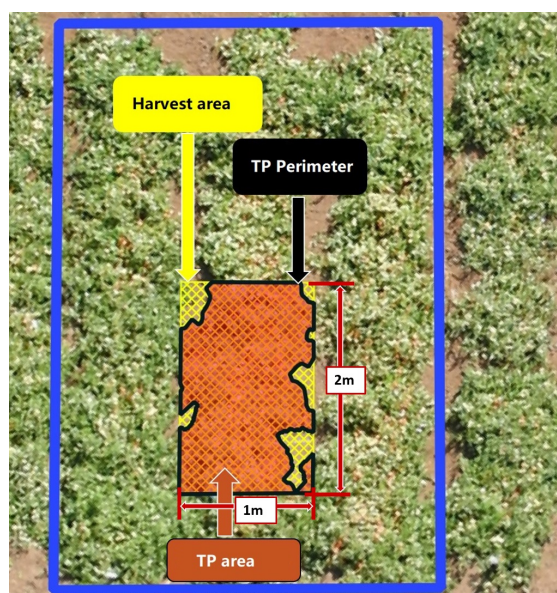
where  $A_{TP}$  and  $P_{TP}$  are the area and perimeter of the tomato plant class, respectively, and  $HA$  is the harvest area. These descriptors are shown in Figure 5

Later, determinations were made about the cumulative behavior of attributes, through their temporal integration for a time window given by the number of flights considered (Equation (7)). This allowed us to obtain the  $FFC'$ ,  $FFS'$ , and  $FFD'$  attributes, and with the accumulated value of the spectral bands, the  $NDVI'$  and  $NDRE'$  vegetation indices were obtained.

$$\Phi' = \sum_{i=1}^n \Phi_i \quad (7)$$

where  $n$  is the flight until the cumulative attribute value is desired. and  $\Phi$  represents the attribute values for a regular flight in the harvest zone. Before training the models, the attributes were scaled between 0 and 1 to avoid some attributes having more weight in the construction of the model.

With the characteristics  $NDVI'$ ,  $NDRE'$ ,  $FFC'$ ,  $FFS'$ , and  $FFD'$ , 3 training sets were created, for time windows of 6, 4, and 2 weeks before harvest (WBH). These training sets were used to forecast the yield of processing tomato through a DTE [37]. With the additional goal of comparing the strategy of using training sets with characteristics accumulated over time, results were obtained for training sets made up of the same characteristics ( $NDVI$ ,  $NDRE$ ,  $FFC$ ,  $FFS$ , and  $FFD$ ) for specific dates, that is, on flight dates previous to 6 WBH, 4 WBH, and 2 WBH (Table 2).



**Figure 5.** Harvest area and their components

## 2.6. Forecasting Models for Processing Tomato Yield

A DTE algorithm was used to generate forecasting models for tomato yield. Models trained on this algorithm were used by other authors in production for both processing tomato [19,20] and other crops [33,38,39]. The strategy of generating models using DTE consists of obtaining a set of individual decision trees (DT), each of which is trained with a sample slightly different from the training data. The prediction of a new observation is obtained by adding the predictions of all individual trees that form the model. For a more detailed description of the algorithm, we recommend the work presented in [33,37]. Bagging and boosting are common ways to create ensembles of DTs. Bagging is a technique used to reduce the variation of predictions via a combination of results from various classifiers, each modeled with different subsets taken from the same population. Boosting consists of changing the results from various weak classifiers to obtain a robust classifier [33].

The advantage of DTE models is that they are robust regarding atypical values and noise apart from not overfitting as more trees are added, rather they produce an error generalization limit value. This generalization error depends on the force of the individual trees in the ensemble and the correlation between them, that is, the precision of the individual classifiers and the dependence between them [37].

To create predictive models, we tested two different DTE algorithms: bagging (DTE-Bag) and boosting (DTE-Boost). Input yield data were converted to quartiles to normalize among varieties. From each quartile, data were randomly selected for training, test, and validation sets; 70% of the data from each quartile was used to train and calibrate the model, with a distribution of 70% for training and 30% for calibration, while the remaining total of 30% of the data from each quartile was used for model testing.

Each tree of the training group was created using 80% of the characteristics, which were randomly chosen. This process is with replacement; that is, the same characteristic can be used many times, generating a random vector, independent of previous random vectors, but with the same distribution [37]. To analyze the test error, we used *RMSE* (root-mean-square error). The number of trees needed for the model to achieve convergence was also defined, and the most important predictors were determined within the characteristic entered into the model obtained from the DTE algorithm [37].

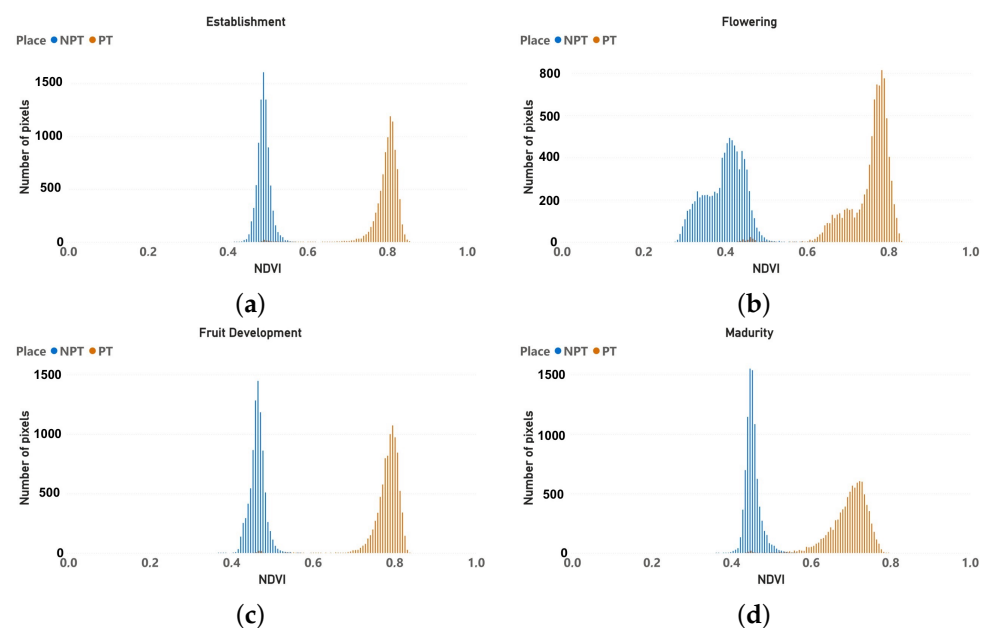
## 3. Results and Discussion

Based on the segmentation of *NDVI* using the *SLIC* algorithm, Figure 6 shows the histograms that highlight the and *NTP* classes for different stages of the phenological devel-

opment of the crops. Table 3 reports the quality of the segmented images. Dice's similarity criterion obtained from segmenting all the images was on average 0.955. The greatest similarities occurred for the maximum phenological development periods of plants (flowering, fruit development, and maturity) from late December to early February. For the dates with available images, the greatest similarity occurred 113 days after transplant (Figure 7a,b). The lowest similarity occurred 44 days after transplant (Figure 7c,d) when plants had a small coverage area, so even small classification errors impacted the final results.

**Table 3.** Quality assessment of the segmentation algorithm. Comparison between manual and automatic NDVI map segmentation for different tomato phenological stages.

Date	Phenological Stage	Dice
21 November 2019	Establishment	0.896
30 November 2019	Vegetative Growth	0.938
11 December 2019	Flowering	0.947
11 January 2020	Fruit Development	0.976
25 January 2020	Fruit Development	0.978
29 January 2020	Maturity	0.979
05 February 2020	Maturity	0.971
20 February 2020	Maturity	0.957

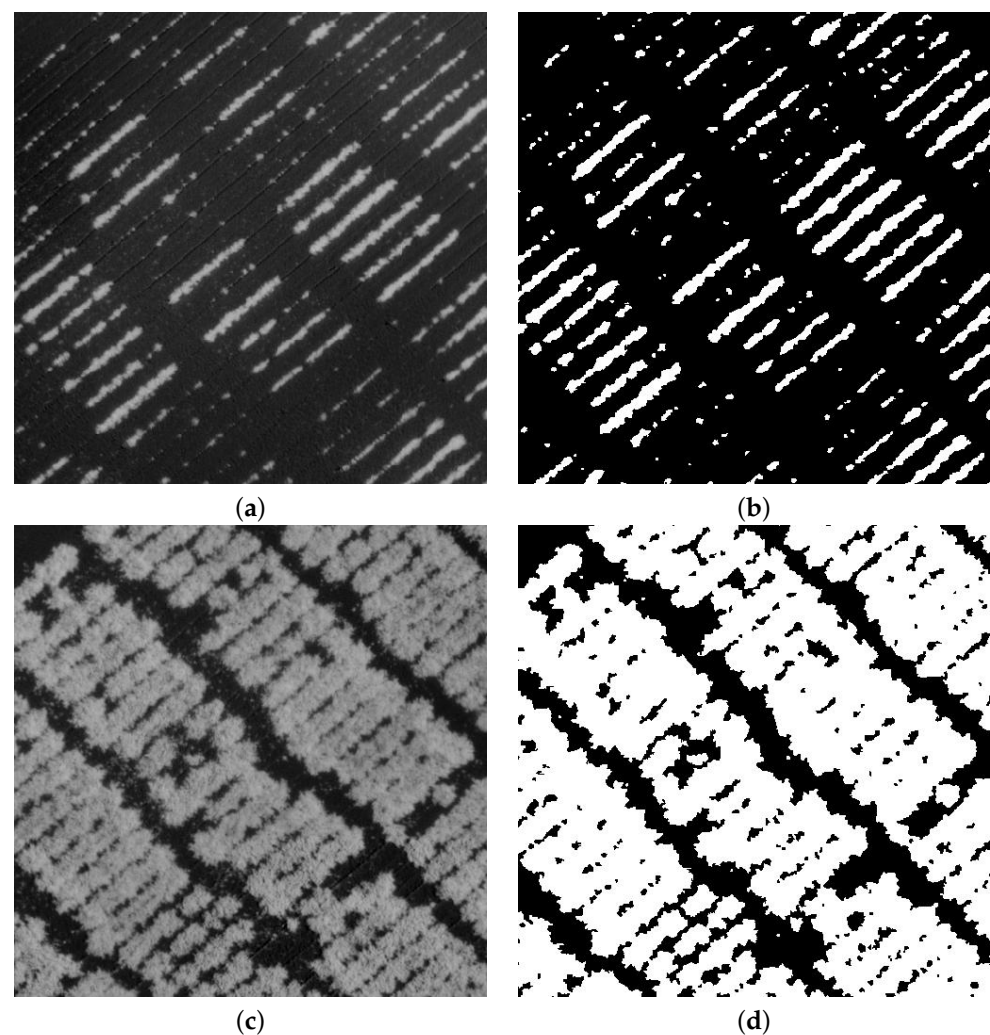


**Figure 6.** Histogram of NDVI maps segmented via SLIC for different tomato phenological conditions: (a) establishment (Vegetative growth) (30 November 2019); (b) flowering (11 December 2019); (c) fruit development (11 January 2020); and (d) maturity (5 February 2020).

The proposed segmentation algorithm allows automatic segmentation of tomato plants, thereby avoiding the intervention of trained personnel to correct classification errors. Refs. [4,19] required manual plant delineation to correct classification errors in the eCognition Developer 9.3 program, and [13] used the Canopedo algorithm [40] to differentiate background pixel plants via a threshold.

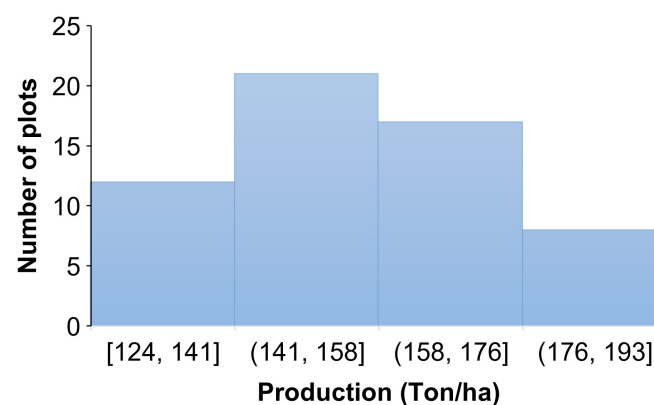
To obtain predictive tomato yield models using the DTE-Bag and DTE-Boost methods, the number of parcels that met the requirement of having a linear 2 m for harvest was analyzed (Section 2.3). Of the 88 parcels, 58 met this requirement; therefore, our training set had 58 observations.





**Figure 7.** NDVI maps and their segmented image (*TP* and *NTP*) for date: 21 November 2019 (a,b), 44 days after transplant, and 29 January 2020 (c,d), 113 days after transplant.

The tomato production histogram of 58 parcels for developing the forecasting model is shown in Figure 8. The predictive models obtained via DTE-Bag converged for more than 30 decision trees for both approaches, i.e., characteristics sets accumulated over time and the sets of specific characteristics. For DTE-Boost, convergence occurred in the 70-tree group for both sets of characteristics accumulated over time and in the sets of specific characteristics.



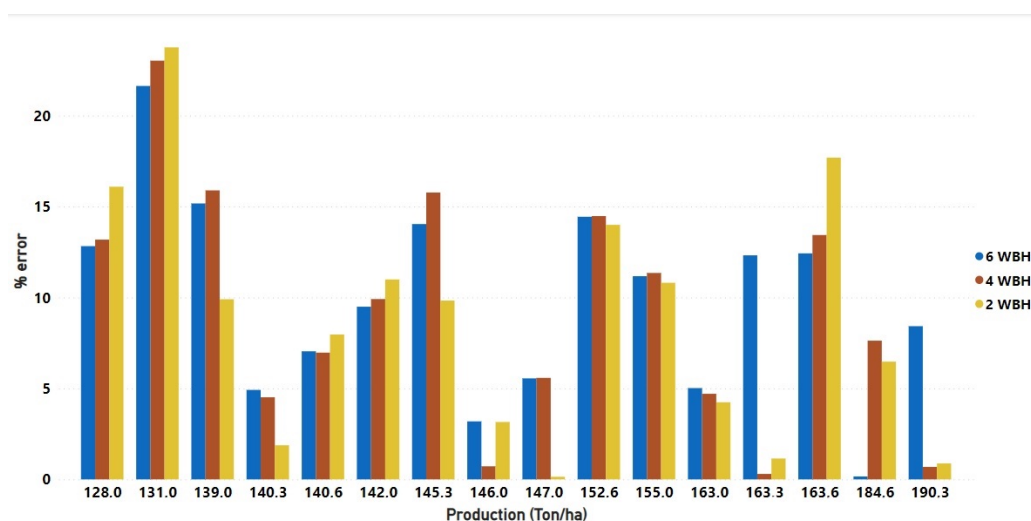
**Figure 8.** The tomato production histogram for the 58 parcels used in the training and validation of the production model based on DTE-Bag.

The results in Table 4 show that with the combination of characteristics accumulated over time and the DTE-Bag method, the *RMSE* of the models decreased as the evaluation dates approached the harvest date. This behavior occurred only when DTE-Bag was used with accumulated characteristics, maintaining an error percentage at <10%. Thus, models trained with accumulated characteristic sets and the DTE-Bag method obtained the best results in forecasting tomato yield at harvest. The model for 2 WBH had an *RMSE* of 12.87 Ton/ha, while the *RMSE* of the model created for 6 WBH was 14.38 Ton/ha, so an estimate at 2 WBH had a benefit of only 1.5 Ton/ha in *RMSE*, compared to an estimate at 6 WBH that allows more time for logistic planning.

**Table 4.** Forecasting tomato yield for models generated via DTE-Bag and DTE-Boost for 6, 4, and 2 WBH.

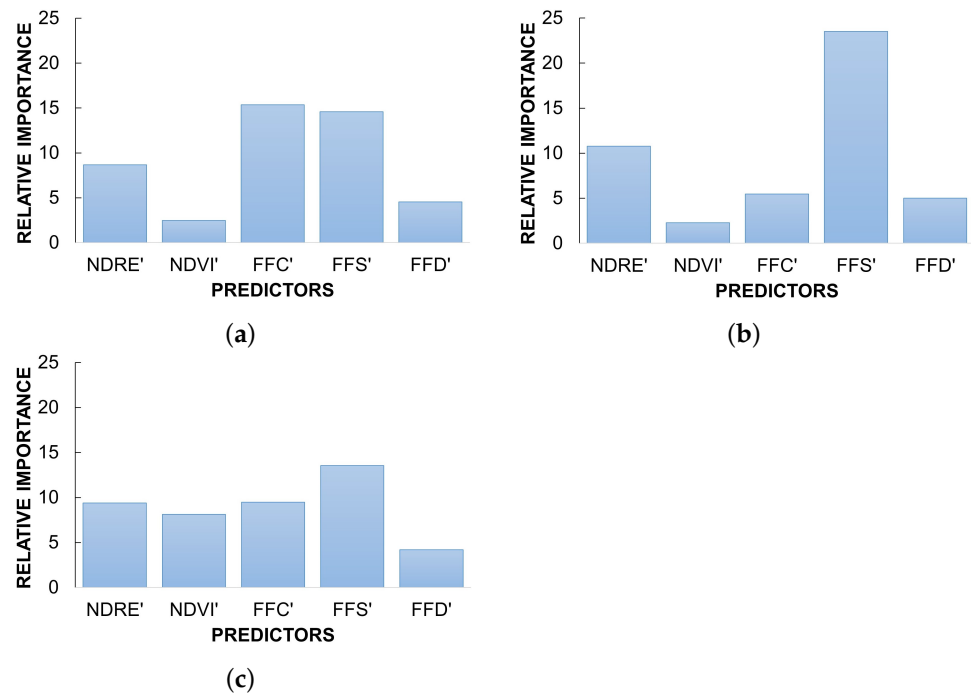
	Accumulated Characteristics		Specific Characteristics	
	DTE-Bag	DTE-Boost	DTE-Bag	DTE-Boost
6 WBH				
<i>RMSE</i> (Ton/ha)	14.38	16.65	14.07	13.34
Percentage error	9.28%	10.82%	8.86 %	8.5 %
Standard deviation (Ton/ha)	9.86	10.61	11.17	8.57
<i>RMSE</i> maximum (Ton/ha)	41.16	40.09	42.29	30.43
4 WBH				
<i>RMSE</i> (Ton/ha)	13.65	13.22	15.67	16.13
Percentage Error	8.81%	8.58%	10.14%	10.26%
Standard deviation (Ton/ha)	11.09	10.67	8.25	13.20
<i>RMSE</i> maximum (Ton/ha)	43.81	36.88	29.28	41.52
2 WBH				
<i>RMSE</i> (Ton/ha)	12.87	14.47	14.30	17.05
Percentage error	8.17%	9.14%	9.26%	11.33%
Standard deviation (Ton/ha)	11.11	13.13	10.80	13.09
<i>RMSE</i> Maximum (Ton/ha)	45.20	53.75	41.91	38.72

In a more detailed analysis of the estimation errors obtained with DTE-Bag and the attribute sets accumulated for validation parcels, we observed that the worst predictions occurred when production levels were below the high ranges (Figure 9). This is due to training data imbalances, as they are mostly close to the average value (154.8 Ton/ha), thus impeding learning for extreme production values (Figure 8).



**Figure 9.** Percentage error in the forecast of tomato yield using the test set.

In Figure 10a–c, we see the characteristics accumulated over time for 6, 4, and 2 WBH, which were the most influential in generating the production model using the DTE-Bag method. We demonstrated that the characteristics that contributed the most information to the model were *FFS'*, *NDRE'*, and *FFC'*. This result is consistent with that obtained by [20], which found that the leaf area of the plant allowed the identification of the best-performing plants.



**Figure 10.** Characteristics accumulated which most influenced forecasting crop yield model generation using the DTE-Bag method, for: (a) 6; (b) 4; and (c) 2 WBH.

In Table 5 the mean values of accumulative attributes for the low (124–141 (Ton/ha)) and high (176–193 (Ton/ha)) range production (Figure 8) are shown. As can be observed, unlike the other four attributes, the *NDVI'* value did not vary for 6, 4, and 2 WBH for both low and high production ranges. This indicates that it is not a good attribute to consider in the training of the predictive model. This is consistent with the result shown in Figure 10.

**Table 5.** Mean value of accumulative attributes for low (124–141 (Ton/ha)) and high 176–193 (Ton/ha) ranges of production, for 6, 4, and 2 WBH.

Attribute	Low Range Production			High Range Production		
	6 WBH	4 WBH	2 WBH	6 WBH	4 WBH	2 WBH
<i>NDVI'</i>	0.56	0.59	0.60	0.71	0.72	0.70
<i>NDRE'</i>	0.51	0.55	0.60	0.68	0.70	0.74
<i>FFC'</i>	0.36	0.52	0.83	0.42	0.59	0.91
<i>FFS'</i>	0.34	0.48	0.76	0.40	0.57	0.86
<i>FFD'</i>	0.47	0.57	0.80	0.39	0.47	0.68

#### 4. Conclusions

Based on the results obtained in this work, we demonstrated that it is possible to forecast tomato yield using an approach based on DTE using information obtained from a UAV.

The cumulative behavior in time of the characteristics used to train a DTE-Bag was useful information for an early forecast of tomato yield.

We were able to forecast tomato yield 6 weeks before harvest, with a percentage error of 9.28% compared to the average production volume (154.5 Ton/ha). We also established that when estimating production 2 weeks before harvest, the RMSE decreases only by 1.5 Ton/ha on average. Therefore, estimating 6 weeks before harvest can directly impact corrective measures for field management during the phenological development of the crop. This study also showed that an early forecast of tomato yield is possible based on a harvest area and not only on individual plants, thus avoiding the need to distance tomato plants from each other to perform estimates. Taking into account our findings and the fact that the flight height in this project was 42.5 m, the developed methodology can be scaled from a small study site to a large-scale plantation.

**Author Contributions:** Conceptualization, M.L.-S., C.S. and E.H.; methodology, M.L.-S., A.E.-S. and A.G.-P.; software, A.G.-P. and A.E.-S.; validation, A.G.-P., A.E.-S. and M.L.-S.; data curation, M.L.-S., A.G.-P.; writing—original draft preparation, M.L.-S. and A.E.-S.; writing—review and editing, M.L.-S., D.R., C.S., M.S.-V., E.H. and C.G.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Water Research Center For Agriculture and Mining, CRHIAM (ANID/FONDAP/15130015).

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

**Sample Availability:** Samples of the compounds ... are available from the authors.

## References

1. FAOSTAT. Foundation and Agricultural Organization of the United States. 2022. Available online: <https://www.fao.org/faostat/en/#data/QI> (accessed on 2 August 2022).
2. The World Processing Tomato Council. WPTC: 2021 Crop Estimated at 38.7 Million Tonnes. 2021. Available online: [https://www.tomatonews.com/en/wptc-2021-crop-estimated-at-387-million-tonnes\\_2\\_1489.html](https://www.tomatonews.com/en/wptc-2021-crop-estimated-at-387-million-tonnes_2_1489.html) (accessed on 2 August 2022).
3. Ashapure, A.; Oh, S.; Marconi, T.G.; Chang, A.; Jung, J.; Landivar, J.; Enciso, J. Unmanned aerial system based tomato yield estimation using machine learning. In Proceedings of the Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV, Baltimore, MD, USA, 15–16 April 2019; Volume 11008, p. 110080.
4. Johansen, K.; Morton, M.J.L.; Malbeteau, Y.M.; Aragon, B.; Al-Mashharawi, S.K.; Ziliani, M.G.; Angel, Y.; Fiene, G.M.; Negrão, S.S.C.; Mousa, M.A.A.; et al. Unmanned Aerial Vehicle-Based Phenotyping Using Morphometric and Spectral Analysis Can Quantify Responses of Wild Tomato Plants to Salinity Stress. *Front. Plant Sci.* **2019**, *10*, 370. [CrossRef] [PubMed]
5. Jongeneel, R.; Gonzalez-Martinez, A.R. Estimating crop yield supply responses to be used for market outlook models: Application to major developed and developing countries. *NJAS-Wageningen. J. Life Sci.* **2020**, *92*, 100327. [CrossRef]
6. Robson, A.; Rahman, M.M.; Muir, J. Using worldview satellite imagery to map yield in avocado (Persea americana): A case study in Bundaberg, Australia. *Remote. Sens.* **2017**, *9*, 1223. [CrossRef]
7. Gao, F.; Zhang, X. Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. *J. Remote. Sens.* **2021**, *2021*, 8379391. [CrossRef]
8. Wei, M.C.F.; Maldaner, L.F.; Ottoni, P.M.N.; Molin, J.P. Carrot yield mapping: A precision agriculture approach based on machine learning. *AI* **2020**, *1*, 229–241. [CrossRef]
9. Cuaran, J.; Leon, J. Crop monitoring using unmanned aerial vehicles: A review. *Agric. Rev.* **2021**, *42*, 121–132. [CrossRef]
10. Velusamy, P.; Rajendran, S.; Mahendran, R.K.; Naseer, S.; Shafiq, M.; Choi, J.G. Unmanned Aerial Vehicles (UAV) in precision agriculture: Applications and challenges. *Energies* **2021**, *15*, 217. [CrossRef]
11. Deng, L.; Mao, Z.; Li, X.; Hu, Z.; Duan, F.; Yan, Y. UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. *ISPRS J. Photogramm. Remote. Sens.* **2018**, *146*, 124–136. [CrossRef]
12. Yang, Q.; Shi, L.; Han, J.; Chen, Z.; Yu, J. A VI-based phenology adaptation approach for rice crop monitoring using UAV multispectral images. *Field Crop. Res.* **2022**, *277*, 108419. [CrossRef]
13. Enciso, J.; Avila, C.A.; Jung, J.; Elsayed-Farag, S.; Chang, A.; Yeom, J.; Landivar, J.; Maeda, M.; Chavez, J.C. Validation of agronomic UAV and field measurements for tomato varieties. *Comput. Electron. Agric.* **2019**, *158*, 278–283. [CrossRef]
14. Kwak, G.H.; Park, N.W. Impact of Texture Information on Crop Classification with Machine Learning and UAV Images. *Appl. Sci.* **2019**, *9*, 643. [CrossRef]
15. Singhal, G.; Bansod, B.; Mathew, L.; Goswami, J.; Choudhury, B.; Raju, P. Chlorophyll estimation using multi-spectral unmanned aerial system based on machine learning techniques. *Remote. Sens. Appl. Soc. Environ.* **2019**, *15*, 100235. [CrossRef]
16. Rakesh, D.; Kumar, N.A.; Sivaguru, M.; Keerthivasan, K.; Janaki, B.R.; Raffik, R. Role of UAVs in Innovating Agriculture with Future Applications: A Review. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 8–9 October 2021; pp. 1–6.



17. Hassan, M.A.; Yang, M.; Rasheed, A.; Yang, G.; Reynolds, M.; Xia, X.; Xiao, Y.; He, Z. A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Sci.* **2019**, *282*, 95–103. [CrossRef] [PubMed]
18. Senthilnath, J.; Dokania, A.; Kandukuri, M.; Ramesh, K.N.; Anand, G.; Omkar, S.N. Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosyst. Eng.* **2016**, *146*, 16–32. [CrossRef]
19. Johansen, K.; Morton, M.J.; Malbeteau, Y.; Aragon, B.; Al-Mashharawi, S.; Ziliani, M.G.; Angel, Y.; Fiene, G.; Negrao, S.; Mousa, M.A.; et al. Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest. *Front. Artif. Intell.* **2020**, *3*, 28. [CrossRef]
20. Johansen, K.; Morton, M.J.L.; Malbeteau, Y.; Aragon, B.; Al-Mashharawi, S.; Ziliani, M.; Angel, Y.; Fiene, G.; Negrao, S.; Mousa, M.A.A.; et al. Predicting biomass and yield at harvest of salt-stressed tomato plants using UAV imagery. *ISPRS—Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 407–411. [CrossRef]
21. Tatsumi, K.; Igarashi, N.; Mengxue, X. Prediction of plant-level tomato biomass and yield using machine learning with unmanned aerial vehicle imagery. *Plant Methods* **2021**, *17*, 1–17. [CrossRef]
22. Gil-Docampo, M.L.; Arza-García, M.; Ortiz-Sanz, J.; Martínez-Rodríguez, S.; Marcos-Robles, J.L.; Sánchez-Sastre, L.F. Above-ground biomass estimation of arable crops using UAV-based SfM photogrammetry. *Geocarto Int.* **2020**, *35*, 687–699. [CrossRef]
23. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
24. Rodríguez, R.; Sossa, J.H. *Procesamiento y Análisis Digital de Imágenes*; Alfaomega Grupo Editor: Ciudad de México, México, 2012; ISBN 10:6077072230.
25. Santibáñez, F. *Atlas agroclimático de Chile. Estado Actual y Tendencias del Clima. Tomo III Regiones de Valparaíso, Metropolitana, del Libertador Bernardo O'Higgins y del Maule*; Universidad de Chile, Facultad de Ciencias Agronómicas: Santiago, Chile, 2017.
26. SensFly. Parrot Sequoia+ Cámara Multiespectral. 2020. Available online: <https://www.sensefly.com/es/camera/parrot-sequoia> (accessed on 21March 2021).
27. Pix4D. Pix4D Radiometric Calibration Target. 2020. Available online: <https://support.pix4d.com/hc/en-us/articles/206494883-Radiometric-calibration-target> (accessed on 21March 2021).
28. Grados, D.; Schrevers, E. Cassava NDVI Analysis: A Nonlinear Mixed Model Approach Based on UAV-Imagery. *PFG—J. Photogramm. Remote. Sens. Geoinf. Sci.* **2020**, *88*, 337–347. [CrossRef]
29. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef]
30. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **1979**, *9*, 62–66. [CrossRef]
31. Chen, Y.; Chen, D.; Yang, L.; Chen, L. Otsu's thresholding method based on gray level-gradient two-dimensional histogram. In Proceedings of the 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010), Wuhan, China, 6–7 March 2010; Volume 3, pp. 282–285. [CrossRef]
32. Innani, S.; Dutande, P.; Baheti, B.; Talbar, S.; Baid, U. Fuse-PN: A Novel Architecture for Anomaly Pattern Segmentation in Aerial Agricultural Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2960–2968.
33. Ramos, A.P.M.; Osco, L.P.; Furuya, D.E.G.; Gonçalves, W.N.; Santana, D.C.; Teodoro, L.P.R.; da Silva Junior, C.A.; Capristo-Silva, G.F.; Li, J.; Baio, F.H.R.; et al. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Comput. Electron. Agric.* **2020**, *178*, 105791. [CrossRef]
34. Barnes, E.; Clarke, T.; Richards, S.; Colaizzi, P.; Haberland, J.; Kostrzewski, M.; Waller, P.; Choi, C.; Riley, E.; Thompson, T.; et al. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA, 16–19 July 2000; Volume 1619, p. 6.
35. Jorge, J.; Vallbé, M.; Soler, J.A. Detection of irrigation inhomogeneities in an olive grove using the NDRE vegetation index obtained from UAV images. *Eur. J. Remote. Sens.* **2019**, *52*, 169–177. [CrossRef]
36. Ohashi, Y.; Murai, M.; Ishigami, Y.; Goto, E. Light-Intercepting Characteristics and Growth of Tomatoes Cultivated in a Greenhouse Using a Movable Bench System. *Horticulturae* **2022**, *8*, 60. [CrossRef]
37. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
38. Everingham, Y.; Sexton, J.; Skocaj, D.; Inman-Bamber, G. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* **2016**, *36*, 27. [CrossRef]
39. Fukuda, S.; Spreer, W.; Yasunaga, E.; Yuge, K.; Sardesud, V.; Müller, J. Random Forests modelling for the estimation of mango (*Mangifera indica* L. cv. Chok Anan) fruit yields under different irrigation regimes. *Agric. Water Manag.* **2013**, *116*, 142–150. [CrossRef]
40. Patrignani, A.; Ochsner, T.E. Canopeo: A Powerful New Tool for Measuring Fractional Green Canopy Cover. *Agron. J.* **2015**, *107*, 2312–2320. [CrossRef]