

Article

Estimating Sugarcane Maturity Using High Spatial Resolution Remote Sensing Images

Esteban Rodriguez Leandro ¹, Muditha K. Heenkenda ^{2,*}  and Kerin F. Romero ¹ 

¹ School of Agricultural Engineering, Institute of Technology Costa Rica, Cartago 30101, Costa Rica; ejleandr@lakeheadu.ca (E.R.L.); kromero@itcr.ac.cr (K.F.R.)

² Department of Geography and the Environment, Lakehead University, 955 Oliver Rd., Thunder Bay, ON P7B 5E1, Canada

* Correspondence: muditha.heenkenda@lakeheadu.ca; Tel.: +1-807-343-8010 (ext. 8746)

Abstract: Sugarcane suffers from the increased frequency and severity of droughts and floods, negatively affecting growing conditions. Climate change has affected cultivation, and the growth dynamics have changed over the years. The identification of the development stages of sugarcane is necessary to reduce its vulnerability. Traditional methods are inefficient when detecting those changes, especially when estimating sugarcane maturity—a critical step in sugarcane production. Hence, the study aimed to develop a cost- and time-effective method to estimate sugarcane maturity using high spatial-resolution remote sensing data. Images were acquired using a drone. Field samples were collected and measured in the laboratory for brix and pol values. Normalized Difference Water Index, Green Normalized Difference Vegetation Index and green band were chosen (highest correlation with field samples) for further analysis. Random forest (RF), Support Vector Machine (SVM), and multi-linear regression models were used to predict sugarcane maturity using the brix and pol variables. The best performance was obtained from the RF model. Hence, the maturity index of the study area was calculated based on the RF model results. It was found that the field plot has not yet reached maturity for harvesting. The developed cost- and time-effective method allows temporal crop monitoring and optimizes the harvest time.

Keywords: sugarcane; maturity index; pol and brix; random forest; support vector machine



Citation: Leandro, E.R.; Heenkenda, M.K.; Romero, K.F. Estimating Sugarcane Maturity Using High Spatial Resolution Remote Sensing Images. *Crops* **2024**, *4*, 333–347. <https://doi.org/10.3390/crops4030024>

Received: 29 May 2024
Revised: 6 July 2024
Accepted: 10 July 2024
Published: 11 July 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/>).

1. Introduction

Sugarcane is a tall perennial grass in the genus *Saccharum*, used mainly but not solely for sugar production [1]. The plants are usually 2–6 m tall with stout, jointed, fibrous stalks rich in sucrose, accumulating in the stalk internodes (the space between the joints of the stalk) [2]. Costa Rica is one of the countries that cultivate sugarcane. Costa Rica's sugar agro-industry contributes 0.33% to the national Gross Domestic Product (GDP) and 3.83% to the agricultural GDP, generating more than USD 200 million in revenue annually [3]. The industry is estimated to create 25,000 direct jobs and more than 100,000 indirect jobs, representing 9.3% of agricultural and 1.3% of national employment [3].

Sugarcane crops have a long growth cycle. The temperature and moisture levels should be sufficient for optimal production as the crop is highly weather-dependent. However, over the years, sugarcane has suffered from the increased frequency and severity of droughts and floods, negatively affecting growing conditions [4]. The province of Guanacaste, the primary sugarcane production area in Costa Rica, is one of the most vulnerable provinces in the country in terms of drought and adverse effects of climate change related to its location on the central-American dry corridor, a route associated with a rainfall distribution problem [5]. Climate change has affected the Guanacaste province's sugarcane cultivation and industrial production. Hence, considering the adverse effects of climate change on sugarcane cultivation and, thus, the socioeconomic status of Costa

Rica, it is necessary to identify the most vulnerable points in the production cycle and develop strategies to reduce the vulnerability. The identification of vulnerable points in the production cycle will promote its competitiveness and facilitate its adaptation to minimize the negative impacts [6]. However, the development stages of sugarcane must be identified to understand the most vulnerable points in sugarcane production.

There are four main phenological development stages of sugarcane: (i) Germination and establishment—this initial stage is where sugarcane seeds or cuttings start to germinate, and young shoots begin to establish. Roots begin to develop and anchor the plant, and the first shoots break through the soil [7]. (ii) Tillering and canopy development—during this phase, the plant produces additional shoots (known as tillers). These tillers expand to form a canopy, which helps in photosynthesis and plant development [8]. (iii) Grand growth and elongation phase—the phase of rapid growth where the sugarcane stalks grow tall and the internodes elongate. The plant accumulates biomass and stores energy in the form of sucrose in the stalks [7]. (iv) Maturity and ripening—in the final stage, the sugarcane stalks mature, and the sugar content increases until the cane is ready for harvest. The leaves start to dry, and the stalks change color as they ripen [8]. Typically, each stage takes up to three months, completing a cycle of one year. Usually, in Costa Rica, there are four to eight cycles per plot before planting again, when yield and profits diminish [9].

One of the most critical points in sugarcane production is the crop's maturation process or the final phase (fourth phenological development phase) [4]. During the final phase, climate variability strongly affects sugarcane growth, causing damage such as reduced sucrose accumulation in the stalks and lower juice quality [4]. With the recent climate variations and environmental changes, agricultural practices have already been changed to overcome this challenge, and sugarcane cultivators have started to use ripening agents. Ripening agents ensure the plant's correct sucrose accumulation and juice quality at the time of harvesting [10]. However, these agents should be used in the right phenological stage of the plant. As previously stated, the crop's maturation process is one of the most critical points in sugarcane production [4] and thus is the benchmark for applying ripening agents. Therefore, estimating maturity is crucial for optimizing harvest timing to maximize sucrose content and minimize yield losses in the sugar industry [11]. According to Singh et al. [12], proper maturity assessment guides farmers in determining the ideal harvesting time, impacting sugar quality, yield, and profitability. The traditional method to identify the maturity stage is to take and analyze the sucrose concentration of its juices in the laboratory and determine its degree of maturation [6]. This is a time-consuming and costly method due to the density of the plants and extensions of the plantations. Hence, the reduced sampling strategies are implemented to overcome financial burdens. Still, those reduced samples do not represent the field variations and thus create places where the maturity is not accounted for [13]. The complexity and variation of sugarcane and its vital place in the economy drive the development of a time- and cost-effective method for measuring maturity while growing [4].

Simões et al. [13] found that spectral indices such as Simple Ratio (SR), Normalized Difference Vegetation Index (NDVI), and Soil-Adjusted Vegetation Index (SAVI) were highly correlated with agronomic parameters (Leaf Area Index (LAI), Number of Stalks per Meter, Yield, and Total Biomass). The study confirmed those relations are effective in crop monitoring and yield estimation. Similar results were also found by Lofton et al. [14]. The method specified in Zhao et al. [8] could predict sucrose content in sugarcane with 90% accuracy. These statistics were significantly higher than the accuracy of the conventional method (only accurate around 70%). Cruz-Sanabria et al. [15] showed that their method could identify phenological stages with an accuracy of 92.45% using the indices—NDVI, GNDVI, SAVI, Perpendicular Vegetation Index (PVI) and LAI. The study concluded that the proposed method accurately and effectively identifies sugarcane's phenological stages. Yeasin et al. [16] conducted a similar study using machine learning models in combination with multispectral images and concluded 88% accuracy in predicting the phenological stage of sugarcane. Another study conducted by Gunnula et al. [17] showed that ground mea-

surements and remote sensing data can effectively assess sugarcane's growth and maturity. NDVI and Enhanced Vegetation Index (EVI) were good indicators of sugarcane growth, while LAI was a good indicator of sugarcane maturity [18]. However, all these studies used space-borne multispectral images and their spatial resolutions were limited to at least 10 m to 30 m. Hence, there is a demand for efficient methods that use multispectral images of very high spatial resolution due to the significant findings of multiple studies [4,18–20] and the spatial variation of sugarcane fields.

This project aimed to develop a method to estimate sugarcane maturity using high spatial resolution remote sensing data as an alternative to traditional methods. The specific objectives were to (a) identify the optimal remote sensing band combination or indices using the correlation between field samples and remote sensing data; (b) utilize three regression models—two machine learning algorithms and a multiple linear model to estimate sugarcane maturity; and (c) compare these regression models, using performance metrics, for selecting the most suitable model and map the spatial distribution of sugarcane maturity over the field.

2. Materials and Methods

2.1. Study Area and Data

The research was conducted at “Agrícola El Cántaro” farm, located at Bebedero, Guanacaste province, Costa Rica (Figure 1). The extent of the sample field is about 14 ha. Its central coordinates are $10^{\circ}21'11.87''$ N, $85^{\circ}8'16.56''$ W. This field is characterized by loam soil composition with particle distribution of 43% sand, 23% clay and 34% silt, obtained from data captured on the field. A surface irrigation method with a surface slope of 1‰ is employed there. From hourly data captured on the field between the last harvest and the remote sensing data capturing date, the cumulative rainfall was 553 mm, and the average temperature was 28.31°C with a maximum of 40.10°C and a minimum of 15.90°C .

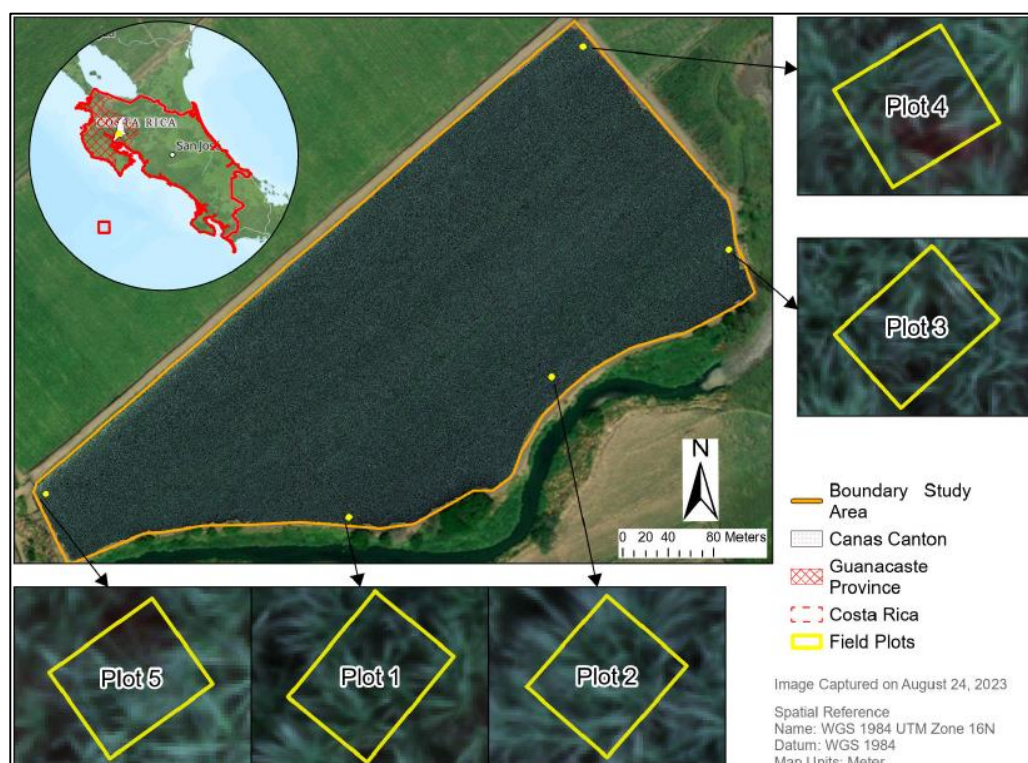


Figure 1. The study area map (upper left with the inset map of Costa Rica) and field plots 1–5 (right and bottom).

The primary purpose of this field is to produce sugar, and the variety cultivated is CP 72-2086. This variety was developed by the United States Sugar Corporation Research Farm in Canal Point (CP), Florida [21]. The first two numbers after CP represent the year the first clonal crop resulting from a cross-pollination was planted; the numbers after the hyphen represent the accession number of that cultivar in the year it was named [21]. CP 70-2086 is adapted to South Florida weather and its tropical climate; it has a preference for sandy organic soils, it yields a sugar content and tonnage from medium to high, and its canopy characteristics are high arch and slow canopy closure with medium tillering and stalk size. Usually, the harvest season is from January to mid-March with it being prone to mechanical harvest, with a fair to good stubbing ability, and this variety has the advantages of late flowering and that the sugar content increases through the harvest season [21].

This field was in its fourth-year production cycle, with a mature age of 255 days at the moment of the field sampling. The planting pattern is single-spaced with 1.7 m spacing between furrows. Consistent cultural practices, irrigation, drainage, and mechanization were applied uniformly across the plot.

Five representative plots with the extent of 1 m × 1.7 m were selected within the study area. Sugarcane plants inside each plot (Figure 1) were cut and weighed to extract fiber, Brix, pol, moisture, Purity, and sugar content measurements. These data were extracted by the Quality Assurance Laboratory at Taboga Sugar Mill using the methodology proposed by the Council of Sugarcane, Sugar and Ethanol Producers of the State of Sao Paulo [22]. Each sampling location was surveyed with a Real Time Kinematics (RTK) global navigation satellite system [23] for higher positional accuracy. Accessibility is one of the criteria when selecting sample plots. As sugarcane matures, the plot becomes denser and denser, making it almost impossible to reach some areas. The second criterion was the spatial distribution of plots within the study area. Although we considered a random distribution of plots, the appearance was checked. Young trees were not ready to harvest in some areas; hence, attention was given to areas with mature trees.

Multispectral images were captured from a Micasense RedEdge-P camera (AgEagle Aerial Systems Inc., Wichita, KS, USA) [24], mounted on a DJI enterprise Matrice-300 drone (DJI, Shenzhen, China) [25]. These images have five multispectral bands and one panchromatic band. The spectral resolution of each multispectral band is available in Table 1. The panchromatic sensor available in this camera was not used for the analysis. The spatial resolution of 5 cm was achieved by flying the drone 80 m above the ground level.

Table 1. Spectral sensitivity of Micasense Red-Edge P camera [24].

Band	Center (nm)	Bandwidth (nm)
Blue	475	32
Green (G)	560	27
Red (R)	668	14
Red-Edge	717	12
Near Infrared (NIR)	842	57

2.2. Method

Figure 2 shows the overall method. The drone images were processed using Pix4Dmapper software (Version 4.9, Pix4D S.A., Prilly, Switzerland) to obtain the radiometrically corrected orthomosaic using the calibration panel and the sensor-specific parameters provided by the manufacturer [26,27]. Then, as shown in Table 2, different vegetation indices were calculated. Different indices were selected based on their spectral response to sugarcane at various phenological stages. For example, when plants mature, they appear greener; thus, indices derived from green and near-infrared bands provide distinct results from those of other spectral bands. Also, when sugarcane ripens, it becomes brownish, and indices derived from the red band are more suitable. After investigating successful studies that used different vegetation indices, a potential, best-fitted list was compiled (Table 2). NDVI indicates vegetation health, more specifically, the level of leaf chlorophyll, and thus

is valuable for vegetation health and yield analysis [8]. Green Normalized Difference Vegetation Index (GNDVI) is an index similar to NDVI but uses the green band instead of the red band [13]; it helps to estimate a plant's chlorophyll content. The Normalized Difference Water Index (NDWI) illustrates moisture and helps to determine the water stress in plants [8]. This is an important aspect to consider regarding sugarcane—when there is a lack of water or water stress, the crop starts accumulating sucrose [28]. The Simple Ratio (SR) and Ratio Vegetation Index (RVI) were used by Simões et al. [13] in their research to estimate sugarcane yield and biomass successfully. INSEY1 and 2 are based on NDVI but adjusted with the Cumulative Growing Degree Days (CGDD) and by Days from Harvest (DFH), respectively (Table 2). This was successfully used by Lofton et al. [14] to estimate sugarcane yield potential.

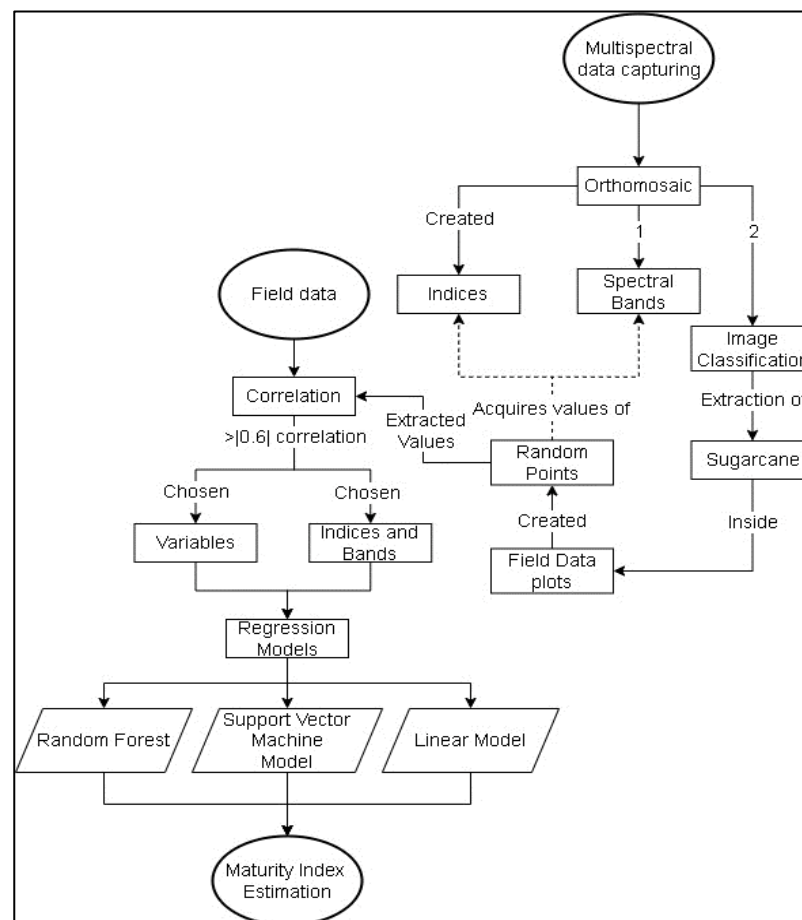


Figure 2. The flow chart of the complete workflow.

Then, the sugarcane coverage was isolated in the study area by classifying the orthomosaic. Four general classes were identified within the study area: sugarcane, shadows, soil, and other vegetation. For each class, over two hundred training polygons were created to ensure higher accuracy in the classification results. After that, the Maximum Likelihood classification algorithm was applied. The classification algorithm was selected based on the characteristics of the orthomosaic and the variety planted. Once the classification was completed, the sugarcane coverage within each field sample was extracted.

Table 2. Vegetation indices used to estimate sugarcane maturity.

Name	Equation	Reference
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR-R}{NIR+R}$	[29]
Green Normalized Difference Vegetation Index (GNDVI)	$\frac{NIR-G}{NIR+G}$	[30]
Normalized Difference Water Index (NDWI)	$\frac{G-NIR}{G+NIR}$	[8]
Simple Ratio (SR)	$\frac{NIR}{R}$	[31]
Ratio Vegetation Index (RVI)	$\frac{R}{NIR}$	[14]
INSEY 1	$\frac{NDVI}{CGDD}$	[15]
INSEY 2	$\frac{NDVI}{DFP}$	[15]
Cumulative Growing Degree Days (CGDD)	$\sum \frac{T_{max}^* - T_{min}^{**}}{2} - T_{base}^{***}$	[32]

* T_{max} : Maximum temperature (°C) ** T_{min} : Minimum temperature (°C) *** T_{base} : 18 °C for sugarcane.

Random points were generated within each sample plot, and spectral values for each band and vegetation indices (VIs) were extracted. The correlations between spectral band values, vegetation indices, and laboratory-analyzed data were calculated using the Pearson correlation coefficient, which provides a statistical measure of the strengths of linear association between these variables. A correlation coefficient greater than ± 0.6 is generally considered a variable with strong relationships [33]. Hence, all combinations greater than ± 0.6 were selected for further analysis. After that, regression models were created to predict brix and pol using two machine learning algorithms: Random Forest [34] and Support Vector Machine [35] and Mixed Linear Model (MLM).

RF algorithm is an ensemble learning method that uses multiple decision trees to classify data, employing a voting scenario to assign labels to unlabeled samples. Each tree depends on the values of a random vector sampled independently and with the same distribution for all the trees in the forest, thus reducing variance without increasing bias [34,36]. The SVM is a binary linear classifier that can be extended to manage non-linearly separable data using the kernel trick and soft margin methods. This algorithm implements the idea that input vectors are non-linearly mapped to a very high-dimension feature space, and a linear decision surface is constructed [35,36]. The algorithm demonstrates its ability to handle large datasets with many features and spectral bands with higher classification accuracy [35,36]. The MLM, as explained by McLean et al. [37], is an equation that offers the base for a methodology that provides flexibility in fitting models with various fixed and random elements with the possible assumption of correlation among random effects; this means that two or more explanatory variables could be used to explain the outcome of a desired variable.

Another variable, purity, was used to predict the maturity of sugarcane. In this study, purity will be called the Maturity Index (MI) and explained in Equation (1), as Barbosa Júnior et al. [38] proposed.

$$MI = \left(\frac{\text{pol}}{\text{brix}} \right) \times 100 \quad (1)$$

According to Araya et al. [39], brix is the concentration of solutes in grams per 100 g of solution at a specific temperature; this is the sucrose on the sugarcane juice with impurities, measured in brix degrees (°Brix). On the other hand, pol is defined as the concentration of sucrose in grams per 100 g of solution at a specific temperature; this is the sucrose on the sugarcane juice without impurities, measured in pol degrees (°Pol) [39]. Therefore, with Equation (1), taking the sucrose as a solution without impurities (pol) divided by the sucrose as a solution with impurities (brix), the purity of the sugarcane juice (maturity) as a percentage can be calculated. The industry-accepted standards for harvesting sugarcane are when $MI > 85\%$ and sucrose content $\text{pol} > 16\%$ [40]. The model performances were evaluated using the coefficient of determination (R^2), Root Mean Squared Error (RMSE), the range of the data and the model deviation.

The overall accuracy was calculated using Equation (2). The accuracy expresses the error obtained relative to the data range; thus, the error will be normalized. This helps to interpret results compared to previous studies and maintain the consistency between existing studies and their evaluation methods.

$$\text{Accuracy} = \frac{(\text{RMSE})}{(\text{Data Range})} \quad (2)$$

The level of maturity at the time of photography can also be checked with the spectral reflectance of sugarcane. It may represent the canopy chlorophyll and moisture levels. Hence, five distinct zones (in terms of the appearance from greener (healthy) to brownish (dry)) within the study area were selected for an area of 17 m × 17 m (Figure 3d). Within each area, random points were created using sugarcane pixels only. Then, the spectral values for each band were extracted, and spectral profiles were created.

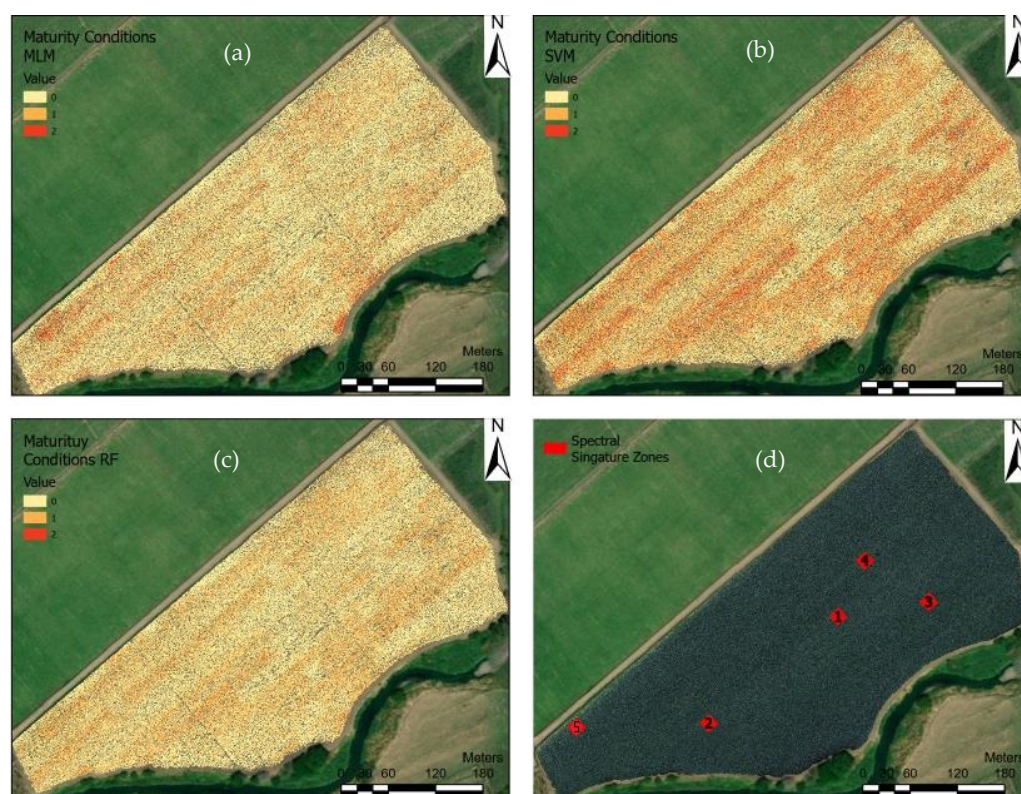


Figure 3. Maturity index maps obtained from: (a) Multiple Linear Regression Model; (b) Support Vector Machine algorithm; (c) Random Forest model; and (d) Locations of zones 1–5 for spectral analysis (red squares). The values 0, 1 and 2 represent maturity conditions D, C and B respectively.

3. Results

The sugarcane area was isolated with a higher degree of accuracy, and the total extent of the sugarcane was 10.23 ha. Figure 4 shows the correlation analysis results. The green band, NDWI and GDNVI were correlated the best, with brix values, providing R^2 of -0.69 , -0.75 and 0.68 , respectively. For pol, the R^2 values were -0.66 , -0.73 and 0.65 , respectively. Hence, green band, NDWI and GDNVI were used to predict brix and pol and, thus, the maturity index. The rest of the indices were discarded since they did not meet the threshold (± 0.6) established in the method.

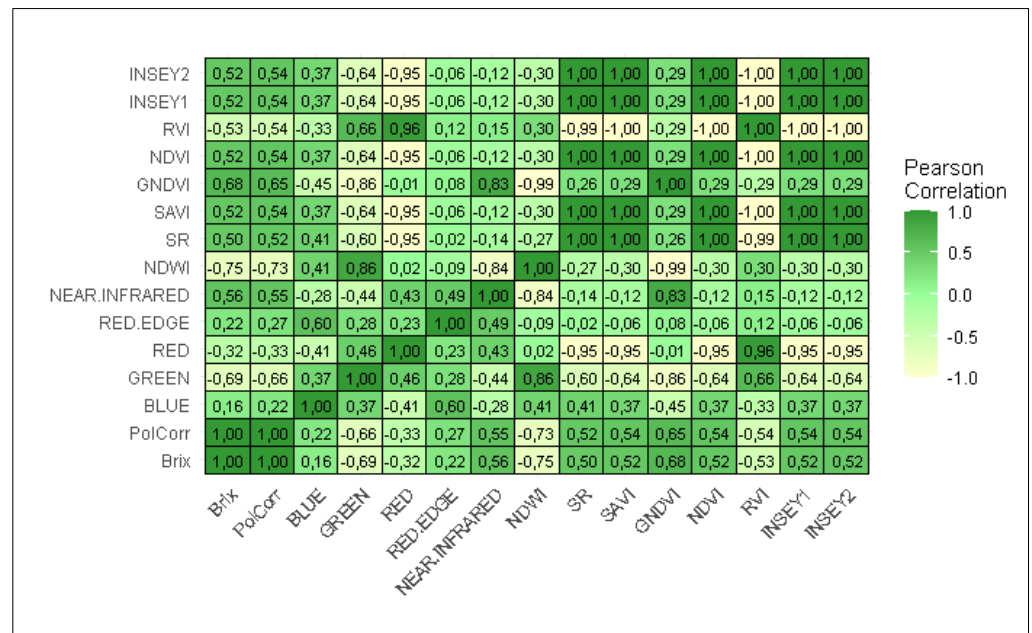


Figure 4. Pearson correlation matrix for each variable, spectral band, and indices.

The MLR models for brix and pol are shown in Equations (3) and (4). They were used to predict the spatial distribution of Pol and Brix maps.

$$\text{Brix} = 12.80412 - 5.27616 \times (\text{NDWI}) + 2.1931 \times (\text{GNDVI}) + 0.00002903 \times (\text{GREEN}) \quad (3)$$

$$\text{Pol} = 8.93896 - 6.75543 \times (\text{NDWI}) + 2.69357 \times (\text{GNDVI}) + 0.00003741 \times (\text{GREEN}) \quad (4)$$

The extent of the percentage of the maturity index (Figure 5a) was grouped into four intervals: 65–70% (denoted as D_{MI}), 70–75% (C_{MI}), 75–80% (B_{MI}) and >80% (A_{MI}). For the A_{MI} category, SVM predicted an area of 6.12 ha, RF = 3.21 ha and MLM = 2.28 ha. B_{MI} was 3.95 ha, 6.92 ha and 7.91 ha for SVM, RF and MLM, respectively. The RF algorithm did not detect the C_{MI} category; the rest were minimal. D_{MI} values were zero or very small for these algorithms (Figure 5a).

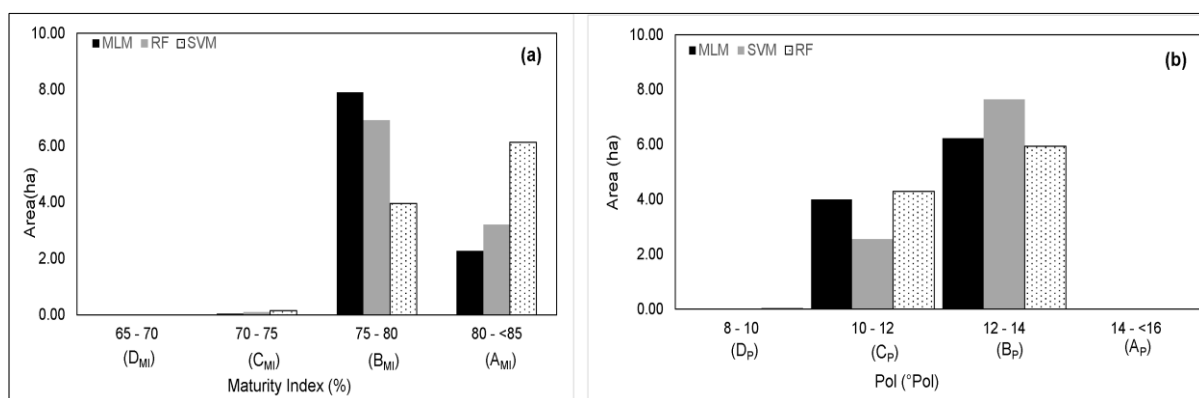


Figure 5. The sugarcane extent related to pol and MI: (a) areas predicted for pol using different prediction models; (b) the percentage of the area of the maturity index from different prediction models.

For pol prediction, the results were categorized into four intervals: 8–10 Pol° (denoted as D_P), 10–12 Pol° (denoted as C_P), 12–14 Pol° (denoted as B_P) and all values > 14 Pol° (denoted as A_P) (Figure 5b). SVM predicted 0.63 m² for A_P , no area was detected from RF, and MLM resulted in 3.73 m². B_P obtained the highest values from SVM (7.65 ha) and the lowest from RF (5.94 ha) (Figure 5b). For the C_P category, the SVM prediction was 2.56 ha,

RF was 4.30 ha, and MLM predicted 4.00 ha. The predicted values for D_p were 0.0160 ha for SVM, 0.0013 ha for RF and 0.0025 ha for MLM.

A classification system was established to assess the sugarcane maturity (Figure 6), where “A” represents $MI > 85\%$ AND $Pol^\circ > 16$, “B” represents $MI > 83\%$ AND $Pol^\circ > 13$, “C” represents $MI > 83\%$ OR $Pol^\circ > 13$, and “D” represents neither condition. This is the standard practice of any quality assurance laboratory in Costa Rica and the general practice among farmers [35]. No area was predicted for “A” from any of the methods. Regarding category B, the MLM estimated the largest area (7.3 ha), followed by the RF algorithm (6.7 ha), for C, SVM, RF, and MLM predicted areas 4.2 ha, 3.4 ha and 4.2 ha, respectively. The lowest values were predicted for category D from all three models.

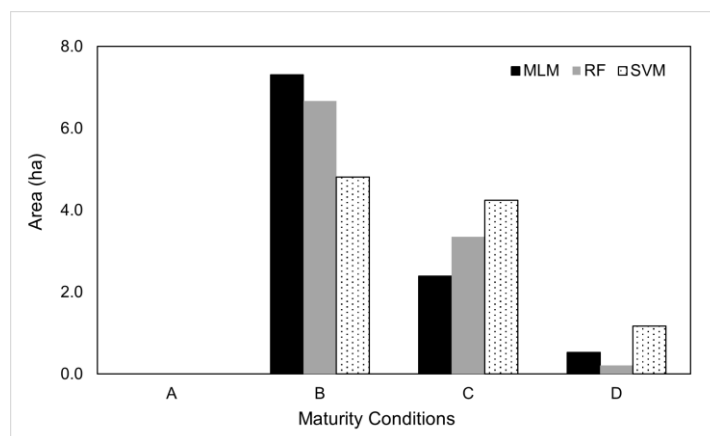


Figure 6. Areas for different maturity conditions (A–D) from different prediction models (mixed linear model (black), random forest (gray), support vector machine (white with dot pattern)).

When predicting the brix values, the highest accuracy was obtained from the RF model (82.68%) (Table 3). The SVM model demonstrated a precision of 21.89% and an accuracy of 66.64% with a deviation of 33.36%. The MLM results showed the lowest accuracy and precision.

Table 3. Performance metrics for the variable brix ($^\circ$ Brix), pol ($^\circ$ Pol) and maturity index.

	Method	MLM	SVM	RF
Brix ($^\circ$ Brix)	Precision	12.61%	21.89%	77.57%
	Accuracy	65.74%	66.64%	82.68%
	Deviation	34.26%	33.36%	17.32%
Pol ($^\circ$ Pol)	Precision	11.81%	21.02%	77.31%
	Accuracy	66.25%	67.46%	82.91%
	Deviation	33.75%	32.54%	17.09%
Maturity Index	Precision	12.21%	21.45%	77.44%
	Accuracy	65.99%	67.05%	82.80%
	Deviation	34.01%	32.95%	17.20%

For the pol prediction results (Table 3), the highest precision and accuracy were obtained from the RF. Both SVM and MLM models showed lower accuracies. Their deviations were approximately similar. The RF model also received the highest accuracy for the maturity index (Table 3). The SVM and MLM showed identical results (approximately 65% accuracy). Hence, the best performance was achieved using the RF algorithm.

As shown in Figure 7, the spectral signatures were distinct from each other and thus their sugarcane parameters. This is explained by the different MI and Pol values of zones (Table 4). For instance, Zone 5 exhibited greater values for all bands than other zones, and

its MI is the highest among them. Approximately similar values were observed for Zones 1 to 4, where an increase in the MI and °Pol resulted in decreased spectral values.

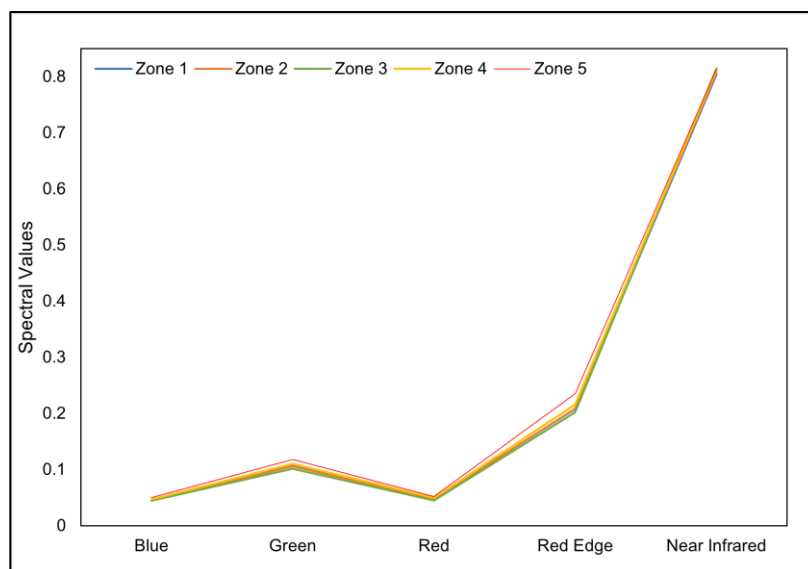


Figure 7. Spectral signatures for Zones 1 to 5.

Table 4. Maturity Index (MI) and pol results for Zones 1 to 5.

Zone	MI	Pol
Zone 1	78.84%	12.00
Zone 2	79.53%	12.29
Zone 3	79.91%	12.44
Zone 4	78.03%	11.68
Zone 5	80.15%	12.55

Figure 3a–c shows the spatial distribution of predicted maturity conditions over the study area using RF, MLM and SVM. SVM predictions are visually different from the other two methods; the map shows more category B areas than others. However, the prediction from the RF model is similar to the field photos or reality; for instance, closer to the edges of the study area, there were lots of category B sugarcane plants.

4. Discussion

This study developed a novel method to determine the sugarcane brix and pol values and, thus, evaluate the maturity of the sugarcane. This can be applied at the operational level to estimate the sugarcane maturity and may reduce the time and cost associated with existing ground sampling-based methods. The robust results generated from this study also confirmed the data acquisition plan, including sensors for remotely sensed data acquisition, data acquisition parameters (shutter speed, flying height and overlaps, etc.) and data processing workflow for higher accuracy.

Green band, NDWI and GNDVI exhibited the strongest correlations with brix and pol values (Figure 4). The green band and NDWI showed a high negative correlation due to the lower canopy chlorophyll in sugarcane when maturing and becoming brownish. Typically, the visible green band maximizes the reflectance of water surfaces. Hence, NDWI and GNDVI are sensitive to the moisture level, and when crops reach maturity, the moisture level will change [8]. Chea et al. [41] demonstrated that combining VIs and visible spectral bands related to the chemical and physical changes inside the plant resulted in greater accuracy in predicting the brix, pol, and subsequently, the maturity index. They suggested the same spectral band and indices as in this study (green band, NDWI and GNDVI).

Additionally, Lebourgeois et al. [42] used similar spectral combinations and proved that they allowed for a significant quantitative characterization of nitrogen status on sugarcane, impacting the physicochemical status of the plant. Moreover, considering the relationship between canopy nitrogen, greenness and biomass with maturity index and pol, Chaves, Lofton et al. Villegas and Larrahondo [15,28,43] and Erdle et al. [44] decided to use spectral indices that incorporate NIR and green band. Further, de Oliveira et al. [38] also supported the efficacy of the combination of these bands and indices.

When considering the overall prediction results, most areas were within the 12–14 pol interval and 75–80% MI range. They can be categorized as either “B” or “C”, and each is about 4–6 ha. That means the study area has not yet reached the harvesting level (“A”). However, the SVM model predicted larger areas for MI > 80%. However, there are no or very small areas for pol > 16 to be classified as “A” according to the standards given by the quality assurance laboratory and to Barbosa Júnior et al. [19]. However, studies from [43], Chaves [28] and Araya et al. [39] used different criteria for maturity—they suggest at least pol = 13 and MI = 80% for category “B”.

The best prediction model performances were obtained from the RF model. It was followed by SVM and the MLM for brix, pol, and MI, respectively. Notably, RF precision ranged between 77.31% and 77.57%, while SVM ranged from 21.02% to 21.89%, and MLM was even lower. This study first assessed the correlation between VIs and pol and brix values. Then, VIs and spectral bands with the highest correlation coefficients were selected as predictor variables for regression models. Hence, the variable importance was not considered within the RF model, but the parameter optimization (and feature importance) was performed. This might allow the RF model to perform well while increasing the predictive power and computing speed. When predicting brix values, Chea et al. [36] reported a precision of 45% for SVM and RF; however, they employed different sets of VI's such as $Cl_{rededge}$, inverted reflectance green and Plant Pigment Ratio (PPR). This reflects that this study obtained better results for brix prediction when combining the selected VIs and the green band.

In comparison with Zhao et al. [8] and Cruz-Sanabria et al. [16], the results obtained for the pol (Figure 5b) were lower since their study obtained a precision of 90.4% and 92.45%, respectively. They used a more extensive data set to test the proposed methodology. As a general rule of thumb, the precision that can be obtained from a traditional method is about 70% [8,16]. Hence, even with a small number of samples, the RF model of this study showed 7.31% higher precision than the traditional method for Pol estimation. Therefore, it is evident that the developed method can be used in the field (or production) as it provides relatively accurate and efficient results (time and cost) compared to conventional methods. A similar result was obtained by Barbosa Júnior et al. [19] in their study, where they obtained a precision of 77% for brix. However, the MI results were much higher, about 90%, compared to the 77.44% of our study. When comparing the results of the MLM, Gunnula et al. [18] obtained about 60% higher accuracy than this study. This is mainly due to the limited number of samples this study used. The considerable outperforming results of RF compared with SVM were proven by Sheykhmousa et al. [36]. They concluded that the MLM suffers from a small number of samples, causing underfitting and adding to the problem that these models struggle with non-linear patterns. Finally, comparing the results of existing studies, although this study used a small number of samples, yielded greater precision and accuracy for predictions. When selecting samples, the study considered uniform conditions such as variety, geographic zone, soil texture, and agricultural and cultural practices, compared with the other studies that varied one or more of them.

The spectral signatures for zones 1 to 4 exhibited similar results with slight Pol and MI value changes. They are visually identical. A lower spectral value for each band indicates their preparation for maturity as this crop's natural response to chlorophyll and moisture levels when maturing. The dry matter of crops started to increase, which was related to sucrose accumulation [43]. Also, as shown in Figure 3, the lower the value of MI and pol, the farther from the center of the plot they are located, a trend found by Som-ard et al. [4], since

the center zones start to dry out sooner than the edges of the plot. Zone 5 was an outlier to both trends mentioned here, but this is due to the proximity of this zone to drainage and water availability, allowing the development of these plants and resulting in high-density, vigorous leaves. This correlates with the results of Som-ard et al. [45], who found that a high yield can be found on the edge of the field since it is more photosynthetically efficient.

One of the most prominent potential benefits of using this developed method is optimizing the harvest timing. As a control parameter within the factory, the search for the optimal maturity point is performed, thus maximizing sugar yield and, consequently, profitability for producers. The developed method significantly reduces the costs associated with traditional sampling. Conventional methods require intensive and repeated field labor, resulting in high operational costs. In contrast, this remote sensing method can quickly cover large cultivation areas, minimizing the need for field personnel and reducing logistical expenses. Data collected from drones can be consistently analyzed across different regions and scales, facilitating comprehensive production management at a macro level. Moreover, it allows for temporal monitoring, enabling the observation of trends for future predictions and strengthening the method with more data series over time. However, there are limitations to consider. The implementation of the method can be affected by adverse weather conditions, the initial cost of implementing these technologies, and the need for technical skills to interpret the data. Future research on estimating the maturity of sugarcane using remote sensing should focus on overcoming these limitations and improving the precision and accessibility of these technologies. Synergy and collaboration among researchers, technologists, and producers will be vital in developing solutions tailored to the specific needs of different regions and growing conditions.

Agriculture is moving into an area of combined operations, where an adjustment of a practice makes a difference to the yield of the crop, and for this reason, the agriculture management could influence the behavior of the crop. Normally a sugarcane field consists of a five-year cycle until it renews. The value of sugarcane reduces at each cycle, affecting the results of the method across the years. Conducting longitudinal studies to validate the method over multiple growing seasons (years) will help to refine the models and confirm their robustness. Increasing the number of samples will further enhance its applicability and reliability at the production level by improving the effectiveness of the method.

5. Conclusions

The study developed a method to estimate sugarcane maturity at the farm level using high spatial resolution remote sensing images and a machine learning algorithm. Many studies used satellite images to assess the vegetation traits of sugarcane around the world. However, the spatial resolution of the most widely available and free satellite images limits the ability to extract fine details about sugarcane fields. Hence, a reliable method that can be actionable at the farm level using high spatial resolution images is in demand. Therefore, the research aimed to develop a method to estimate sugarcane maturity using images acquired from drones as an alternative to traditional methods. The images were captured using a Micasense RedEdge-P camera, and field samples of sugarcane were collected and analyzed in the laboratory. Different vegetation indices (nine indices that can be generated from the MicaSense RedEdge-P sensor) were calculated, and their correlation was tested against laboratory-generated “pol” and “brix” values. The highest correlation was obtained for the Normalized Difference Water Index (NDWI), Green Normalized Difference Vegetation Index (GNDVI) and the green band. The green band, GNDVI and NDWI are sensitive to the canopy chlorophyll available in sugarcane and, thus, showed a high negative correlation between the variables (“pol” and “brix”) when maturing. Therefore, the green band, GNDVI and NDWI were used to predict the spatial distribution of pol and brix values over the study area using a multiple linear regression model (MLM) and two machine learning models (Random Forest (RF) and Support Vector Machine (SVM)). The performance matrices for the brix model demonstrated that the RF had greater accuracy and precision than the rest, with 77.57% and 82.68%,

respectively. The second best was SVM, followed by MLM, which had precision and accuracy of 21.89%, 66.64%, 12.61%, and 65.74%, respectively. The RF models showed the best results for the pol models and maturity index. Therefore, the RF demonstrates better accuracy than SVM in cases of increasing dimensionality and small training data samples. MLM suffers from fewer samples, causing the model to be underestimated. This indicates the potential of machine learning models, mainly RF, to handle complex agricultural datasets and provide reliable predictions. The Maturity Index (MI) or purity of sugarcane was estimated based on the percentage of pol per brix values. The spatial distribution of MI over the study area helps to determine areas where the level of maturity is reached for harvesting or not. The method offers a practical tool for farmers to optimize harvest timing, which can enhance sugar yield and profit by providing a cost-effective and scalable alternative to traditional methods. Also, this state-of-the-art technique enables farmers to make data-driven decisions, improving efficiency and reducing the time it takes to collect samples and harvest the crop at the right time. Additionally, it promotes sustainable agricultural practices by enabling more precise resource management, thus minimizing environmental impact. Therefore, based on the results of this study, we can conclude that (a) the optimal band/indices combination for estimating sugarcane maturity is NDWI, GNDVI and green; (b) according to the results of the regression models, the plot is not ready for harvesting since it did not achieve the minimum values for the pol and maturity index, and (c) according to the performance metrics and the conditions of the study, the most suitable regression model for estimating sugarcane maturity is the random forest model. Future research should explore the scalability of this method across different geographical regions and weather conditions. Small variations in agricultural practices, weather or geographic regions could highly affect the research outcome. Integrating this approach with other precision agriculture technologies, such as soil moisture sensors, irrigation systems, and different management practices, could prove the effectiveness of the method across various scenarios.

Author Contributions: Conceptualization, E.R.L. and K.F.R.; methodology, E.R.L., K.F.R. and M.K.H.; software, E.R.L., K.F.R. and M.K.H.; formal analysis, E.R.L. and M.K.H.; writing—original draft preparation, E.R.L.; writing—review and editing, M.K.H.; supervision, M.K.H.; project administration, K.F.R.; funding acquisition, K.F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Mitacs-CALAREO Globalink Research Award, grant number IT36425.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: We would like to express our sincere gratitude to José Torres and the “Agrícola El Cántaro” farm team for their invaluable collaboration in collecting data for this study. In addition, we would like to thank our colleagues who have provided support and guidance during the development of this project.

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

References

1. Singh, A.; Lal, U.; Mukhtar, H.; Singh, P.; Shah, G.; Dhawan, R. Phytochemical profile of sugarcane and its potential health aspects. *Pharmacogn. Rev.* **2015**, *9*, 45–54. [CrossRef]
2. Kltimor, K. Stages of Sugarcane Production: A Review. *Res. Gate* **2022**. [CrossRef]
3. Chaves, M.; Bermúdez, L. Agroindustria Azucarera Costarricense: Un Modelo Organizacional y Productivo Efectivo Con 75 Años de Vigencia. Cali, Colombia. 2015. Available online: <https://servicios.laica.co.cr/laica-cv-biblioteca/index.php/Library/download/DSCmSdyqoIjhQAUmeqVMAwOPjrySXJdh> (accessed on 9 December 2023).
4. Som-ard, J.; Atzberger, C.; Izquierdo-Verdiguier, E.; Vuolo, F.; Immitzer, M. Remote Sensing Applications in Sugarcane Cultivation: A Review. *Remote. Sens.* **2021**, *13*, 4040. [CrossRef]
5. Wake, B. Central America drying. *Nat. Clim. Change* **2021**, *11*, 900. [CrossRef]

6. Vignola, R.; Poveda, K.; Walter, W.; Vargas, A.; Berrocal, Á. *Cultivo de Caña de Azúcar en Costa Rica*; San José, Costa Rica, 2018; Available online: <https://www.mag.go.cr/bibliotecavirtual/F01-8327.pdf> (accessed on 27 February 2023).
7. Wiedenfeld, R.P. Water stress during different sugarcane growth periods on yield and response to N fertilization. *Agric. Water Manag.* **2000**, *43*, 173–182. [[CrossRef](#)]
8. Zhao, Y.; Della Justina, D.; Kazama, Y.; Rocha, J.V.; Graziano, P.S.; Lamparelli, R.A.C. Dynamics modeling for sugar cane sucrose estimation using time series satellite imagery. In Proceedings of the SPIE Remote Sensing, Edinburgh, UK, 25 October 2016; p. 99980J. [[CrossRef](#)]
9. Leon, R.G.; Agüero, R.; Calderón, D. Diversity and Spatial Heterogeneity of Weed Communities in a Sugarcane Cropping System in the Dry Tropics of Costa Rica. *Weed Sci.* **2017**, *65*, 128–140. [[CrossRef](#)]
10. Alfaro, R.; Ocampo, R.; Barrantes, J. Respuesta Productiva de Diferentes Madurantes en El Cultivo de La Caña de Azúcar, Zona Sur, Costa Rica. Alajuela, Costa Rica. 2019. Available online: <https://servicios.laica.co.cr/laica-cv-biblioteca/index.php/Library/download/bbaVIFwLDpaHsTAihTAIVysfTnLAZZcL> (accessed on 14 March 2023).
11. Pereira, L.F.; Ferreira, V.M.; DE Oliveira, N.G.; Sarmiento, P.L.; Endres, L.; Teodoro, I. Sugars levels of four sugarcane genotypes in different stem portions during the maturation phase. *An. Acad. Bras. Cienc.* **2017**, *89*, 1231–1242. [[CrossRef](#)]
12. Singh, P.; Singh, S.N.; Tiwari, A.K.; Pathak, S.K.; Singh, A.K.; Srivastava, S.; Mohan, N. Integration of sugarcane production technologies for enhanced cane and sugar productivity targeting to increase farmers' income: Strategies and prospects. *3 Biotech* **2019**, *9*, 48. [[CrossRef](#)]
13. Rahimi, M.; Liaghat, A.; Mirzaei, F. Optimization of Sugarcane Harvest Using Remote Sensing. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-4/W18*, 857–861. [[CrossRef](#)]
14. Simões, M.D.S.; Rocha, J.V.; Lamparelli, R.A.C. Spectral variables, growth analysis and yield of sugarcane. *Sci. Agricola* **2005**, *62*, 199–207. [[CrossRef](#)]
15. Lofton, J.; Tubana, B.S.; Kanke, Y.; Teboh, J.; Viator, H.; Dalen, M. Estimating Sugarcane Yield Potential Using an In-Season Determination of Normalized Difference Vegetative Index. *Sensors* **2012**, *12*, 7529–7547. [[CrossRef](#)]
16. Cruz-Sanabria, H.; Sanchez, M.G.; Rivera-Caicedo, J.P.; Avila-George, H. Identification of phenological stages of sugarcane cultivation using Sentinel-2 images. In Proceedings of the 2020 9th International Conference on Software Process Improvement (CIMPS), Mazatlan, Sinaloa, Mexico, 21–23 October 2020; pp. 110–116. [[CrossRef](#)]
17. Yeasin, M.; Haldar, D.; Kumar, S.; Paul, R.K.; Ghosh, S. Machine Learning Techniques for Phenology Assessment of Sugarcane Using Conjunctive SAR and Optical Data. *Remote Sens.* **2022**, *14*, 3249. [[CrossRef](#)]
18. Gunnula, W.; Kositrakun, M.; Righetti, T.; Weerathaworn, P.; Prabpan, W. Evaluating Sugarcane Growth and Maturity Using Ground-Based Measurements and Remote Sensing Data. *Thai J. Agric. Sci.* **2012**, *2015*, 17–28. Available online: https://www.researchgate.net/publication/280132988_Evaluating_Sugarcane_Growth_and_Maturity_Using_Ground-Based_Measurements_and_Remote_Sensing_Data (accessed on 5 September 2023).
19. Júnior, M.R.B.; Moreira, B.R.d.A.; de Oliveira, R.P.; Shiratsuchi, L.S.; da Silva, R.P. UAV imagery data and machine learning: A driving merger for predictive analysis of qualitative yield in sugarcane. *Front. Plant Sci.* **2023**, *14*, 1114852. [[CrossRef](#)]
20. Akbarian, S.; Jamnani, M.R.; Xu, C.; Wang, W.; Lim, S. Plot level sugarcane yield estimation by machine learning on multispectral images: A case study of Bundaberg, Australia. *Inf. Process. Agric.* **2023**, *in press*. [[CrossRef](#)]
21. Schueneman, T.J.; Miller, J.D.; Gilbert, R.A.; Harrison, N.L. *Sugarcane Cultivar CP 72-2086 Descriptive Fact Sheet*; University of Florida: Gainesville, FL, USA, 2001.
22. 2006. Available online: https://edisciplinas.usp.br/pluginfile.php/4277152/mod_resource/content/1/ManualdoConsecana.pdf (accessed on 28 May 2024).
23. CHCNAV. CHCNAV i50 GNSS User Guide. Shanghai. 2018, pp. 1–68. Available online: <https://chcnv.es/es/productos/gps-chc-i50-gnss/chc-manual-gps-centimetrico-i50-en/view> (accessed on 7 January 2024).
24. MicaSense. MicaSense Series RedEdge-P. 2021. Available online: <https://ageagle.com/drone-sensors/rededge-p-high-res-multispectral-camera/> (accessed on 7 January 2024).
25. DJI. Matrice 300 RTK. 2020. Available online: <https://enterprise.dji.com/matrice-300> (accessed on 1 April 2023).
26. PIX4D. PIX4Dmapper. 2023. Available online: <https://www.pix4d.com/product/pix4dmapper-photogrammetry-software/> (accessed on 4 January 2023).
27. Kedzierski, M.; Wierzbicki, D.; Sekrecka, A.; Fryskowska, A.; Walczykowski, P.; Siewert, J. Influence of Lower Atmosphere on the Radiometric Quality of Unmanned Aerial Vehicle Imagery. *Remote Sens.* **2019**, *11*, 1214. [[CrossRef](#)]
28. Chaves, M. La Maduración, su Control y la Cosecha de la caña de Azúcar. In *Seminario de Tecnología Moderna de la Caña de Azúcar*; San José, Costa Rica, 1982; Volume 2, pp. 28–40. Available online: <https://servicios.laica.co.cr/laica-cv-biblioteca/index.php/Library/download/FAmCPeflIHfQCIJhAhnzDdGwAStzOdssA> (accessed on 14 March 2023).
29. Rouse, J.W. Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation. Texas. 1974. Available online: <https://ntrs.nasa.gov/citations/19740022555> (accessed on 14 January 2024).
30. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
31. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [[CrossRef](#)]

32. Ferchault de Réamur, R.-A. Observations du Thermometre, Faites a Paris Pendant l'annees 1735, Comparees a Celles qui ont ete Faites sous la Ligne, a l'Isle de France, a Alger et en Quelques-Unes de nos Isles de l'Amerique. 1735. Available online: https://www.academie-sciences.fr/pdf/dossiers/Reaumur/Reaumur_pdf/p545_576_vol3532m.pdf (accessed on 14 January 2024).
33. Sedgwick, P. Pearson's correlation coefficient. *BMJ* **2012**, *345*, e4483. [[CrossRef](#)]
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
35. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
36. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [[CrossRef](#)]
37. McLean, R.A.; Sanders, W.L.; Stroup, W.W. A Unified Approach to Mixed Linear Models. *Am. Stat.* **1991**, *45*, 54–64. [[CrossRef](#)]
38. Araya, J.; Alpizar, E.; Valverde, W.; Chavarria, E. Aplicación del índice de madurez de la caña como criterio de cosecha: Un método sencillo con un fundamento fisiológico sólido y un enfoque práctico. *Entre Cañeros* **2016**, *6*, 51–59. Available online: https://www.researchgate.net/publication/335003001_Aplicacion_del_indice_de_madurez_de_la_cana_como_criterio_de_cosecha_un_metodo_sencillo_con_un_fundamento_fisiologico_solido_y_un_enfoque_practico (accessed on 15 January 2023).
39. Quality Assurance Laboratory at Taboga Mill, Cañas, Guanacaste, Cost Rica. Industry Standards for Harvesting Sugarcane. *Private Communication*, 2023.
40. Chea, C.; Saengprachatanarug, K.; Posom, J.; Saikaew, K.; Wongphati, M.; Taira, E. Optimal models under multiple resource types for Brix content prediction in sugarcane fields using machine learning. *Remote Sens. Appl. Soc. Environ.* **2022**, *26*, 100718. [[CrossRef](#)]
41. Lebourgeois, V.; Bégué, A.; Labbé, S.; Houles, M.; Martiné, J.F. A light-weight multi-spectral aerial imaging system for nitrogen crop monitoring. *Precis. Agric.* **2012**, *13*, 525–541. [[CrossRef](#)]
42. Villegas, F.; Larrahondo, J. Control y características de maduración. In *El Cultivo de la Caña en la Zona Azucarera de Colombia*; Centro de Investigación de la Caña de Azúcar de Colombia, Ed.; Centro de Investigación de la Caña de Azúcar de Colombia: Cali, Colombia, 1995; pp. 297–313.
43. Erdle, K.; Mistele, B.; Schmidhalter, U. Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars. *Field Crop. Res.* **2011**, *124*, 74–84. [[CrossRef](#)]
44. de Oliveira, R.P.; Júnior, M.R.B.; Pinto, A.A.; Oliveira, J.L.P.; Zerbato, C.; Furlani, C.E.A. Predicting Sugarcane Biometric Parameters by UAV Multispectral Images and Machine Learning. *Agronomy* **2022**, *12*, 1992. [[CrossRef](#)]
45. Som-ard, J.; Hossain, M.D.; Ninsawat, S.; Veerachitt, V. Pre-harvest Sugarcane Yield Estimation Using UAV-Based RGB Images and Ground Observation. *Sugar Tech.* **2018**, *20*, 645–657. [[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.