



## Article

# Exploring the Relationship Between Very-High-Resolution Satellite Imagery Data and Fruit Count for Predicting Mango Yield at Multiple Scales

Benjamin Adjah Torgbor<sup>1,2</sup>, Priyakant Sinha<sup>1,3,\*</sup>, Muhammad Moshir Rahman<sup>1</sup>, Andrew Robson<sup>1</sup>, James Brinkhoff<sup>1</sup> and Luz Angelica Suarez<sup>1</sup>

<sup>1</sup> Applied Agricultural Remote Sensing Centre, University of New England, Armidale, NSW 2351, Australia; btorgbor@myune.edu.au (B.A.T.); mrahma37@une.edu.au (M.M.R.); arobson7@une.edu.au (A.R.); james.brinkhoff@une.edu.au (J.B.); lsuarezc@une.edu.au (L.A.S.)

<sup>2</sup> Forestry Commission, Accra P.O. Box MB 434, Ghana

<sup>3</sup> Ecosystem Management, School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia

\* Correspondence: psinha2@une.edu.au; Tel.: +61-2-6773-3849

**Abstract:** Tree- and block-level prediction of mango yield is important for farm operations, but current manual methods are inefficient. Previous research has identified the accuracies of mango yield forecasting using very-high-resolution (VHR) satellite imagery and an '18-tree' stratified sampling method. However, this approach still requires infield sampling to calibrate canopy reflectance and the derived block-level algorithms are unable to translate to other orchards due to the influences of abiotic and biotic conditions. To better appreciate these influences, individual tree yields and corresponding canopy reflectance properties were collected from 2015 to 2021 for 1958 individual mango trees from 55 orchard blocks across 14 farms located in three mango growing regions of Australia. A linear regression analysis of the block-level data revealed the non-existence of a universal relationship between the 24 vegetation indices (VIs) derived from VHR satellite data and fruit count per tree, an outcome likely due to the influence of location, season, management and cultivar. The tree-level fruit count predicted using a random forest (RF) model trained on all calibration data produced a percentage root mean squared error (PRMSE) of 26.5% and a mean absolute error (MAE) of 48 fruits/tree. The lowest PRMSEs produced from RF-based models developed from location, season and cultivar subsets at the individual tree level ranged from 19.3% to 32.6%. At the block level, the PRMSE for the combined model was 10.1% and the lowest values for the location, seasonal and cultivar subset models varied between 7.2% and 10.0% upon validation. Generally, the block-level predictions outperformed the individual tree-level models. Maps were produced to provide mango growers with a visual representation of yield variability across orchards. This enables better identification and management of the influence of abiotic and biotic constraints on production. Future research could investigate the causes of spatial yield variability in mango orchards.

**Keywords:** mango (*Mangifera indica* L.) fruit count; vegetation indices (VIs); machine learning; random forest; satellite imagery; yield prediction



**Citation:** Torgbor, B.A.; Sinha, P.; Rahman, M.M.; Robson, A.; Brinkhoff, J.; Suarez, L.A. Exploring the Relationship Between Very-High-Resolution Satellite Imagery Data and Fruit Count for Predicting Mango Yield at Multiple Scales. *Remote Sens.* **2024**, *16*, 4170. <https://doi.org/10.3390/rs16224170>

Academic Editors: Jorge Martínez-Guanter, Salah Er-Raki and Akash Ashapure

Received: 25 August 2024

Revised: 30 October 2024

Accepted: 6 November 2024

Published: 8 November 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

Over the past few decades, the yield of mango (*Mangifera indica* L.), one of the most important fruit crops with global recognition due to its nutritional benefits and economic value, has been on the rise [1]. Mango yield is influenced by a range of factors, including soil quality, weather (i.e., rainfall and temperature) and management practices, among others [2–4]. Accurate mango yield prediction promotes the efficient use of farm inputs and resources such as irrigation, fertiliser, labour and machinery [5,6]. Additionally, with an understanding of the spatial variability in yield across a farm (trees and blocks), growers are

able to optimise production by implementing site-specific management programmes on the farms. This is the essence of the concept of precision agriculture, which aims at providing support for growers to apply the right amount of input at the right time in the right place (i.e., per tree or management zone) [7]. Thus, the ability to accurately estimate mango yield in a timely manner at varying scales is crucial for effective agricultural planning and resource allocation [5,6]. This also help growers in planning harvesting logistics (labour, transport, storage and processing requirements) and total production estimates to support forward selling [8–10].

Presently, although the industry has tested different in-season mango yield estimations at the tree and block levels using techniques such as manual counting, hyperspectral imaging and machine vision, such trials are often costly, time- and labour-intensive and lack scalability [11–14]. These trials are usually conducted on small plot(s) in the orchard and the results are often extrapolated to the entire orchard or farm [6,11,15]. This sometimes results in inaccurate estimates due to occlusion and the inability to adequately describe spatial variability. There is therefore the need for a technology that predicts annual yield using the RS-based vegetation indices (VIs) approach that has the potential of covering a larger area with relatively less labour and time investment.

Currently, the study of vegetation dynamics and their relationship with agricultural productivity has gained a significant interest globally. Remote sensing (RS) technologies using satellite imagery provide valuable tools for monitoring and analysing VIs that are closely associated with crop yield [7]. Past studies have demonstrated the potential of VIs, such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), and Transformed Chlorophyll Absorption Reflectance Index (TCARI), in estimating crop yield in various agricultural systems [7,16]. However, limited research has focused specifically on mango orchards and the potential of very-high-resolution (VHR) satellite imagery to assess mango yield at multiple scales using machine learning (ML) algorithms. By filling this research gap, we aim to enhance our understanding of the relationship between VIs and mango productivity, ultimately providing valuable information to mango growers and stakeholders regarding spatial yield variability, targeted input application and more. The availability of VHR satellite imagery, such as WorldView-2 (WV2) and WorldView-3 (WV3), has opened up immense possibilities for understanding the relationship between VIs and crop yield at various spatial scales [6,10,13]. WV imagery offers spectral bands and spatial resolutions suitable for detailed vegetation analysis at the tree, block and farm levels. By leveraging the strengths of this technology, insights into the spatial patterns and variations in VIs within mango orchards and their association with mango yield can be gained.

A number of existing studies in the horticultural tree crop sector have tested the use of different statistical and ML approaches for the in-season yield estimation of different fruit trees [17–19]. For example, Matese and Di Gennaro [19] applied a Gaussian process regression and NDVI acquired from a UAV to evaluate the performance of traditional linear and ML regression in forecasting in-season grape yield with an accuracy of 85.95%. Gan et al. [20] detected and counted fruits using a combination of a thermal imaging method (from a vehicle mounted camera (FLIR A655SC)) and ML (Region-Convolutional Neural Network (Faster R-CNN)) to tackle problems associated with colour similarity between immature citrus fruits and leaves, achieving an accuracy of 96%. Additionally, Apolo-Apolo et al. [18] applied Faster R-CNN to detect small target fruits from top-view RGB images of apple trees captured by UAVs with accuracy above 90%. In the mango industry specifically, only a handful of studies have conducted in-season counting of fruits on trees. Notable among them are the studies by Rahman et al. [6] and Anderson et al. [10], who predicted in-field, in-season (for two seasons) mango yield for mango with an “18 tree calibration” approach using VHR WV3 imagery. Although the approaches discussed above produced accurate results, they require physical measures of fruit counts/weights from many individual trees for model calibration [6,10,15,17].

The current study is thus meant to expand the initial findings from those studies, particularly the work of Rahman et al. [6], who primarily used an Artificial Neural Network (ANN) on single-capture WV3 imagery over three mango orchards in the Australian Northern Territory. That method was tested for only one region, two seasons and one cultivar. Although there is scope for the application of other machine learning and deep learning approaches to model mango yield using remote sensing, the current study, firstly, aims to expand the scope of that research using multiple farms, growing seasons and regions in Australia. This will help to explore the potential of predicting mango yield without the need for in-field calibration. The application of deep learning approaches has the potential for generalisation across a wider range of geographical coverage and such methodologies could be explored. Secondly, the current study aims to investigate whether a generic model is possible using this large dataset. Thus, the Random Forest approach, which has been widely used due to its user friendliness in practical applications with the ability to produce accurate predictions, as reported by numerous studies by describing more variability in the data, was explored [2,21–23]. The RF model will be particularly useful in capturing the spatial variability in mango yield across different numbers of farms, growing seasons and regions in Australia. This will contribute to the understanding of the spatial variability of yield. The successful implementation of this approach would not only promote adoptability for individual growers but also potentially support scalability to multiple scales (block, farm, regional and national levels).

The objectives of this study are as follows:

1. Explore the relationships between VIs derived from WV2 and WV3 imagery and fruit count at the individual tree level, using data sourced from different growing seasons, locations and cultivars.
2. Evaluate a range of analytics to determine if a generic crop load (yield) model can be derived between canopy reflectance and yield.
3. Validate the accuracies of a generic model for estimating fruit number at the individual tree and orchard block level.
4. Produce tree-level yield variability maps.

## 2. Materials and Methods

### 2.1. Study Area

The study was conducted from 2015 to 2021, including 1958 individual trees from 55 orchard blocks across 14 farms in three mango growing regions of Australia, namely, the Northern Territory (NT), Northern Queensland (N-QLD) and South East Queensland (SE-QLD). The study regions are located between longitudes 131.02°E and 152.41°E and latitudes 12.73°S and 27.91°S. Table 1 describes the study sites with information including regions, the location of individual farms, cultivar, number of blocks, etc. Eight cultivars including Calypso, Kensington Pride (KP), Honey Gold (HG), Parvin, R2E2, Keitt, Lady Jane (LJ) and Lady Grace (LG) were studied. The regions are characterised by a tropical climate with distinct dry (May to October) and wet (November to April) seasons. For NT, average daily minimum and maximum temperatures range between 21 °C and 35 °C in the dry season, with average annual rainfall of 1570 mm and wet season temperatures ranging between 25 °C and 33 °C [24]. For the Queensland regions, minimum and maximum temperatures range between 21 °C and 30 °C with average annual rainfall of 1030 mm in SE-QLD and 2010 mm in N-QLD [25]. These study areas, apart from being the regions in Australia with the highest mango production volumes, are also locations that have benefited significantly from previous research and have available data for exploring various forms of analytics. Figure 1 shows the spatial distribution of the study farms.

Figure 2 show the flowchart that outlines the procedural steps used to generate the results. Each component of the flowchart refers to a section describing the methodology in greater detail.

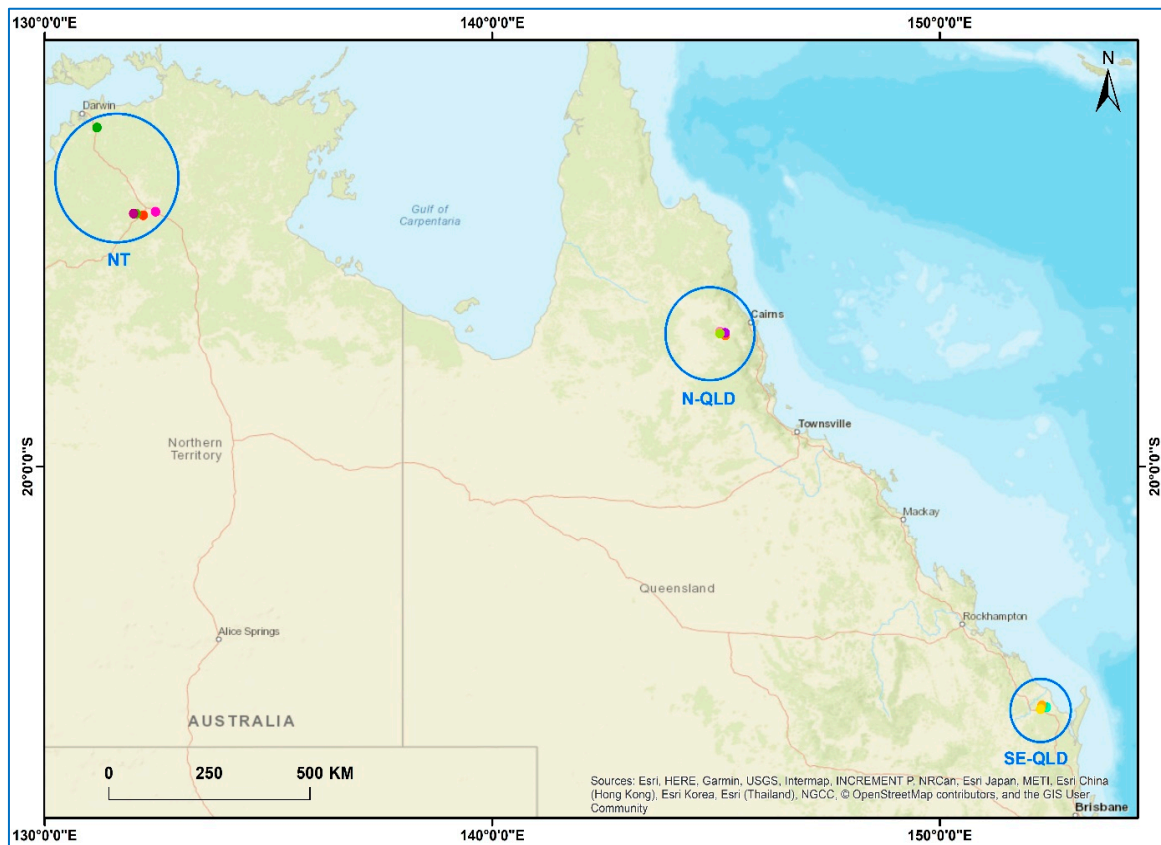


Figure 1. Location of mango farms in the three mango growing regions of Australia.

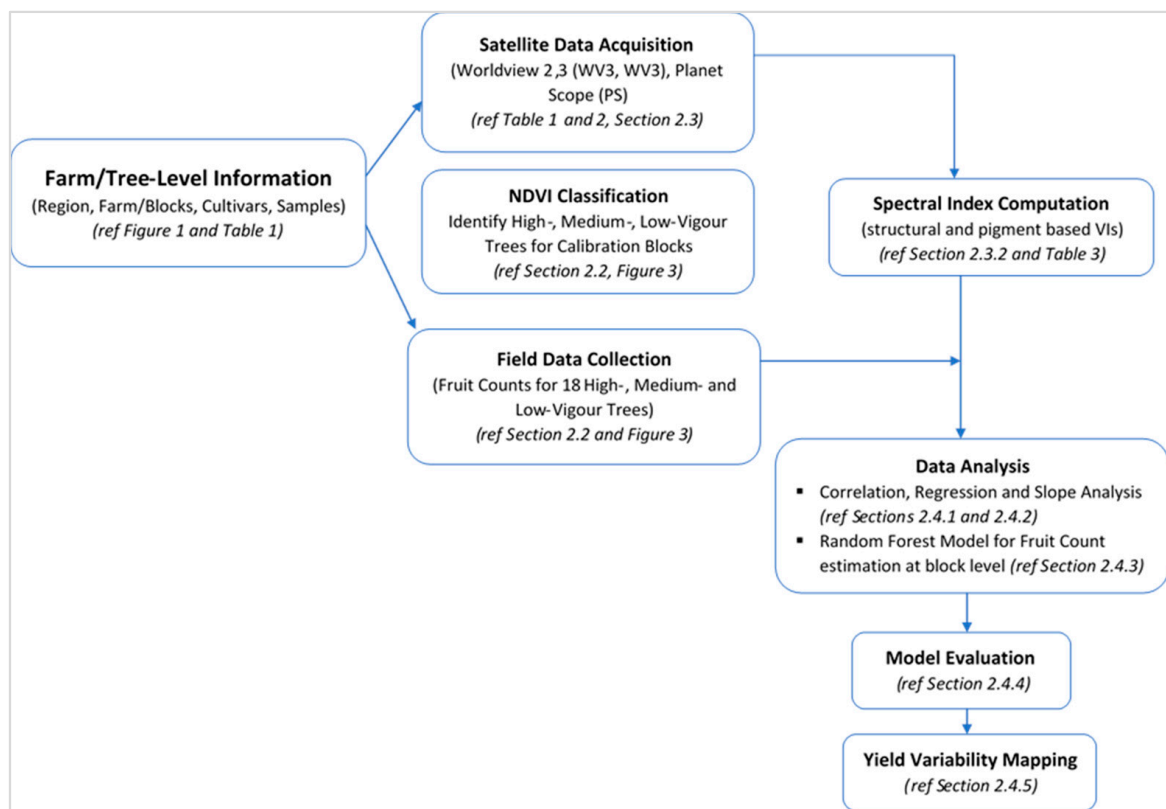


Figure 2. Flowchart showing the sequence of procedure steps used in this study to generate the results.

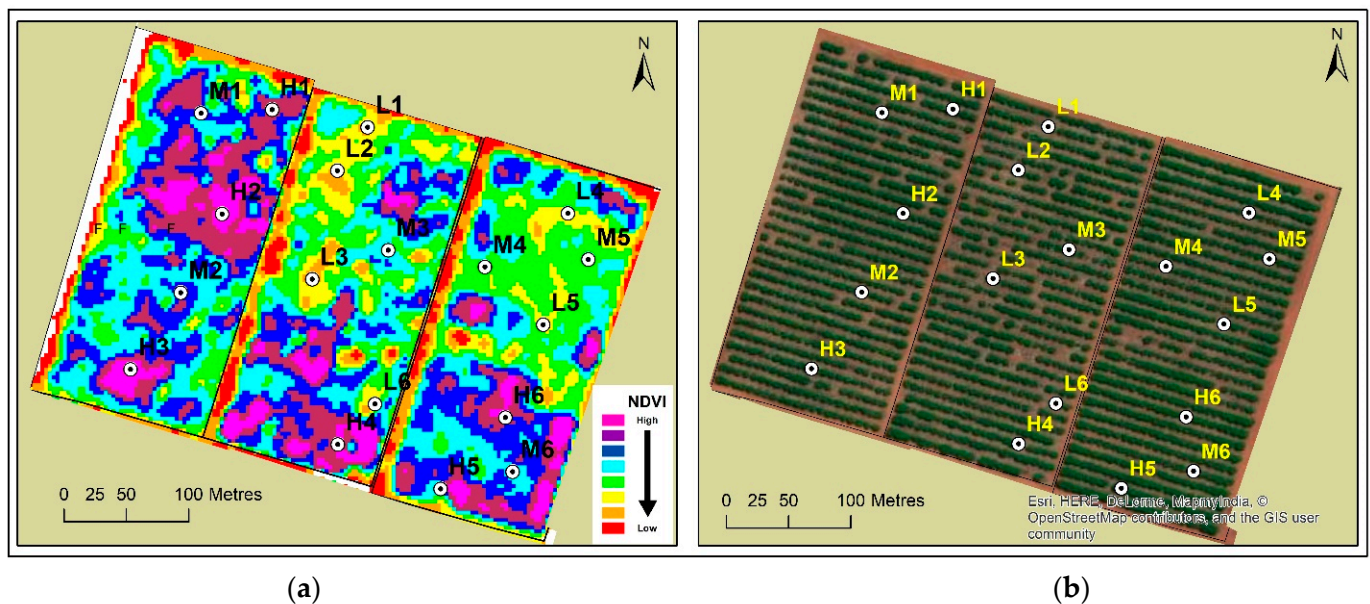
**Table 1.** Description of study farms and their respective WV2 and WV3 imagery capture dates.

Region	Farm	No. of Blocks	Season	Cultivar	No. of Sampled Trees	Satellite	Image Acquisition Date
NT	Farm 1	6	2016	Calypso	432 *	WV3	23-10-2016
			2017			WV3	16-08-2017
			2019			WV3	27-08-2019
			2020			WV3	27-08-2020
			2021			WV2	01-09-2021
	Farm 2	5	2020	KP, R2E2, Parvin	158 *	WV3	27-09-2020
			2021			WV3	23-09-2021
	Farm 3	6	2020	KP, R2E2	144 *	WV3	27-09-2020
			2021			WV3	23-09-2021
	Farm 4	5	2020	KP, R2E2, LG, LJ	180	WV3	04-11-2020
			2021			WV2	16-11-2021
	Farm 5	2	2020	Calypso	36 *	WV3	27-09-2020
			2021			WV3	23-09-2021
Farm 6	2	2020	HG	72	WV3	04-11-2020	
		2021			WV2	16-11-2021	
Farm 7	2	2016	KP, R2E2	72	WV3	23-10-2016	
		2019			WV3	27-08-2019	
Farm 8	5	2019	Calypso, HG	126 *	WV3	08-12-2019	
		2020			WV3	07-11-2020	
Farm 9	4	2020	KP, R2E2, Keitt	72	WV3	07-11-2020	
N-QLD	Farm 10	4	2019	Calypso	144	WV3	08-12-2019
			2021			WV3	07-11-2020
	Farm 11	4	2019	KP, HG, R2E2	108 *	WV3	08-12-2019
			2020			WV3	07-11-2020
	Farm 12	2	2020	KP, R2E2	36	WV3	07-11-2020
Farm 13	3	2019	Calypso	108	WV3	06-12-2019	
		2020			WV3	06-12-2020	
SE-QLD	Farm 14	5	2015	Calypso, HG, R2E2	270 *	WV2	02-09-2015
			2016			WV3	23-09-2016
			2017			WV3	14-05-2017
			2019			WV3	06-12-2019
			2020			WV3	06-12-2020
Total		55			1958		

\* Numbers of 18 tree datapoints per farm vary in some cases as a result of variation in the number of years with available data per block. For example, out of the six blocks in Farm 1, three blocks had 18 tree sample datapoints for only three years out of the five-year period of data collection.

## 2.2. Field Data Collection

For each mango farm, a representative block (calibration block) was selected for the targeted ground truthing of 18 trees representing high, medium and low vigour (i.e., 6 replicate trees per vigour zone). Initially, either PlanetScope or WV satellite data were used to derive a Normalized Difference Vegetation Index (NDVI)-based mango tree vigour map, classified into 8 classes from low to high vigour. The tree vigour map was superimposed on a high-resolution ESRI basemap image in ArcGIS [26] to identify 6 trees in each of high-, medium-, and low-vigour categories (total of 18 trees per block). These trees thus represented the spatial variability in tree vigour in each block. Figure 3 shows an example of 18 tree locations on a classified NDVI map and ESRI basemap image in ArcGIS.



**Figure 3.** Example of 18 tree locations on the classified NDVI map (a) and on the ESRI basemap image (b). The points with L, M and H prefixes represent the different tree vigour classes of low, medium and high, respectively.

The latitude and longitude of sample trees were extracted, along with their positions in terms of row and tree numbers. The coordinates and row/tree numbers were used in DGPS to identify the same trees in the field. For each identified tree, field measurements of fruit numbers (fruit count) were carried out manually at the stone hardening stage, at least six weeks before harvest timing (August–December, depending upon the growing region and the harvest timing). The counting was performed by three different individuals, and was then averaged to obtain the total number of fruits/trees. The same trees were used for sampling across the different seasons. In total, tree-level data collected over 55 blocks between 2015 and 2021 were used for model development and validation. Additionally, the actual total yield (fruit count at harvest) of 29 unique blocks over the same period was obtained from the growers. This information was used for further comparisons at the block level.

### 2.3. Satellite Data

VHR WorldView-2 (WV2) and WorldView-3 (WV3) satellite imagery was acquired for the study regions from MAXAR (<https://earth.esa.int/eogateway/missions/worldview> (accessed on 28 October 2024)). Images had a spatial resolution of 1.6 m and 1.24 m, respectively, suitable for tree-level analysis, with eight multispectral visible near-infrared (VNIR) bands [27], used for the derivation of Vis in this study (Table 2). The pansharpened (PS) images with spatial resolutions of 0.4 m and 0.31 m (at nadir) for WV2 and WV3, respectively, were used for tree crown delineation. Despite the very high spatial resolution of the WV2 and WV3 sensor systems used, their spectral and radiometric capabilities are largely limited [27].

A total of 17 images covering 14 farms captured between 2 September 2015 and 16 November 2021 (Table 1) were acquired. The image capture dates were timed to coincide with the fruit development (FRD) or stone hardening stage of the mango fruit [16]. In cases where WV3 images were unavailable, WV2 data were used as replacement.

**Table 2.** Spectral characteristics of the WV2 and WV3 imagery used in this study.

Image Band	Band Name	Wavelength (nm)
1	Coastal (C)	400–450
2	Blue (B)	450–510
3	Green (G)	510–580
4	Yellow (Y)	585–625
5	Red (R)	630–690
6	Red-edge (RE)	705–745
7	NIR-1	770–895
8	NIR-2	860–900/1040 *

\* NIR-2 wavelength for WV2 and WV3 ranges from 860 to 1040 and from 860 to 900 nm, respectively.

### Spectral Data Extraction and VI Calculation

The canopy reflectance data for each sampled tree were extracted for each spectral band (Table 2), which were then used to derive 24 structural and pigment-based Vis (Table 3). The Vis were calculated taking advantage of the spectral characteristics of each band in the multispectral WV2 and WV3 data applying the mathematical formulas displayed in Table 3 in the R statistical package. Typically, indices that are critical for the assessment of vegetation properties as shown in Table 3 were considered.

**Table 3.** Formula and characteristics of spectral vegetation indices used in this study.

Vegetation Index	Formula	Reference
Red-edge Normalized Difference Vegetation Index (RENDVI)	$(RE - R)/(RE + R)$	[28]
Normalized difference Red-edge index (N1/RENDVI)	$(NIR1 - R)/(NIR1 + RE)$	[29]
Normalized difference Red-edge index 1 (N1RENDVI)	$(NIR1 - RE)/(NIR1 + RE)$	[29]
Normalized difference Red-edge index 2 (N2RENDVI)	$(NIR2 - RE)/(NIR2 + RE)$	[29]
Transformed Chlorophyll Absorption in Reflectance Index (TCARI)	$3 \times ((RE - R) - 0.2 \times (RE - G) \times (RE/R))$	[30]
Structure Insensitive Pigment Index (SIPI)	$(NIR1 - B)/(NIR1 + R)$	[31]
Structure Insensitive Pigment Index (CB SIPI)	$NIR1 - CB)/(NIR1 + CB)$	[31]
Normalized Difference NIR Index (N1/N2NDVI)	$(NIR1 - R)/(NIR1 + NIR2)$	[32]
Green Normalized Difference Vegetation Index (N1GNDVI)	$(NIR1 - G)/(NIR1 + G)$	[33]
Normalized Difference Vegetation Index (N1NDVI)	$(NIR1 - R)/(NIR1 + R)$	[34]
Normalized Difference Vegetation Index (N2NDVI)	$(NIR2 - R)/(NIR2 + R)$	[34]
Renormalized Difference Vegetation Index 1 (RDVI1)	$(NIR1 - R)/(\text{SQRT}(NIR1 + R))$	[32]
Renormalized Difference Vegetation Index 2 (RDVI2)	$(NIR2 - R)/(\text{SQRT}(NIR2 + R))$	[32]
Modified Simple Ratio (MSR)	$(NIR1/R - 1)/(\text{SQRT}((NIR1/R) + 1))$	[35]
Transformed Difference Vegetation Index 1 (TDVI1)	$1.5 \times ((NIR1 - R)/(\text{SQRT}(NIR1^2 + R + 0.5)))$	[36]
Transformed Difference Vegetation Index 2 (TDVI2)	$1.5 \times ((NIR2 - R)/(\text{SQRT}(NIR2^2 + R + 0.5)))$	[36]
Ratio Vegetation Index (RVI)	$(NIR1)/(R)$	[37,38]
Yellow Soil Adjusted Vegetation Index (Yellow SAVI)	$(NIR1 - CB) \times (1 + 0.5)/(NIR1 + CB + 0.5)$	[39]
Enhanced Vegetation Index 1 (EVI2N1)	$2.5 \times ((NIR1 - R)/(1 + NIR1 + (2.4 \times R)))$	[40]
Enhanced Vegetation Index 2 (EVI2N2)	$2.5 \times ((NIR2 - R)/(1 + NIR2 + (2.4 \times R)))$	[40]
Chlorophyll Index Green 1 (CIg_1)	$(NIR1)/(G) - 1$	[41]
Chlorophyll Index Green 2 (CIg_2)	$(NIR2)/(G) - 1$	[41]
Chlorophyll Index Red-edge 1 (CIRe_1)	$(NIR1)/(RE) - 1$	[41]
Chlorophyll Index Red-edge 2 (CIRe_2)	$(NIR2)/(RE) - 1$	[41]

## 2.4. Data Analysis

### 2.4.1. Correlation Analysis

Exploratory data analysis was conducted to understand the data and its distribution. Thereafter, correlation analysis between the Vis and fruit count was performed, first for the entire dataset and then its subsets (location, cultivar and season). This was carried out to assess the direction, strength and statistical significance of their relationship. The analysis was conducted using the R statistical software version 4.1.2 which also assessed the Pearson's correlation coefficient for all the relevant combinations. The correlations were then plotted using the "ggplot" package in the R software version 4.1.2 to facilitate visual interpretation of the relationships and their respective strengths.

### 2.4.2. Linear Regression and Slope Analysis

To explore the relationship between the response (fruit count) and predictor variables (Vis), a multiple linear regression analysis was conducted in Python using the scipy.stats library (<https://www.scipy.org/>). The slopes of the regression lines were analysed using a two-tailed paired *t*-test. The *t*-test was conducted to test the null hypothesis that the means of the slopes from the relationships between Vis and fruit count are equal to zero. This analysis was conducted for the combined dataset containing all individual trees and subsets of this dataset (location (region), cultivar and season). To satisfy normality requirements for statistical testing, the sampling framework for selecting mango trees was random.

### 2.4.3. Random Forest Prediction of Fruit Count

Fruit count was predicted using an RF model. Knowing that most of the Vis used were highly autocorrelated, the RF model was optimised to make it less susceptible to overfitting. Thus, to parameterize the model to select the optimum number of trees and mtry and to perform independent evaluation, a k-fold (with  $k = 10$ ) cross validation (CV) with 5 iterations was conducted during model training [42] using the "caret" package in the R software version 4.1.2. This reduced the detrimental impact of multicollinearity between predictor variables to ensure the models were stable [2,43]. The model developed using all datasets (combined model) was trained on 80% of the dataset using the CV approach mentioned above and tested on the remaining 20%. The sample selection was performed randomly. Additionally, models were trained and tested on subsets of the data, including season, location and cultivar, using the 80%/20% approach to assess which method produces better accuracy at the individual tree level. The location scale used in this study was "region" and, thus, all location references made are in that respect. A different approach, the leave-one-year-out (LOYO) approach, following Torgbor et al. [2], was used for the seasonal prediction model in which one season is held out as a test set and all other seasons are used for model training. All 24 Vis were used as predictors in the models.

The results from these models were graphically displayed and compared. Subsequently, the best approach was applied in validating fruit count (yield) prediction at the block level. The predicted block-level yield,  $Y_{pred}$  (number of fruits), was computed from the product of the predicted average fruit count ( $FC_{pred}$ ) per tree and the total number of trees per block ( $N$ ) [10] as shown in equation 1. The  $FC_{pred}$  per tree was obtained from the final model calibrated using a total of 1940 individual trees' data and excluding the 18 trees of the actual block being validated for a given season.

$$Y_{pred} = FC_{pred} \times N \quad (1)$$

This was compared with the grower-obtained actual yield (fruit count for blocks with available yield data at harvest spanning the period from 2016 to 2021) from the respective blocks to assess the performance of the model at the block level.



#### 2.4.4. Model Evaluation

Several model evaluation metrics, including mean absolute error (MAE) and root mean squared error (RMSE), have been used in the past [8,44,45]. The models in this study were evaluated using MAE and percentage root mean squared error (PRMSE) as shown in Equations (2) and (3), respectively. The MAE was used because, compared to the RMSE, it is less sensitive to outliers [46,47].

$$\text{MAE}(\text{fruits/tree}) = \frac{1}{N} \times \sum_{i=1}^N |\bar{y}_i - y_i| \quad (2)$$

where  $N$ ,  $\bar{y}_i$  and  $y_i$  are the number of observations and the predicted and actual fruit count, respectively.

The PRMSE is an easy-to-interpret metric that explains the performance of the model in relation to the predicted feature (i.e., fruit count). It is the ratio of the model's RMSE to the absolute difference of the predicted minimum ( $|P_{\min}|$ ) and maximum ( $|P_{\max}|$ ) fruit count expressed as a percentage. The sampling framework for selecting the mango trees was randomised to satisfy normality requirements for statistical testing.

$$\text{PRMSE}(\%) = \frac{\text{RMSE}}{(|P_{\max}| - |P_{\min}|)} \times 100 \quad (3)$$

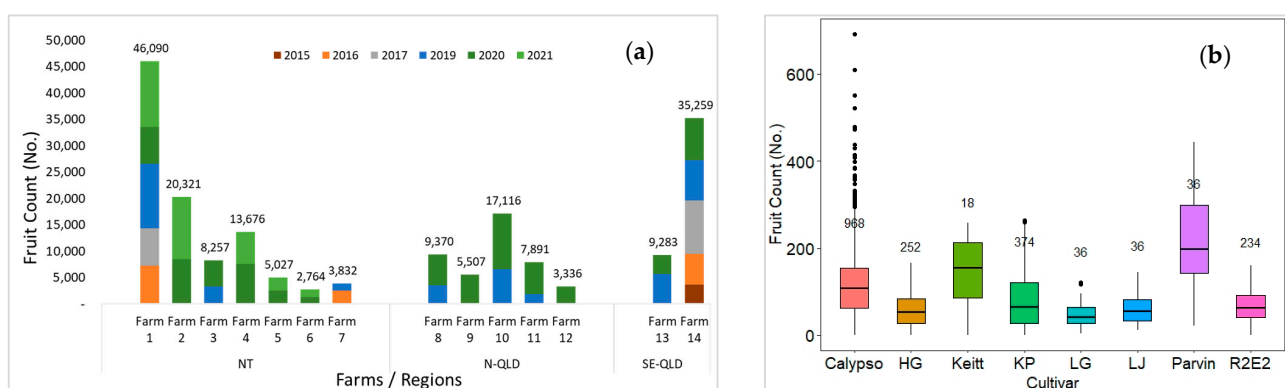
#### 2.4.5. Yield Variability Mapping

To show the spatial variability of yield across orchard blocks for grower management decision support on the precise application of farm inputs, among others, a surrogate yield map was produced from the region-based RF model. The map was produced with ESRI ArcGIS version 10.8 [26], using the predicted yield obtained from the RF model. The yield predicted per tree from the RF model was grouped and assigned a unique colour. This helped categorise trees with similar yield profiles and offered an opportunity for site productivity assessment.

### 3. Results

#### 3.1. Exploratory Data Analysis

The NT had the highest number of farms and datapoints. Farm 1 contributed the highest number of trees sampled (432) and fruits counted (46,090), followed by Farm 14 with 270 sampled trees and 35,259 fruits counted across the six-year period. The least number of trees sampled and fruits counted was found on Farms 5 and 12 with 36 trees each and respective fruit counts as shown in Figure 4a. While majority of the cultivars planted produced median number of fruits below 100, Calypso, Keitt and Parvin produced median fruit counts above 100 fruits/tree (Figure 4b).



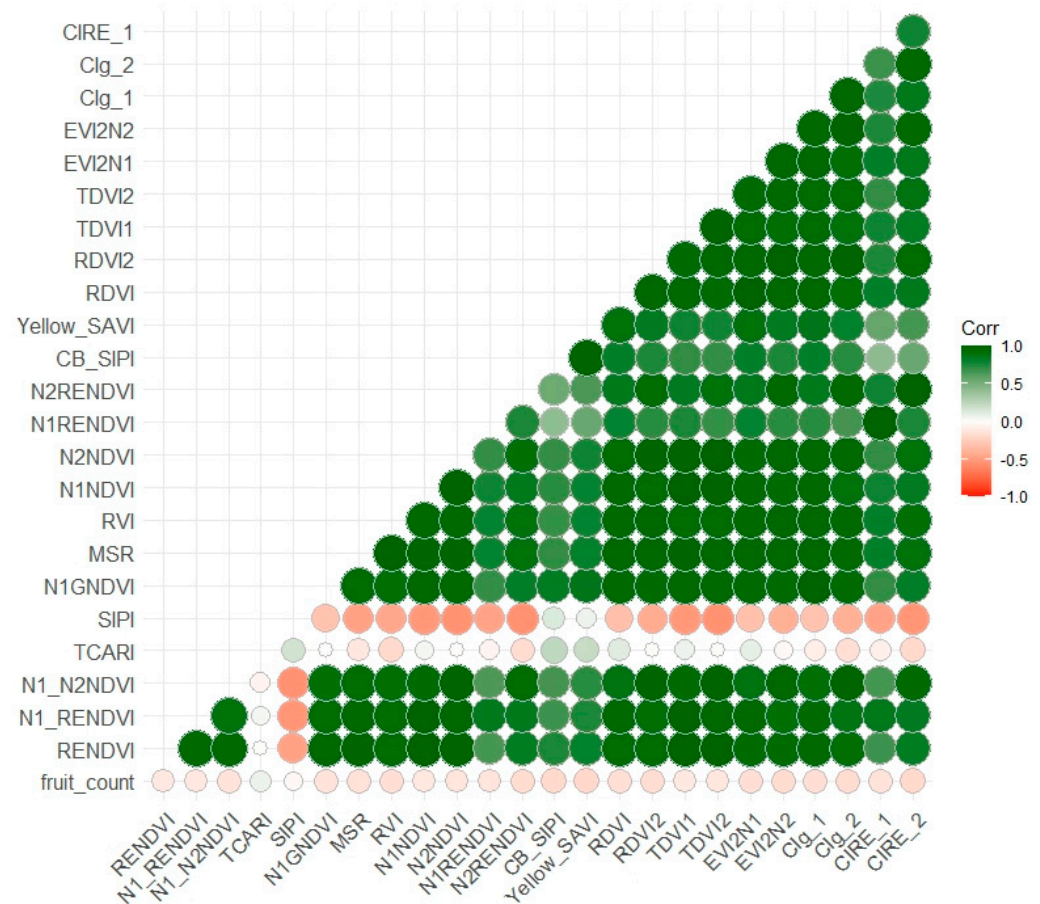
**Figure 4.** Summary of fruits counted (a) per farm and (b) heterogeneity of cultivar yield distribution from 2015 to 2021. The numerical values and black dots associated with each boxplot represent the number of trees of that particular cultivar and outliers, respectively.

### 3.2. Exploring the Relationship Between VIs and Fruit Count

The results of the three separate experiments conducted to assess the relationship between VIs and fruit count are presented in this section. They include (1) regression analyses for all datasets combined, (2) separate cultivar and region subset regressions and (3) meta-regression analyses of the yield vs. VI slopes of all blocks.

#### 3.2.1. All Data Aggregation Results

The correlation coefficients ( $r$ ) of the VIs with the highest correlation of fruit count for all blocks and all seasons as well as cultivar and location data subsets were assessed. The results show that considering all blocks, including their respective locations, cultivars and seasons, N1RENDVI, CIRE\_1 and SIPI produced the highest correlation with  $r = 0.14$ ,  $0.14$  and  $-0.14$ , respectively (Figure 5).



**Figure 5.** Correlation between fruit count and the 24 VIs using the entire datasets of 1958 datapoints. The green and red colour ramps show the strength and direction of the correlation being positive and negative, respectively.

#### 3.2.2. Separate Cultivar and Region Regression Results

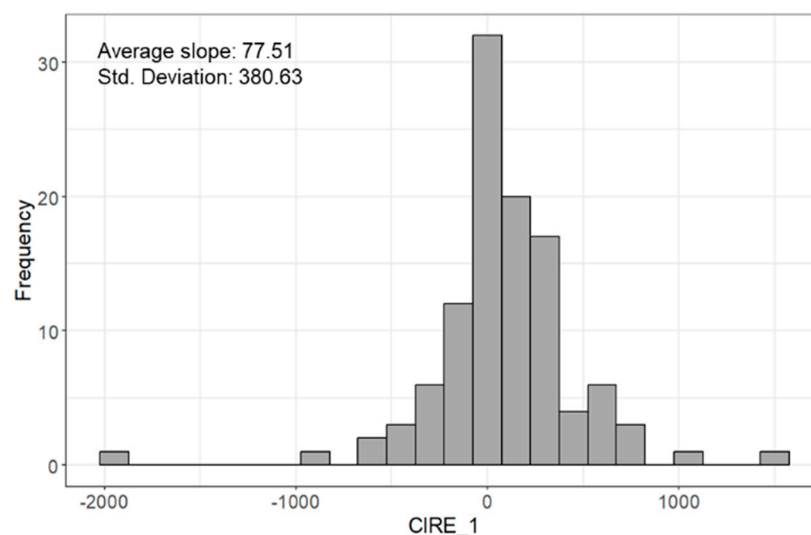
Subsequently, for the cultivar and location subsets, N2RENDVI produced the highest correlation with  $r = 0.76$  (maximum from the Keitt cultivar) and  $0.34$  (from SE-QLD), respectively. Table 4 provides details on the correlation analysis results from the cultivar and location subsets. In general, the correlation coefficient obtained for the highest correlating VI(s) was higher in the subset models than in the combined model.

**Table 4.** Output from the correlation analysis on cultivar and location subsets.

Subset	Description	Best Correlation Coefficient (r)	Best Contributing VI (s)
Cultivar	Calypso	0.24	CIRE_2
	KP	−0.19	CIRE_2 and CB-SIPI
	HG	0.39	CIRE_2, N2RENDVI, TDVI1 and N1/N2NDVI
	Parvin	−0.69	EVI2N2
	R2E2	0.35	CB-SIPI
	LJ	0.15	Yellow-SAVI
	LG	0.51	CB-SIPI
	Keitt	0.76	N2RENDVI
Region	NT	−0.18	SIPI
	N-QLD	−0.15	Yellow-SAVI
	SE-QLD	0.34	CIRE_2 and N2RENDVI

### 3.2.3. Individual Block Meta-Regression Analysis Results

In an attempt to better quantify the influence of location, cultivar and season on the relationship between VIs and fruit count, a meta-regression analysis of slopes was conducted using a two-tailed *t*-test of the regression lines per block for all datasets and subsets. The results show that for all 24 VIs tested on the entire datasets, only two (N1RENDVI with  $p = 0.044$  and CIRE\_1 with  $p = 0.036$ ) provided enough evidence for the rejection of the null hypothesis at  $p < 0.05$ . Figure 6 shows a histogram of CIRE\_1 with the average slope and standard deviation. It was therefore concluded that the mean of the slopes for the relationship of fruit count with these indices was significant.

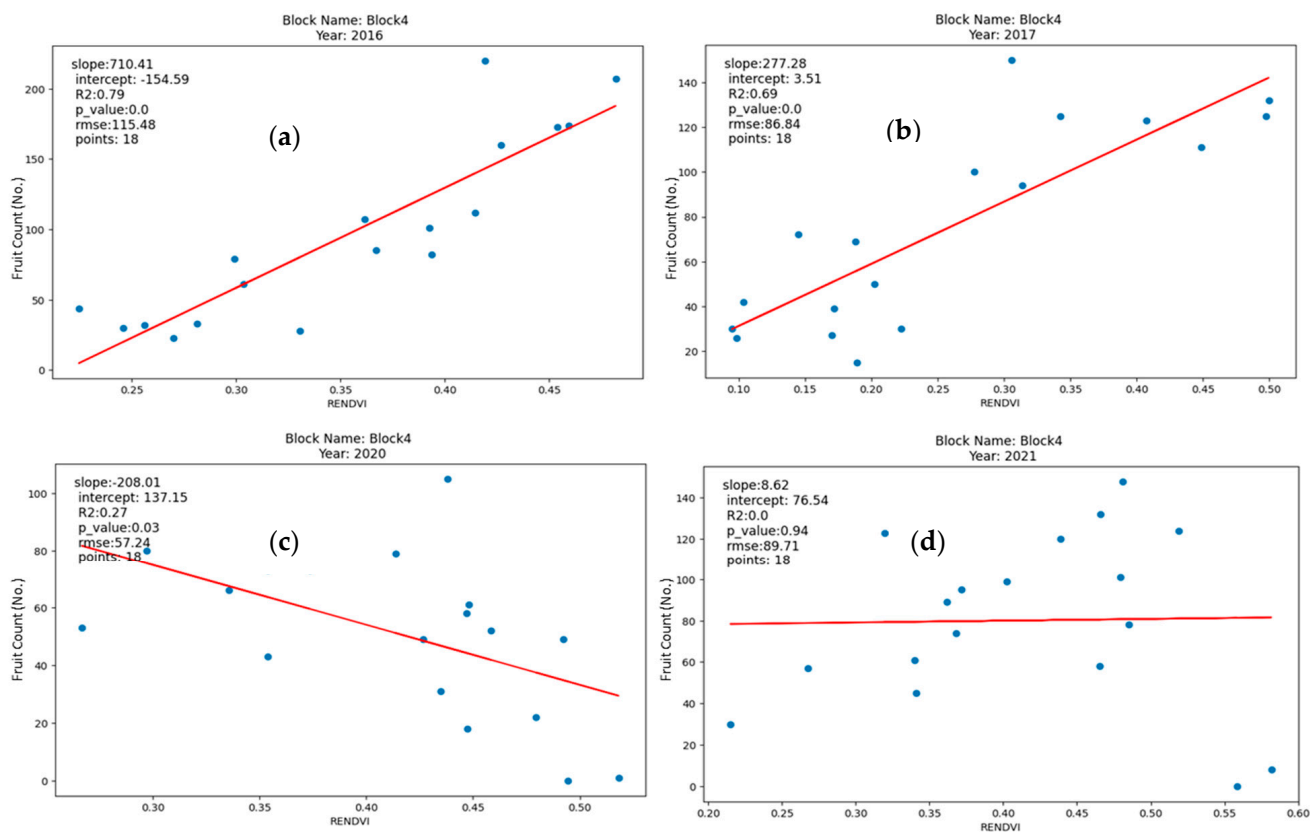
**Figure 6.** Distribution of slopes for CIRE\_1 with average slope and standard deviation.

Further analysis conducted on subsets of the data to investigate the effect of location, cultivar and season on the relationship between VIs and fruit count using all the datapoints for each group revealed the following:

1. Location (region): the mean of all slopes tested was not significantly different from zero in the NT, whereas one VI (SIPI with  $p = 0.023$ ) in the N-QLD and two VIs (N1RENDVI with  $p = 0.041$  and CIRE\_1 with  $p = 0.038$ ) in the SE-QLD regions were significantly different from zero.

- Cultivar: there were significant differences in the mean of the slopes for all VIs except SIPI for both KP and Calypso and N1\_N2NDVI for LJ.
- Season: The mean of the slopes for 2015, 2020 and 2021 was not significantly different from zero and the null hypothesis was therefore accepted. However, slopes were significantly different in the 2016, 2017 and 2019 seasons.

Generally, three different kinds of relationships were identified between VIs and fruit count for all the VIs tested across different seasons. VIs were either positively, negatively or not related with fruit count as shown by the regression lines. Figure 7 shows the relationships obtained for Block 4 from Farm 1 using RENDVI over four seasons.



**Figure 7.** Relationships identified between RENDVI and fruit count: (a) and (b) were positive for 2016 and 2017, (c) negative for 2020 and (d) non-existent for 2021.

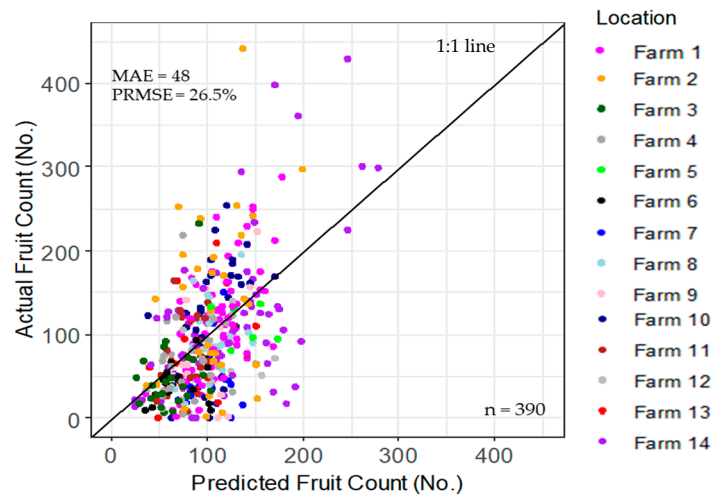
Different relationships were observed in different seasons for the same trees. Similar relationships were identified for the other VIs and blocks as well. There is therefore no universal relationship between all VIs and fruit count across different seasons and blocks.

### 3.3. Random Forest Prediction of Fruit Count at the Individual Tree Level

#### 3.3.1. Fruit Count Prediction Using Combined Datasets

Using the RF model on the entire dataset and splitting it into 80% for training and 20% for testing to predict fruit count produced PRMSE = 26.5% with an MAE of 48 fruits/tree (Figure 8). Zero (0) predicted fruit counts were recorded due to the inclusion of calibration trees without fruits in some blocks across the period. This was accounted for by the model.

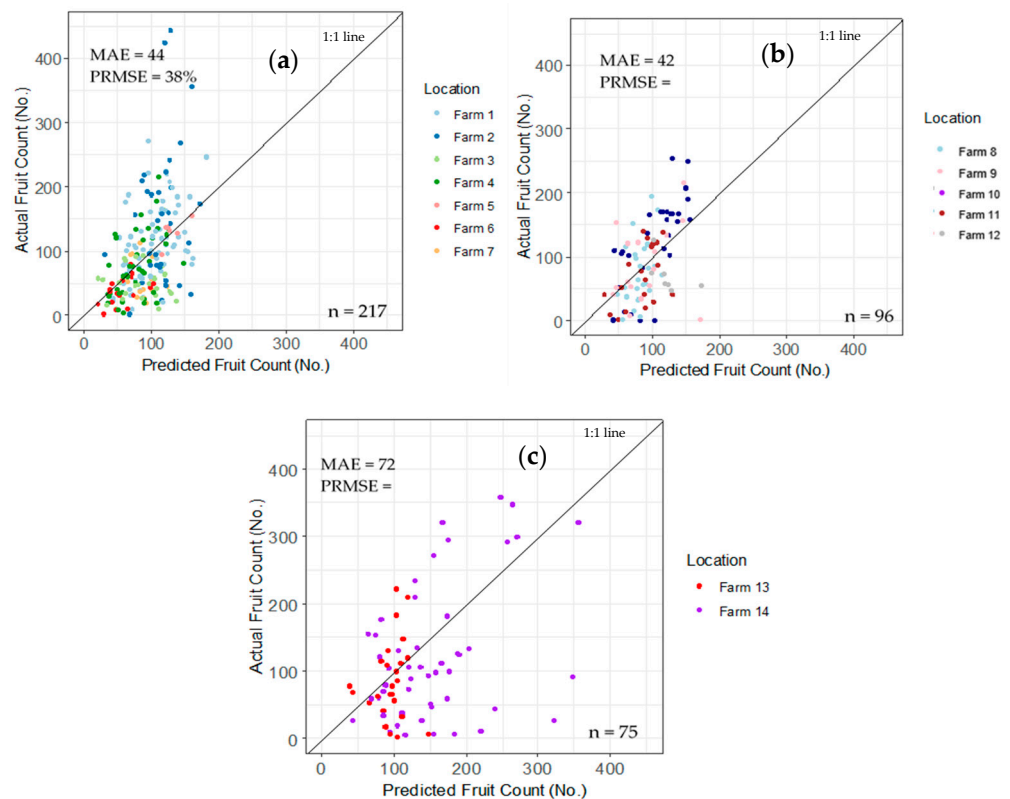
To explore the possibility of improving the prediction, subsets of the data including location, cultivar and seasons were tested. The following section presents the outcome of this trial.



**Figure 8.** RF prediction of fruit count using all individual tree datasets (combined model). The different coloured points represent the sampled trees from the respective farms and regions. n = 390 represents the number of datapoints (20%) used for model validation.

### 3.3.2. Individual Tree Fruit Count Prediction Using Data Subsets (Location, Cultivar and Season)

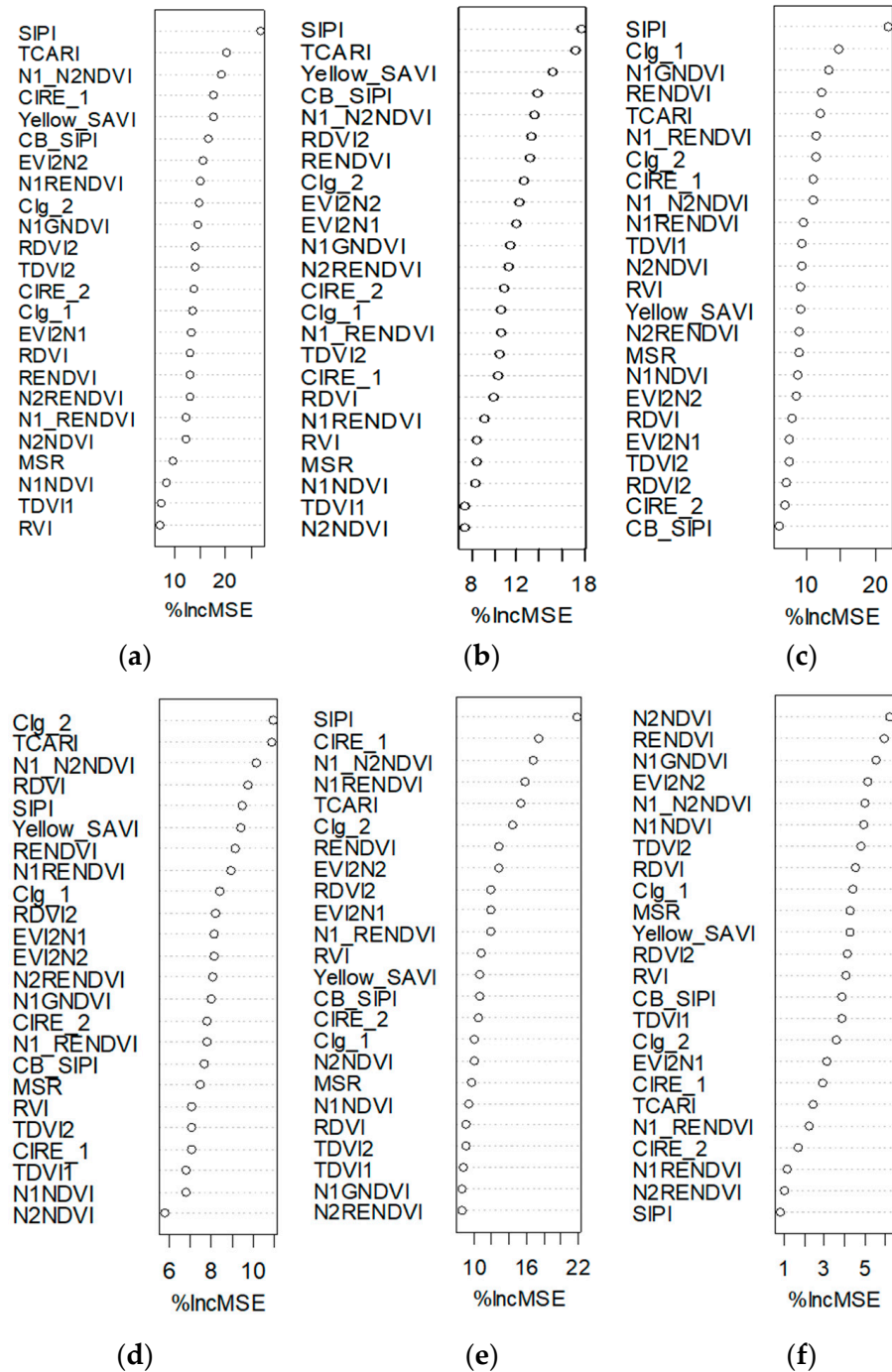
In this section, the results of the individual tree fruit count prediction using the location (region), cultivar and season subsets are shown. Figure 9 shows the RF models for the predictions of fruit number from individual trees based on their growing region from the locational subset model.



**Figure 9.** RF-based location (region) prediction of fruit count in the (a) Northern Territory (NT), (b) Northern Queensland (N-QLD) and (c) South East Queensland (SE-QLD). The different coloured points represent the sampled trees on a given farm in the respective regions.

The PRMSE associated with the SE–QLD region model was 29%. Similar results were produced for the NT region with PRMSE and MAE values of 38% and 44 fruits/tree and for the N–QLD region with PRMSE and MAE values of 36% and 42 fruits/tree, respectively.

We compared models built using all 24 predictors to the top 10 and 6 predictors (as determined by feature importance analysis, Figure 10). It was shown that using all 24 VIs produced slightly better results. Table 5 shows a summary of the feature selection experiment from the regional subset model.



**Figure 10.** RF-based variable importance plots for models from (a) combined datasets, (b) Northern Territory (NT), (c) Northern Queensland (N–QLD) and (d) South East Queensland (SE–QLD) and the best (e) seasonal and (f) cultivar models.

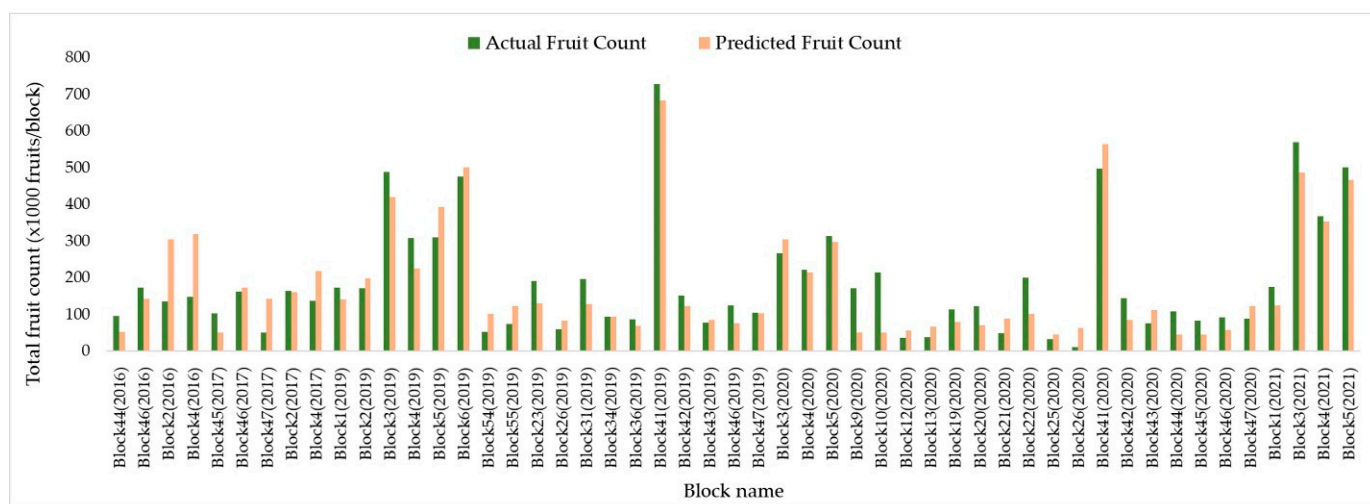
**Table 5.** Regional model PRMSE (%) comparison using different numbers of predictors (all 24 VIs, top 10 and top 6) based on RF-based feature importance ranking.

Regional Model	All 24 Predictors	10 Top Ranked Predictors	6 Top Ranked Predictors
NT	38.1%	42.5%	40.8%
N-QLD	36.3%	37.9%	36.6%
SE-QLD	29.0%	30.0%	29.2%

For the season-based model which was developed using the LOYO approach, 2017 and 2015 produced the best (PRMSE = 32.6%) and worse (PRMSE = 59%) predictions, respectively. From the cultivar-specific modelling experiments, the Parvin cultivar produced the lowest PRMSE of 19.3%, followed by the R2E2 cultivar with a PRMSE of 26.6%. The Lady Jay (LJ) cultivar produced the worse prediction with a PRMSE over 90%. Overall, the results show that the subset model approach is not consistently better than the combined model described in Section 3.3.1.

#### 3.4. Validation of Combined and Subset (Location, Season and Cultivar) Predicted Fruit Count Models at the Block Level

This section focuses on the validation of the combined model developed in Section 3.3.1 using actual harvest data (51 validation points, blocks per season), spanning a five-year period (2016 to 2021, excluding 2018) and covering 29 unique blocks located within the three regions (NT = 26, N-QLD = 8 and SE-QLD = 17 validation blocks per season). The predicted total block yield (fruit count), which is a product of the predicted average number of fruits per tree and the respective total number of trees per block (Equation (1)), was compared with the actual total yield at harvest. The performances of the models (combined datasets and NT-, N-QLD- and SE-QLD-specific models) were compared to test the effect of location on the combined model. Figure 11 shows the results from the validation for all 51 validation points.

**Figure 11.** Comparison of total actual and predicted yield for the 51 validation points (blocks per season) obtained from 29 unique blocks with available actual harvest data from 2016 to 2021.

The accuracy of the combined model on a block-by-block level showed median accuracy of 65.5%, with 45% of the blocks producing accuracies greater than 70%. This holds great promise for the industry as a performance guide for general season block-level prediction.

The PRMSE associated with the combined dataset model was 10.1%. Subsequently, the NT-, N-QLD- and SE-QLD-specific models produced PRMSEs of 16.8%, 61.0% and 7.2%, respectively. Table 6 details the performance of the location-specific models. Additionally, the model was validated for seasons (irrespective of region and cultivar) and cultivar (irrespective of region and season). The results of this assessment are also shown in Table 6.

**Table 6.** Model validation for region, season and cultivar at the block-level.

Model	Description	PRMSE (%)	MAE (No. of Fruits/Block) *	R <sup>2</sup>	No. of Calibration Datapoints	No. of Validation Blocks
Location (Region)	NT	16.8	59.2	0.75	1940	26
	N-QLD	61.0	41.9	0.78		8
	SE-QLD	7.2	40.6	0.93		17
Season	2016	46.2	103.4	0.18	1940	4
	2017	35.8	47.8	0.17		5
	2019	7.7	39.7	0.93		18
	2020	12.2	50.6	0.77		20
	2021	14.4	45.7	0.97		4
Cultivar †	Calypso	10.0	50.7	0.86	1940	35
	KP	174.5	76.3	0.90		6
	HG	44.5	33.4	0.19		7
	R2E2	72.3	38.7	0.98		2

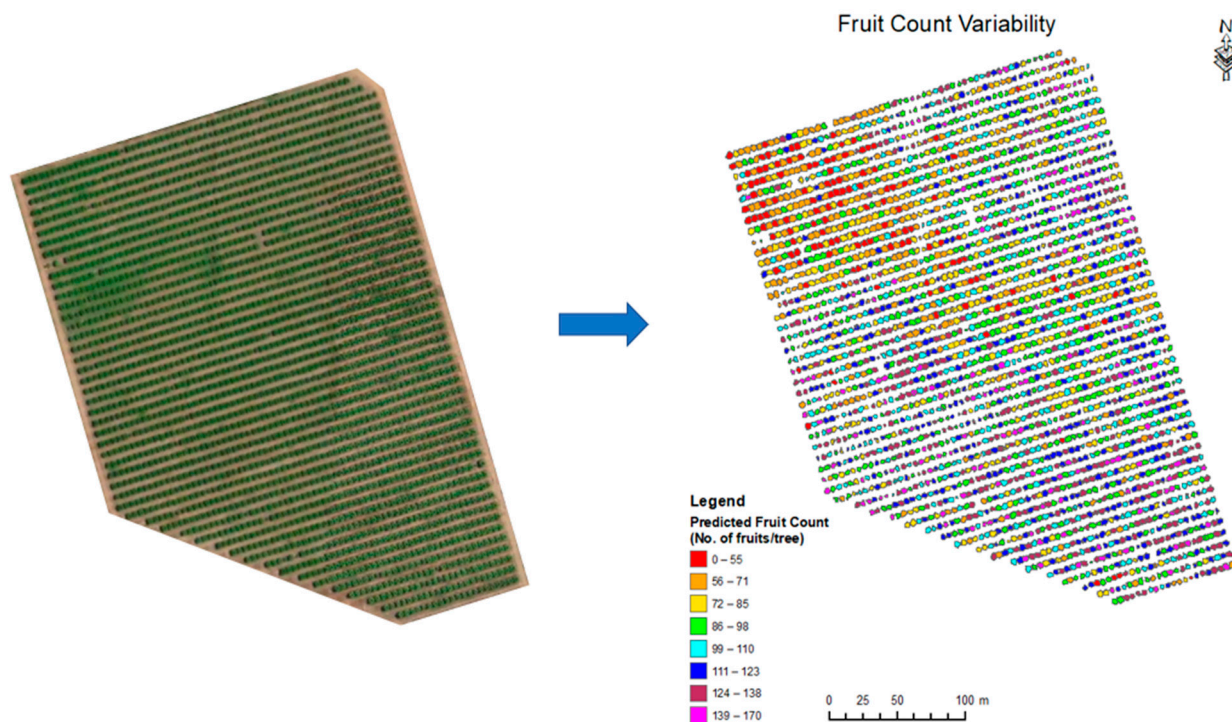
\* MAE values are in  $\times 1000$  of fruits/block per respective region, season or cultivar. † Data were available for the Calypso, HG, KP and R2E2 cultivars. Only one validation point was available for the LJ cultivar.

The 2019 seasonal model produced the lowest PRMSE of 7.7% compared to all other seasons, followed by the 2020 season with a PRMSE of 12.2%. The 2016 and 2017 seasons, with small numbers of validation points (<10), produced PRMSEs of 46.2% and 35.8%, respectively (Table 6). Similarly, in the block-level validation for the cultivar model, Calypso was the most dominant cultivar produced the lowest PRMSE of 10.0%. The KP, HG and R2E2 cultivars produced PRMSEs above 44%, whilst LG, LJ, Keitt and Parvin had either one or no validation points at all for conducting any analysis.

### 3.5. Yield Variability Mapping for a Block at the Tree Level

An example variability map showing the tree-level predicted fruit count per tree is shown in Figure 12, generated using the combined model developed in Section 3.3.1. The map shows that the north-western portion of the block generally had low productivity (red-orange colour) as compared to the south-western and eastern portions of the orchard (magenta). The overall pattern across the field is that 2–3 neighbouring trees have similar productivity levels and then change. Pockets of individual low-performing trees in terms of productivity are found even in the higher potential areas (Figure 12).





**Figure 12.** An example of a tree-level yield variability map derived from the RF-based combined model (right). The RGB image of the mango orchard mapped is shown on the (left). The legend presents an industry-based categorization of yield variability ranging from low (0–55) to high (139–170) for this study.

## 4. Discussion

### 4.1. Relationship Between VIs and Fruit Count

With the aim of developing an accurate mango yield model, the correlation between VIs and yield (fruit count) was tested using all datasets including all blocks and all seasons. Subsequently, subsets of the data in terms of cultivar and location (region) were evaluated. Overall, four VIs (CIRE\_1, N1RENDVI, N2RENDVI and SIPI) were found to correlate better with fruit count at both the combined data and subset scales of analysis. Generally, the correlation improved when the data were split based on cultivar and location. This observation is consistent with Rahman et al. [6], who also found red-edge NDVI and SIPI as the highest correlating VIs with fruit number and weight. Their findings align with Anderson et al. [10], who noted that the observed performance of different VIs in different blocks (and also in different regions and cultivars) is a result of factors such as tree age, cultivar, seasonal and locational difference, and management activities [6,48]. These differences are reflected in the variation in the spectral reflectance characteristics of the tree canopies in different blocks and regions. While this study focused on mango, it has the potential to predict the yield of other horticultural tree crops. For example, Zhu et al. [49] developed a machine learning model that is able to accurately detect the fruit tree canopy and subsequently count fruit. Their study demonstrated the ability to apply UAV-derived images in combination with other machine and deep learning algorithms to improve the accuracy and efficiency of such models. To ensure scalability and transferability, our study, by its nature, shows potential in adopting such methods in combination with very-high-resolution satellite imagery to cover larger geographical areas. Furthermore, Zheng et al. [50] demonstrated the capabilities of multi-view remote sensing (UAV) image detection in combination with Faster R-CNN and FaceNet models in improving the accuracy of recognising strawberry flowers and fruits in central Florida, USA. This method improved the recognition accuracy of strawberry flowers and unripe and ripe strawberry fruit from 76.3% to 97%, 71.6% to 99.1% and 69.8% to 97.2%, respectively.

Three key relationships were identified between VIs and fruit count across the six-year period (Figure 6). The identification of these relationships (positive, negative and non-existent) is well aligned with the observation of Robson et al. [32], who identified similarly variable relationships in Avocado and macadamia in Australia using 18 structural and pigment-based VIs derived from WV3 imagery. For a given block, the relationship could be positive in one season and negative or non-existent in another season as shown in Figure 6. This led to the establishment of the fact that no consistent generic relationship between the predictor and response variables across the period exists [10,48]. This situation could be a result of the irregular bearing habits (biennial bearing (BB)) exhibited by a mango tree [11,51].

#### 4.2. RF Prediction Using All Datasets and Subsets

The lowest prediction PRMSEs from the location (region), cultivar and seasonal subsets were 29%, 19.3% and 32.6%, respectively. Using all the individual tree data combined, the PRMSE was 26.5%. Therefore, in situations where location-, cultivar- or season-specific models are required, the subset models could produce appreciable accuracies. However, it would be more valuable to have a single generic model that is capable of predicting yield with satisfactory accuracy without restrictions from factors such as the cultivar, among others. Additionally, SE-QLD produced a relatively higher average number of fruits per tree (115) than the NT (94) and N-QLD (91) regions. These predicted average fruit counts and associated errors are consistent with the findings of Anderson et al. [10] and Payne et al. [52].

Assessing the number of predictors in the locational models demonstrated that, out of the 24 VIs used, 6 of them (SIPI, TCARI, Yellow\_SAVI, N1/N2NDVI, CIg\_2 and CIRE\_1) were found to contribute more to the prediction from the variable importance plot list (Figure 9) in the RF models for both the combined and subset models. Out of these six VIs, SIPI and TCARI were usually in the top two out of all the models as the most important predictors. Although TCARI is mainly a pigment-related VI, it is also structure-sensitive and resistant to variations in LAI; therefore, it has the ability to capture vital information in the canopy that other VIs may not resolve [30]. Generally, the accuracies achieved using the top six predictors were marginally lower than those obtained using all 24 VIs in the RF model. It is therefore better to use all 24 VIs in a machine learning model like RF to draw as much information as possible from the variability in the data. For the number of predictors involved in this analysis, computational time was not a challenge for the RF model. If a multiple linear regression model were used instead of RF, it is likely that a model with fewer predictors would have been optimal [53].

#### 4.3. Validation of Predicted Fruit Count Models at the Block Level

Although the performance of the block-level combined model, with PRMSE = 10.1% and an R2 of 0.84, was higher than the regional model (SE-QLD) with the lowest PRMSE of 7.2% and an R2 of 0.93, it was comparable with the seasonal and cultivar models with the lowest PRMSE. This observation, which is expressively better than the finding in fruit count prediction at the individual tree level, agrees with the acceptable range of errors for tree crops as reported by a number of studies [2,54]. Additionally, it lays credence to the findings of Brinkhoff and Robson [55], Filippi et al. [56], Deines et al. [57] and Torgbor et al. [2] that the application of yield prediction models developed at finer resolution (e.g., at the tree level) to predict yield at a coarser scale (e.g., at the farm level) tends to improve the accuracy of the model, as overprediction and underprediction errors at finer scales end up cancelling each other out.

Generally, sub-setting the data based on location (region) did not improve the model. Thus, applying the combined model to predict block-level yield in any of the regions could produce appreciable accuracy that can aid grower logistics planning for harvesting and forward selling. This aligns well with the findings of Rahman et al. [6], who concluded in a study that split datasets into a training and a test set in two growing seasons, compared

with combining the seasons, produced better results and showed the influence of location and seasonal variation on the orchards studied. The poor performance of the 2016 and 2017 seasons, the N-QLD regional model and the KP, HG and R2E2 cultivar models with high PRMSEs could be explained by the limited number of datapoints available for the block-level validation. Furthermore, factors such as biennial bearing, tree age and tree density could have contributed to the performance of the models. For example, the effect of soil background and tree size resulting from tree age and tree density can influence the accuracy of VIs derived from the satellite imagery.

#### *4.4. Mapping the Spatial Variability of Tree Yield in an Orchard Block*

In this work, we predicted the spatial distribution of yield in addition to productivity summaries. This is a crucial aspect of yield modelling as it provides information to growers that aids their farm management decisions in applying precision agricultural principles in a number of ways. Firstly, this yield variability map provides a spatial and graphical aid to growers to identify portions of their orchard blocks that are less productive and require attention. For example, fertiliser regimes can be adjusted for the following seasons to increase production in low-productivity areas as was described for mango and macadamia by Robson et al. [32]. Secondly, it helps to provide growers with areas to focus monitoring activities to identify yield-limiting factors such as soil fertility, moisture stress, pests, diseases, etc. [58,59]. This ensures a more efficient use of resources as growers target specific portions of the orchard that require specific attention instead of the whole field input application, which results in wastage [7]. For instance, in the yield variability map produced in this study (Figure 11), the north-western portion of the block had low productivity as compared to the south-western and eastern portions of the orchard. This could be explained by the variation in soil fertility across the field as climate, cultivar, management and other factors are similar across the field. The grower is aided by the map to identify individual trees in the orchard that require attention to improve their productivity. Thirdly, and perhaps most importantly, the growers are able to allocate labour and transport resources for farm operations such as harvesting in a more efficient way when they know what is going to be harvested from the different parts of the fields. In this way, farm costs will be reduced, which will increase the profitability of the farming operations.

Furthermore, the novelty of this research is in the extraction of canopy reflectance values and VIs for individual trees and using an RF prediction model to predict yield at the individual tree and block levels. The yield variability map subsequently has a direct benefit for growers in terms of targeted agronomic management. Thus, it helps in field-based manual measurement. The significance of this research is in its ability to reduce the amount of field work required for manual estimation (i.e., the current traditional method involves the selection of 5–10% of trees in a given block for manual yield estimation). This is through the strategic use of 18 trees of different vigour (low, medium and high) for field sampling and model development, and subsequently predicting yield for the entire block. The new approach not only estimated the yield with satisfactory accuracy but also reduced the cost, labour and time required for such operations. This is therefore a significant contribution to the understanding of mango yield variability at multiple scales using remote sensing and machine learning approaches.

#### *4.5. Limitations of the Study*

Difficulty in obtaining more validation datasets is one of the limitations of this study [60]. The effect of this was seen in the validation of the seasonal and cultivar model, which had a high error rate. This could be due to the limited number of datapoints available, especially for the 2016 and 2017 seasons as well as the KP, HG and R2E2 cultivars (Table 6). The errors were minimised when those seasons and cultivars with a smaller number of observations were grouped together. The assessment of the subset model at the block level was thus limited due the lack of actual block-level data. Therefore, further research could consider validating such models on more blocks in multiple seasons and

cultivars when such data are available. Another key limitation of this study is the use of relatively costly high-resolution satellite data for predicting fruit count. In situations where budgetary allocation for similar projects is limited, implementing such a study could be challenging. Additionally, delineation of individual tree crown areas was a daunting task, although it was a key prerequisite to the development of the yield variability map produced in this study. The delineation of tree crowns can only be performed using a high/very-high-resolution image, which is often expensive. It would therefore be useful in the future to develop an approach that will establish the relationship between freely available medium-resolution and very-high-resolution data for such tasks. This will be particularly useful because, although VHR data have high spatial resolution, they are sometimes limited in spectral and radiometric resolution. The use of other forms of data with varying resolutions will take advantage of the strengths of the other to complement the VHR imagery.

## 5. Conclusions

The current study explored the relationships between vegetation indices and mango yield (fruit count) at the individual tree scale and applied a random forest approach that predicted fruit count at both the tree and block level. The study identified three key relationships between VIs and fruit count, including positive, negative and non-existent, for the different seasons and blocks assessed. It thus demonstrates that no consistent generic relationship exists between the predictor (VIs) and response variables (fruit count) over the six-year period (2015–2021, excluding 2018). For example, a given block could exhibit any of the three relationships in different seasons. Generally, the performance of the RF-based combined and subset models was better at the block level than at the individual tree level. Sub-setting the data based on region, season or cultivar did not improve model performance. Thus, applying the combined model to predict block-level yield in any of the regions could produce appreciable accuracy that could aid grower logistics planning for harvesting and forward selling. We demonstrated that the developed model could be applied to all trees in an orchard, mapping yield variability, which provides a range of commercial benefits to mango growers. It offers them an opportunity to spatially assess productivity at the tree and block level for harvest segregation. Additionally, it affords growers the opportunity to identify potential effects of management, soils, diseases and pests on the productivity of orchards. Future research could thus explore the application of the methods in this study to other horticultural tree crops and also investigate the causes of spatial yield variability in mango orchards. Furthermore, with the availability of other machine and deep learning methods, future studies could explore the application of our method in combination with other approaches that could leverage the use of either UAV data or freely available medium-resolution data like Sentinel-2 to develop accurate and cost-effective yield prediction models. This will reduce the cost of acquiring VHR satellite imagery like the WV2/WV3 used in our study. There is also the potential of detecting and predicting the yield of small fruits like strawberry and others in different geographical regions.

**Author Contributions:** B.A.T.: Conceptualization, Investigation, Methodology, Validation, Data curation, Formal analysis, Visualisation, Writing—original draft, Software. P.S.: Conceptualization, Methodology, Supervision, Visualisation, Validation, Data curation, Writing—review and editing. M.M.R.: Conceptualization, Methodology, Supervision, Validation, Data curation, Writing—review and editing. J.B.: Conceptualization, Methodology, Supervision, Writing—review and editing. A.R.: Funding acquisition, Conceptualization, Methodology, Supervision, Writing—review and editing. L.A.S.: Methodology, Software, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Australian Government Department of Agriculture and Water Resources as part of its Rural R&D for Profit program and by Horticulture Innovation Australia Ltd. [grant number ST15002]; additional support came from a Remote Sensing scholar-

ship granted by the Applied Agricultural Remote Sensing Centre (AARSC) of the University of New England, Australia.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author due to confidentiality agreements with mango farmers.

**Acknowledgments:** The authors are grateful to the Australian Government through the Destination Australia Program (DAP) Scholarship initiative for their support. The support and contribution of the Management of the Forestry Commission, Ghana, as well as the Management and relevant staff of all the mango farms from which data for this study were obtained is highly appreciated.

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

## References

1. FAOSTAT. Value of Agricultural Production. Available online: <https://www.fao.org/faostat/en/#data/QV> (accessed on 8 January 2023).
2. Torgbor, B.A.; Rahman, M.M.; Brinkhoff, J.; Sinha, P.; Robson, A. Integrating Remote Sensing and Weather Variables for Mango Yield Prediction Using a Machine Learning Approach. *Remote Sens.* **2023**, *15*, 3075. [CrossRef]
3. Song, T. Characterization of Soil-Plant Leaf Nutrient Elements and Key Factors Affecting Mangoes in Karst Areas of Southwest China. *Land* **2022**, *11*, 970. [CrossRef]
4. Sarron, J.; Malézieux, É.; Sané, C.; Mango, É.F. Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV. *Remote Sens.* **2018**, *10*, 1900. [CrossRef]
5. Hoffman, L.A.; Etienne, X.L.; Irwin, S.H.; Colino, E.V.; Toasa, J.I. Forecast performance of WASDE price projections for U.S. corn. *Agric. Econ.* **2015**, *1*, 157–171. [CrossRef]
6. Rahman, M.M.; Robson, A.; Bristow, M. Exploring the Potential of High Resolution WorldView-3 Imagery for Estimating Yield of Mango. *Remote Sens.* **2018**, *10*, 1866. [CrossRef]
7. Zude-Sasse, M.; Fountas, S.; Gemtos, T.A.; Abu-Khalaf, N. Applications of precision agriculture in horticultural crops. *Eur. J. Hortic. Sci.* **2016**, *1*, 78–90. [CrossRef]
8. Zhang, Z.; Jin, Y.; Chen, B.; Brown, P. California Almond Yield Prediction at the Orchard Level With a Machine Learning Approach. *Front. Plant Sci.* **2019**, *1*, 809. [CrossRef] [PubMed]
9. Zarate-Valdez, J.L.; Muhammad, S.; Saa, S.; Lampinen, B.D.; Brown, P.H. Light interception, leaf nitrogen and yield prediction in almonds: A case study. *Eur. J. Agron.* **2015**, *1*, 1–7. [CrossRef]
10. Anderson, N.T.; Underwood, J.P.; Rahman, M.M.; Robson, A.; Walsh, K.B. Estimation of fruit load in mango orchards: Tree sampling considerations and use of machine vision and satellite imagery. *Precis. Agric.* **2018**, *1*, 823–839. [CrossRef]
11. He, L. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. *Comput. Electron. Agric.* **2022**, *195*, 106812. [CrossRef]
12. Marani, R.; Milella, A.; Petitti, A.; Reina, G. Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera. *Precis. Agric.* **2020**, *1*, 387–413. [CrossRef]
13. Robson, A.J.; 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]
14. Gutiérrez, S.; Wendel, A.; Underwood, J. Ground based hyperspectral imaging for extensive mango yield estimation. *Comput. Electron. Agric.* **2019**, *1*, 126–135. [CrossRef]
15. Sinha, P.; Robson, A.J. Satellites Used to Predict Commercial Mango Yields. McPherson Media Group (MMG). Available online: <https://www.treecrop.com.au/news/satellites-used-predict-commercial-mango-yields/> (accessed on 12 August 2022).
16. Torgbor, B.A.; Rahman, M.M.; Robson, A.; Brinkhoff, J.; Khan, A. Assessing the Potential of Sentinel-2 Derived Vegetation Indices to Retrieve Phenological Stages of Mango in Ghana. *Horticulturae* **2022**, *1*, 11. [CrossRef]
17. Stein, M.; Bargoti, S.; Underwood, J. Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. *Sensors* **2016**, *16*, 1915. [CrossRef]
18. Apolo-Apolo, O.E.; Perez-Ruiz, M.; Martinez-Guanter, J.; Valente, J. A Cloud-Based Environment for Generating Yield Estimation Maps From Apple Orchards Using UAV Imagery and a Deep Learning Technique. *Front. Plant Sci.* **2020**, *1*, 1086. [CrossRef]
19. Matese, A.; Di Gennaro, S.F. Beyond the traditional NDVI index as a key factor to mainstream the use of UAV in precision viticulture. *Sci. Rep.* **2021**, *11*, 2721. [CrossRef]
20. Gan, H.; Lee, W.S.; Alchanatis, V.; Abd-Elrahman, A. Active thermal imaging for immature citrus fruit detection. *Biosyst. Eng.* **2020**, *1*, 291–303. [CrossRef]
21. Bai, X. Comparison of Machine-Learning and CASA Models for Predicting Apple Fruit Yields from Time-Series Planet Imageries. *Remote Sens.* **2021**, *13*, 3073. [CrossRef]
22. Jeong, J.H. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *1*, e0156571. [CrossRef]
23. Fukuda, S.; Spreer, W.; Yasunaga, E.; Yuge, K.; Sardud, 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**, *1*, 142–150. [CrossRef]

24. NTG. Soils of the Northern Territory—Factsheet. Department of Land Resource Management. Available online: <https://depwv.nt.gov.au/rangelands/technical-notes-and-fact-sheets/land-soil-vegetation-technical-information> (accessed on 8 January 2023).
25. BOM. Regional Weather and Climate Guide. In the last 30 years in South East Queensland. Bureau of Meteorology. Available online: [www.bom.gov.au/climate/climate-guides/guides/044-South-East-QLD-Climate-Guide.pdf](http://www.bom.gov.au/climate/climate-guides/guides/044-South-East-QLD-Climate-Guide.pdf) (accessed on 5 July 2023).
26. ESRI. *ArcGIS Version 10.8 for Desktop*; Environmental Systems Research Institute Inc.: West Redlands, CA, USA, 2019.
27. Jawak, S.D.; Luis, A.J.; Fretwell, P.T.; Convey, P.; Durairajan, U.A. Semiautomated Detection and Mapping of Vegetation Distribution in the Antarctic Environment Using Spatial-Spectral Characteristics of WorldView-2 Imagery. *Remote Sens.* **2019**, *11*, 1909. [[CrossRef](#)]
28. Gitelson, A.; Merzlyak, M.N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *1*, 247–252. [[CrossRef](#)]
29. Barnes, E. 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.
30. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* **2002**, *1*, 416–426. [[CrossRef](#)]
31. Penuelas, J.; Baret, F.; Filella, I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica* **1995**, *1*, 221–230.
32. Robson, A.J.; Rahman, M.M.; Muir, J.; Saint, A.; Simpson, C.; Searle, C. Evaluating satellite remote sensing as a method for measuring yield variability in Avocado and Macadamia tree crops. *Adv. Anim. Biosci.* **2017**, *1*, 498–504. [[CrossRef](#)]
33. 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**, *1*, 289–298. [[CrossRef](#)]
34. Rouse, J., Jr.; Haas, R.; Schell, J.; Deering, D. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA Special Publication: Washington, DC, USA, 1974; Volume 351.
35. Gitelson, A.A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.* **2004**, *1*, 165–173. [[CrossRef](#)]
36. Bannari, A.; Asalhi, H.; Teillet, P.M. Transformed difference vegetation index (TDVI) for vegetation cover mapping. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002.
37. Pearson, R.L.; Miller, L.D. Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. *Remote Sens. Environ.* **1972**, *VIII*, 1355.
38. Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* **1969**, *1*, 663–666. [[CrossRef](#)]
39. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *1*, 295–309. [[CrossRef](#)]
40. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *1*, 195–213. [[CrossRef](#)]
41. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *1*, 271–282. [[CrossRef](#)]
42. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
43. Litvinenko, V.S.; Eckart, L.; Eckart, S.; Enke, M. A brief comparative study of the potentialities and limitations of machine-learning algorithms and statistical techniques. *E3S Web Conf.* **2021**, *266*, 02001. [[CrossRef](#)]
44. Brinkhoff, J.; Houborg, R.; Dunn, B.W. Rice ponding date detection in Australia using Sentinel-2 and Planet Fusion imagery. *Agric. Water Manag.* **2022**, *273*, 107907. [[CrossRef](#)]
45. Dong, X.; Zc, Z.; Yu, R.; Tian, Q.; Zhu, X. Extraction of Information about Individual Trees from High-Spatial-Resolution UAV-Acquired Images of an Orchard. *Remote Sens.* **2020**, *1*, 133. [[CrossRef](#)]
46. Perez, R.; Cebecauer, T.; Šúri, M. Chapter 2—Semi-Empirical Satellite Models. In *Solar Energy Forecasting and Resource Assessment*; Kleissl, J., Ed.; Academic Press: Boston, MA, USA, 2013; pp. 21–48.
47. Piñeiro, G.; Perelman, S.; Guerschman, J.P.; Paruelo, J.M. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Model.* **2008**, *1*, 316–322. [[CrossRef](#)]
48. Suarez, L.A.; Robson, A.; Brinkhoff, J. Early-Season forecasting of citrus block-yield using time series remote sensing and machine learning: A case study in Australian orchards. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103434. [[CrossRef](#)]
49. Zhu, Y. Rapid Target Detection of Fruit Trees Using UAV Imaging and Improved Light YOLOv4 Algorithm. *Remote Sens.* **2022**, *14*, 4324. [[CrossRef](#)]
50. Zheng, C.; Liu, T.; Abd-Elrahman, A.; Whitaker, V.M.; Wilkinson, B. Object-Detection from Multi-View remote sensing Images: A case study of fruit and flower detection and counting on a central Florida strawberry farm. *Int. J. Appl. Earth Obs. Geoinformation.* **2023**, *123*, 103457. [[CrossRef](#)]
51. Donovan, J. Australian Mango Varieties. Available online: <https://lawn.com.au/australian-mango-varieties/> (accessed on 29 March 2023).
52. Payne, A.B.; Walsh, K.B.; Subedi, P.P.; Jarvis, D. Estimation of mango crop yield using image analysis—Segmentation method. *Comput. Electron. Agric.* **2013**, *1*, 57–64. [[CrossRef](#)]
53. Smith, J.; Smith, P. *Environmental Modelling: An Introduction*; Oxford University Press: Oxford, UK, 2007.

54. Anderson, N.T.; Walsh, K.B.; Wulfsohn, D. Technologies for Forecasting Tree Fruit Load and Harvest Timing—From Ground, Sky and Time. *Agronomy* **2021**, *11*, 1409. [[CrossRef](#)]
55. Brinkhoff, J.; Robson, A.J. Block-level macadamia yield forecasting using spatio-temporal datasets. *Agric. For. Meteorol.* **2021**, *303*, 108369. [[CrossRef](#)]
56. Filippi, P.; Whelan, B.M.; Vervoort, R.W.; Bishop, T.F.A. Mid-season empirical cotton yield forecasts at fine resolutions using large yield mapping datasets and diverse spatial covariates. *Agric. Syst.* **2020**, *184*, 102894. [[CrossRef](#)]
57. Deines, J.M.; Patel, R.; Liang, S.-Z.; Dado, W.; Lobell, D.B. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sens. Environ.* **2021**, *1*, 112174. [[CrossRef](#)]
58. Bana, J.; Kumar, S.; Sharma, H. Diversity and nature of damage of mango insect-pests in south Gujarat ecosystem. *J. Entomol. Zool. Stud.* **2018**, *1*, 274–278.
59. Kumar, K.; Adak, T.; Singha, A.; Shukla, S.; Singh, V. Appraisal of soil fertility, leaf nutrient concentration and yield of mango (*Mangifera indica* L.) at Malihabad region, Uttar Pradesh. *Curr. Adv. Agric. Sci. (Int. J.)* **2012**, *1*, 13–19.
60. van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [[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.