



*Research article*

## **Peanut yield prediction with UAV multispectral imagery using a cooperative machine learning approach**

**Tej Bahadur Shahi<sup>1,\*</sup>, Cheng-Yuan Xu<sup>2</sup>, Arjun Neupane<sup>1</sup>, Dayle B. Fleischfresser<sup>3</sup>, Daniel J. O'Connor<sup>4</sup>, Graeme C. Wright<sup>4</sup> and William Guo<sup>1,\*</sup>**

<sup>1</sup> School of Engineering and Technology, Central Queensland University, Rockhampton, Australia

<sup>2</sup> School of Health, Medical, and Applied Sciences, Central Queensland University, Bundaberg, Australia

<sup>3</sup> Queensland Department of Agriculture and Fisheries, Townsville, Australia

<sup>4</sup> Peanut Company of Australia, Kingaroy, Australia

\* **Correspondence:** Email: [t.shahi@cqu.edu.au](mailto:t.shahi@cqu.edu.au), [w.guo@cqu.edu.au](mailto:w.guo@cqu.edu.au).

**Abstract:** The unmanned aerial vehicle (UAV), as a remote sensing platform, has attracted many researchers in precision agriculture because of its operational flexibility and capability of producing high spatial and temporal resolution images of agricultural fields. This study proposed machine learning (ML) models and their ensembles for peanut yield prediction using UAV multispectral data. We utilized five bands (red, green, blue, near-infra-red (NIR) and red-edge) multispectral images acquired at various growth stages of peanuts using UAV. The correlation between spectral bands and yield was analyzed for each growth stage, which showed that the maturity stages had a significant correlation between peanut yield and spectral bands: red, green, NIR and red edge (REDE). Using these four bands spectral data, we assessed the potential for peanut yield prediction using multiple linear regression and seven non-linear ML models whose hyperparameters were optimized using simulated annealing (SA). The best three ML models, random forest (RF), support vector machine (SVM) and XGBoost, were then selected to construct a cooperative yield prediction framework with both the best ML model and the ensemble scheme from the best three as comparable recommendations to the farmers.

**Keywords:** unmanned aerial vehicle; multispectral imagery; machine learning; simulated annealing; peanut yield prediction; random forest; support vector machine; XGBoost

---

## 1. Introduction

Maximizing crop yield by keeping the cost as low as possible is one of the main goals of many precision agriculture systems. Early identification and prediction of crop traits such as crop disease, biomass and yield are beneficial as they allow the farmer to manage crop growth and harvesting well in advance [1]. Therefore, the estimation of yield and related parameters such as biomass, disease, plant health, nitrogen status and soil conditions has been a frequent topic in the literature [2–8]. Early detection and management of problems associated with farming can help increase yield and subsequent profit, and better estimation of the yield offers farmers and processors numerous benefits in terms of harvest planning, storage and transportation scheduling, sale and price negotiation and other business decisions.

The traditional yield prediction models are based on ground samples, collected from the farm, and extrapolating these samples throughout the field to estimate the yield [9]. These methods are not only costly and labour-intensive but also poorly represent the spatial variability of yield over the field. An alternative approach is a non-destructive sampling method for yield estimation which uses a remote sensing platform to acquire field images and employs various vegetation indices (VIs) to establish a regression model for crop yield [10]. Recent works on UAV-based remote sensing [11–14] showed the efficiency of crop traits such as yield estimation using multispectral images and ML methods [15]. For instance, Guo et al. [14] utilized the multispectral images of maize with a Mini-MCA camera embedded in the drone to estimate the soil and plant analyzer development (SPAD) values. They also implemented various ML methods such as SVM and RF where SVM outperformed the RF with an  $R^2$  of 0.81 in estimating SPAD value.

For crop yield estimation using UAVs, the VIs derived from the multispectral and RGB images were extensively utilized by various works [12,16,17]. These studies established a strong correlation between crop yield and VIs. For instance, the normalized difference vegetation index (NDVI) is linearly related to wheat yield [17]. Similarly, a yield map for rice and wheat crops was developed using NDVI from multispectral images [12]. Since the NDVI has a saturation issue with high biomass at the early growing stages of the crop, a few other VIs such as enhanced vegetation index (EVI) and soil-adjusted vegetation index (SAVI) were also assessed for yield estimation [16].

Since UAV has the flexibility in revisiting the field and can capture high-resolution imagery in comparison to satellite imagery, it has opened possible avenues for cheaper and more frequent image acquisition to support more accurate estimates of crop traits using predictive approaches such as ML methods [18,19]. For instance, Zhou et al. [18] implemented a convolutional neural network (CNN) for soybean yield estimation with high-resolution UAV imagery. They used crop features such as plant height, canopy colour and canopy texture to train the neural network. Their model achieved an  $R^2$  of 0.78 with a root mean square of 391.0 kg/ha. Similarly, Guo et al. [19] implemented four ML models, a backpropagation neural network (BPNN), SVM, RF and extreme learning machine (ELM), for maize yield predictions using VIs. They showed that SVM with a modified red-blue vegetation index (MRBVI) was effective in monitoring maize yield. Besides the image feature, Guo et al. [20] employed the combination of phenology, climate and geography data to estimate rice yield with statistical and ML methods. However, their proposal of building the yield prediction model with an individual ML method missed the cooperative nature of the ensemble approach where if one method fails to capture the correct prediction, another ML method can pick the right prediction. Considering such limitations, this study first establishes the relationship between UAV images and peanut yield at the individual

growth stage. Based on such a relationship and existing ML methods, an accurate and cooperative ensemble method for yield prediction is proposed and validated using peanuts as a study crop.

Peanut is an oilseed crop grown in many countries over the world. In Australia, the peanut is mainly grown in Queensland, in the northeast of Australia. Its growth cycles include various stages: planting, emergence, emergence to first flower (FF), flowering (F), pegging, pod-filling and harvest maturity (HM). It takes around three to five months from planting to maturity [21]. It is important to monitor peanut growth to assure the quality and quantity potential of peanuts. Owing to such successes of UAV-based remote sensing for crop yield estimation, this study aims to develop peanut yield estimation models based on UAV multispectral images at the late growth stages in Queensland. This study intends to

- a) investigate the relationship between spectral information acquired with UAV and peanut yield at different peanut growth stages.
- b) evaluate multiple linear regression and seven existing ML (non-linear) models for yield prediction using SA-based hyperparameter optimization.
- c) select the best learning models and design an ensemble approach for better yield prediction.
- d) compare the performance of the best ML model and the ensemble approach for yield prediction.

The paper is organized as follows. Related works are reviewed in Section 2. The study site, data collections, experimental design and methodologies are presented in Section 3. Experimental results and discussion are reported in Section 4. Finally, conclusions and future works are summarized in Section 5.

## 2. Related works

Remote sensing has been widely used for crop yield prediction because of its ability to cover large geographical areas from the country level to the continent level [22]. Forecasting with remote sensing tries to build a prediction model in a non-destructive way by capturing field data with sensors [6]. Recent works with UAVs [23,24] showed that it has a great potential to be used in precision agriculture because of their flexibility in flying, ability to capture high-resolution imagery and low cost compared to other airborne imagery such as satellite [25]. However, these features vary in designs and sensors used for imaging on different UAVs [26]. Sensors on UAVs play a vital role in data acquisition. Several types of sensors have been used with UAVs for crop monitoring. These include RGB sensors, multispectral sensors, hyperspectral sensors and thermal sensors [27]. A few studies have proved the effectiveness of using UAV images in yield prediction. For instance, Ramos et al. [28] showed that NDVI, normalized difference NDRE and green normalized difference (GNDVI) were highly ranked indices for maize yield prediction using multispectral UAV imagery.

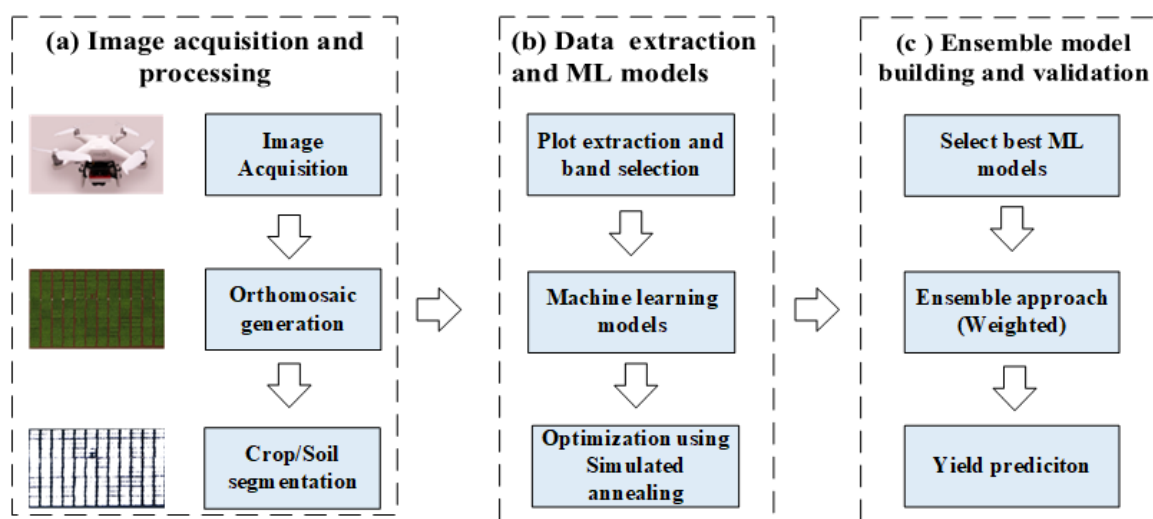
Studies on yield estimation using UAV-based sensors have increased in recent years. Zhou et al. [24] estimated grain yield using RGB as well as multispectral sensors. They investigated six RGB indices and seven multispectral indices at multiple growth stages of rice for yield estimation. Five regression models based on linear, exponential, logarithmic, polynomial and power functions were established. Their results showed that rice yield was best estimated at the booting stage with NDVI and visible atmospherically resistant index (VARI). However, this study did not explore the ML model for yield predictions. Corn grain yield estimation was proposed in [23] using VIs, canopy cover and plant density

acquired through multispectral as well as RGB sensors. Six VIs were examined for grain yield prediction with an RMSE of 0.125 t/ha and a correlation coefficient of 0.99. Similarly, Geipel et al. [29] combined the spectral and spatial indices with a linear regression model for corn yield estimation and achieved an  $R^2$  of 0.74. An artificial neural network (ANN) was implemented by Ashapure et al. [30] for tomato yield estimation using a combination of plant attributes, VI and weather information which achieved an  $R^2$  of 0.70. The various ML models including LR, RF SVM and GPR were implemented by Matese and Di Gennaro [31] for vine yield estimation where the GPR achieved the highest  $R^2$  of 0.80.

A regional regression model for crop yield prediction with UAV multispectral data was implemented by Bian et al. [32]. They explored six ML methods such as SVM, RF and Gaussian process regression (GPR) and showed that GPR achieved the optimal prediction of wheat yield with  $R^2 = 0.87$  at the filling stage. Similarly, multi-sensor data fusion and ML methods for wheat yield prediction were implemented by Fei et al. [33]. They developed regression models using ML algorithms such as SVM, deep neural networks (DNN), ridge regression, RF and cubist. They achieved the highest  $R^2$  values up to 0.69 when data from multiple sensors such as RGB, multispectral and thermal were combined and ensemble learning was implemented.

### 3. Materials and methods

The high-level setup required to carry out this work is depicted in Figure 1. First, the raw UAV images were captured and processed by following the standard UAV image processing pipeline [27,34]. Second, the pre-processed images were divided into a region of interest, and plot-level data extraction was carried out. Third, the highly correlated spectral bands with peanut yield are selected and fed into both linear and non-linear ML models. Furthermore, these ML models' hyperparameters were optimized using SA. Finally, the best-performing ML models were selected to build a cooperative ensemble-based yield prediction framework. A detailed discussion of each activity is provided in the following sections.



**Figure 1.** The high-level model building pipeline used in this work.

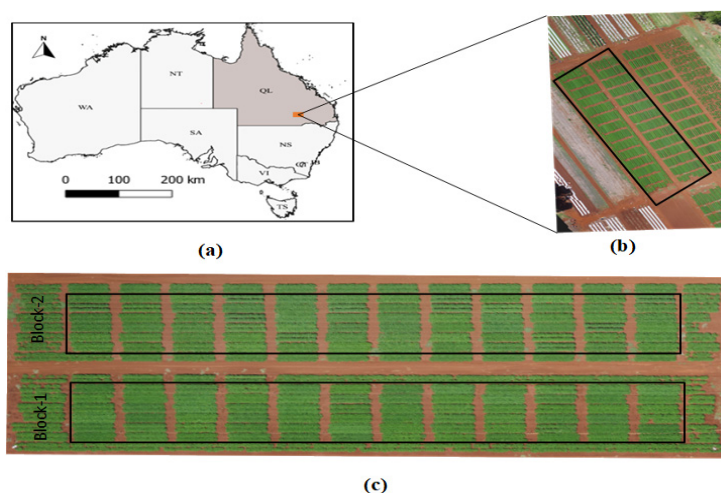
### 3.1. Crop study area and yield data

Field data were collected from the Queensland Department of Agriculture and Fisheries research facility at Bundaberg in S. Queensland, Australia. The regional climate is categorized as sub-tropical with an annual average temperature of 27.8 °C and average precipitation of 742.8 mm for 2018 (<http://www.bom.gov.au/climate/data/>). This study considered two field trials and each trial had 24 treatments/genotypes x 3 replications (72 plots) where each treatment has 2 rows x 5m. Therefore, there were 144 plots in total. Before planting the peanut, the soil sample was collected and sent for analysis. Then, the Gypsum @1.5 t/ha and potassium sulphate @70kg/ha was applied on 2017-11-20 to make the field ready for peanut plantation. The peanuts were planted on 2017-12-19 with inter-row cultivation and no herbicide treatments except some hand weeding and chipping. The soil type was red ferrosol as per the Australian soil classification. The 50mm of irrigation was applied three times dated 2018-01-23, 2018-02-26 and 2018-04-20. Similarly, the fungicide (Bravo @1.8L/ha) and fungicide (Amistar xtra 750ml/ha+ agral 100ml/100l) were applied four and two times respectively throughout the growth periods. No insecticide treatments were applied. The peanut trails were harvested on 2018-06-04 and threshed on 2018-06-19.

A destructive sampling method was used to collect peanut yield data for each plot. For this, a sample from a non-plot area was used to determine when the peanuts have reached full maturity. Once the peanuts reached full maturity, they were dug out with a mechanical digger and the bushes/peanuts were left to dry on the ground for 7–10 days. This allowed the bush and peanuts to dry which helped with the separation process during threshing. In the threshing process, the pods were removed from the bush with the peanuts going into a hessian bag and then labelled. Once the trials were harvested, kernel moisture was determined. If the peanuts were too high in moisture they were put onto bed dryers until they reached safe kernel moisture of around 9–10%. Finally, the extraneous material was removed through a pre-cleaner and each sample was weighed to determine the final yield which was expressed as tons per hectare (t/ha).

### 3.2. UAV image acquisition

The multi-rotor drone, Phantom 3 (DJI, Shenzhen, China) was used to collect the peanut field images. It consists of an integrated MicaSense RedEdge (Mica-sense, Seattle, WA, USA) with five spectral bands: Red (630–690 nm), Blue (460–510 nm), Green (545–575 nm), Near-infrared (820–860 nm) and Red-edge (712–722 nm). The images were captured at the various growth stages of peanuts at the height of 40 meters above the ground along with a parallel camera CCD angle to the ground. The side and forward overlaps of 60% and 90% were maintained in each UAV flight while capturing the images. The geo-referencing was carried out in the World Geodetic System (WGS) 1984 datum, Universal Transverse Mercator (UTM) Zone 55 projection. For this, six ground checkpoints were surveyed and marked with Real-time Kinematic (RTK), Global Positioning System (GPS) and ground data were registered with multispectral images which provide a spatial error of less than 2 cm across the field of study (Figure 2). The five growth stages of peanuts were mapped with the UAV flights listed in Table 1.



**Figure 2.** Study area maps (a) Field location on the Australia map. (b) RGB Orthomosaic of whole UAV trial (c) two peanut trials/blocks (block1 and block2) used in this work.

**Table 1.** UAV Images acquisition of peanut fields\*.

Image acquisition date	Growth stages	Days after planting (DAP)
25/01/2018	FF	37
12/02/2018	F	55
13/03/2018	Pegging	84
25/04/2018	Pod filling (PF)	127
29/05/2018	HM	161

\*Note the planting date for these trials was 19/12/2017.

### 3.3. Plot-level data extraction

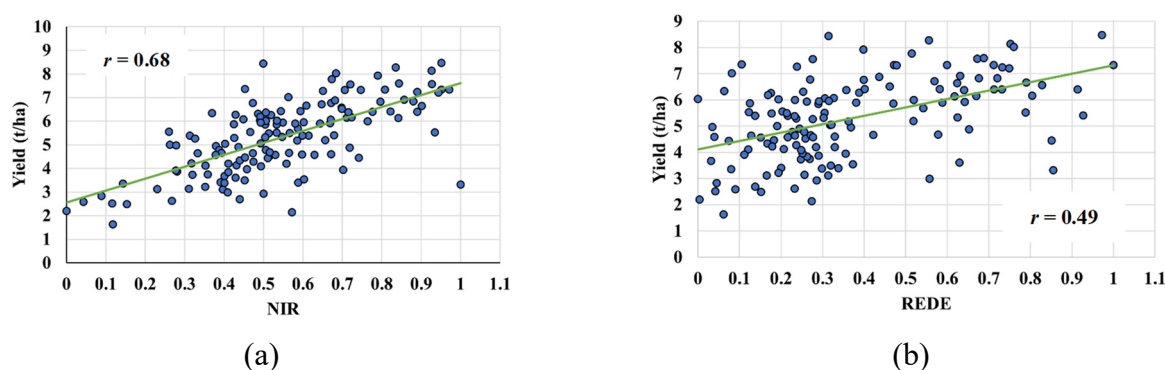
We followed the UAV image processing pipeline as outlined in [27] to extract the plot-level image data. We first transferred the raw UAV image into a computing platform to perform the image stitching using Pix4Dmapper (Pix4D S. A. Prilly, Switzerland) with a specific template “Ag Multispectral” included in the software package to rectify and mosaic the UAV images. Once the orthomosaic of the study area was achieved, the individual orthomosaic for each spectral band was stacked into a virtual raster using quantum geographic information system (QGIS) software [35]. Then, an individual plot shapefile for each block (block1 and block2) was built using an open-source R package-FIELDimageR [36] which divides a whole field into individual plots. Finally, the individual plot-level data extraction was carried out by clipping the individual plot using the given shape file and the average of all pixels included in each plot is considered plot-level spectral information. Furthermore, the soil pixels were segmented from crop pixels using a Green Red Vegetation Index (GRVI) as defined in Eq (1).

$$\text{GRVI} = \frac{(G-R)}{(G+R)} \quad (1)$$

where if  $\text{GRVI} \leq 0.2$ , a pixel was masked out as a soil pixel; otherwise, the pixel was considered as a crop pixel.

### 3.4. Correlation analysis and band selections

The correlation results between the individual spectral band and peanut yield reported in Table 2 show that the first four growth stages (FF, F, P and PF) have a very low correlation with yield. Hence, we filtered out these growth stages from further consideration and choose the HM stage for yield prediction. Considering the individual spectral band correlation at the HM stage, the NIR ( $r = 0.68$ ) and REDE ( $r = 0.49$ ) bands have a higher correlation (greater than 0.40) in comparison to the other three bands. However, the other two bands Red (R) and green (G) have a highly significant correlation ( $r > 0.27$ ) with yield, and the blue (B) band has a poor correlation. For instance, the correlation plots showing a positive relationship between yield and NIR ( $r = 0.68$ ) and REDE ( $r = 0.49$ ) are shown in Figure 3. Hence, the four spectral bands (R, G, NIR and REDE,) at the HM stage were selected to develop the peanut yield prediction model using ML as well as ensemble models.



**Figure 3.** The correlation plot between yield and spectral bands. a) NIR and b) REDE at the HM stage of peanuts. Note that the ‘ $r$ ’ denotes the correlation coefficient between yield and the respective spectral bands.

**Table 2:** Correlation coefficient ( $r$ ) between spectral bands and peanut yield. Note that \* and \*\* represents the significance level of 0.01 and 0.05 for the correlation value respectively.

Growth stage/DAP	R	G	B	NIR	REDE
FF	-0.15	-0.18**	-0.16**	-0.10	-0.12
F	-0.05	-0.01	0.16**	0.27*	0.11
P	-0.16**	-0.05	0.02	0.28*	-0.14
PF	-0.25*	-0.16**	-0.17**	0.31*	0.01
HM	<b>0.27*</b>	<b>0.35*</b>	0.15	<b>0.68*</b>	<b>0.49*</b>

### 3.5. Multiple linear regression

Multiple linear regression represents the linear relationship between a set of several independent variables and a dependent variable. It estimates the regression model by minimizing the sum of squared errors between the dependent variable and prediction by linear approximations. Here, we used individual spectral bands as independent variables and peanut yield as a dependent variable to build the MLR model. If  $X_1$ ,  $X_2$ ,  $X_3$  and  $X_4$  represent the four spectral bands (R, G, NIR and REDE) as independent variables and  $Y$  represent the dependent variable (yield), the multiple regression model

for peanut yield estimation is defined in Eq (2).

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + c \quad (2)$$

where  $a_1, a_2, a_3, a_4$  represent the regression coefficient and  $c$  represents the constant.

### 3.6. ML models

We consider seven existing ML models for yield prediction. These models range from support vector regressor (SVR) to multilayer perceptron neural network (MLP). Here we briefly summarized these models.

#### **Support vector regressor**

The SVM is a binary classifier based on hyperplane to separate multidimensional data into two classes [37]. However, it can be used to resolve the regression problem using a margin of tolerance known as a SVR. It consists of two free parameters as regularization parameter ( $C$ ) and epsilon which need to be optimized.

#### **Decision tree**

A decision tree (DT) is a non-parametric learning method which creates a set of decision rules to predict the target variable using certain criteria such as the Gini index or entropy [38]. The decision tree's hyperparameter such as the maximum depth of the tree, minimum samples to split an internal node, and minimum sample required to be at the leaf need to be optimized for a given dataset.

#### **Random forest**

RF uses the decision tree as a basic regressor with bagging approaches [39]. It built a forest of decision trees with random subsets of training data with the replacement of samples. Finally, the output of all trees is averaged to get the final prediction for a given sample [39]. The random forest's hyperparameters that need to be optimized include a number of estimators, the maximum depth of the tree, minimum samples to split an internal node, the minimum sample required to be at the leaf, etc. [40].

#### **Extra tree classifier**

Extra Tree regressor (ETR) is also a meta estimator that uses the randomized decision trees on the random subsets of training data similar to a RF. However, it is different from RF regressors in the way that trees are constructed. In ETR, further randomness is introduced while constructing the splitting rule. Here, the thresholds for the splitting rule are drawn at random for each candidate feature and the best threshold among these randomly generated thresholds is chosen [40]. It has a similar set of hyperparameters as of RF to be optimized.

#### **AdaBoost**

It is also a meta-estimator based on the adaptive boosting method of ensemble learning, which fits a sequence of weak learning trees such as small decision trees on a modified version of the dataset. A strong learner is obtained by combining all such weak learners using a weighted majority voting in each boosting iteration [41]. The data modification at each boosting iteration consists of applying weights to each of the training samples.

#### **XGBoost**

XGBoost uses a boosting approach for ensemble learning. The combination of a group of weak learners can be performed either by boosting or bagging. XGBoost uses three kinds of boosting: gradient boosting, regularized boosting and stochastic boosting, which surge the overall performance of XGBoost [42].

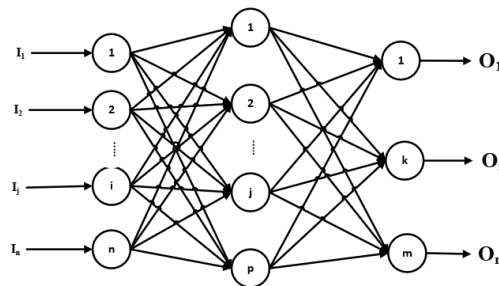


### Multilayer perceptron neural network

A multilayer perceptron (MLP) neural consists of one input layer, one-hidden layer and one output layer (Figure 4). The n-dimensional vector as input to a one-hidden layer (1-h) neural network will be transformed into an m-dimensional output vector using Eq (3).

$$o_m = f(\sum_{j=1}^m B_{kj} g(\sum_{i=1}^n A_{ji} I_i) \quad (k = 1, 2, 3 \dots \dots m) \quad (3)$$

where  $f$  and  $g$  are the activation functions;  $A_{ji}$  represents input-hidden layer weights at the neuron  $j$  and  $B_{kj}$  is the hidden-output layer weights at output unit  $k$ .



**Figure 4.** Multilayer perceptron (MLP) neural network with one-hidden (1-h) layer.

### 3.7. SA for hyper-parameter optimization of ML models

SA is based on the analogy of heating a material and cooling it down slowly to achieve the desired structure. Similarly, it can be used to find the optimal or approximate solution during the iterative process over the search space [43]. The SA iteratively tries to find the best solution with the following steps: initialization, neighbour selection, evaluation and accept/reject the solution. Here, we employed SA to find the optimal set of hyperparameters for each ML model discussed in section 3.6. In each ML model, there are two major stages: model training and evaluation. Model training involves the finding of a set of rules or functions resulting from a given ML method using training data. The training data consists of pair of dependent and independent variables. While training the model, it tries to minimize the objective function. In this work, we utilized the mean square error (MSE) as an objective function (refer to Eq (4)). Once the model is trained, it is evaluated on the validation data. The SA is used to select the best set of hyperparameters while evaluating the ML model as shown in Algorithm-1. Here, we first set the initial temperature and the initial set of hyperparameters randomly. This set of hyperparameters is considered as an initial solution (for instance, NeighborSelection(s) in Algorithm 1, Table 3). Then the ML model is trained with those hyper-parameters and evaluated using Evaluation (V) operator on the validation set. The SA iteratively finds the optimal set of hyper-parameters for a given ML model using steps 3 to 11 in Algorithm 1 (Table 3).

$$MSE = \frac{1}{N} (\sum_{i=1}^N (y_i - \hat{y}_i))^2 \quad (4)$$

where  $y_i$  and  $\hat{y}_i$  are actual and simulated values.

**Table 3.** The high-level pseudocode of the SA algorithm [44].**Algorithm 1** Simulated Annealing

---

```

1: Set the initial temperature  $T \leftarrow T_0$ 
2: Set the initial solution  $S \leftarrow S_0$ 
3: while stopping criterion is not met do
4:    $V = \text{NeighborSelection}(S)$ 
5:    $F = \text{Evaluation}(V)$ 
6:   if  $F$  satisfies the probabilistic acceptance criterion then
7:      $S = V$ 
8:   end if
9:   Update  $T$  according with the annealing schedule
10: end while
11: return  $S$ 

```

---

## 4. Experimental results and discussion

### 4.1. Data compilation and parameter settings

Among the 144 data samples (72 plots in each block), we performed a sanitized check to find any noise or outliers. For this, we used Mahalanobis distance [45] and found four data points as noise or outliers which we removed from the dataset. This is essential as some ML models may not learn the appropriate patterns with the noise present in the dataset. After such noise removal, we ended up with a total of 140 data samples for the training and testing of yield prediction models. The yield distribution among the experimental plots was in the range of 1–9 tons/ha. While dividing the data samples into train and test sets, it is crucial to maintain a similar distribution in both train and test sets. Otherwise, the model evaluation might be biased towards the specific range of yield. The common approach to split the training and test data for model training is to split the dataset into train and test set in the ratio of 9:1 with random sampling which doesn't fit in our case as we need to preserve the yield range of 1–9 t/ha in both training and test set. Therefore, we used a specific technique to split the data into train and test sets which first group the sample data into nine groups on the basis of yield range (yield in the range of [1–2), [2–3), and so on, where '[' denotes the inclusion and ')' represents the exclusion). Then, we performed the stratified sampling to select the train and test set from these groups in the ratio of 9:1. Similarly, 10% of the training set was sampled as validation data while training ML models. The holdout set of 14 data samples was used to evaluate the performance of ML as well as ensemble approaches.

All the ML methods were implemented using a scikit-learn [40] package while the SA and the proposed ensemble approaches are implemented in Python. The empirical simulations were carried out on a PC with an Intel i5-8265 CPU (1.6GHz, 8 cores) with 16 GB of memory running Windows 10. The list of hyperparameters that are optimized using SA for each ML model and their optimal values are listed in Table 4.

**Table 4.** The list of ML and their associated hyperparameters used in this work.

ML method	List of hyperparameters and their optimal value
Decision Tree	Max_depth = 87, min_samples_split = 0.18, max_leaf_nodes = 4, min_samples_leaf = 0.1, splitter = random
SVR	Gamma = 0.001, C = 1000, kernel = linear, epsilon = 0.1
MLR	N/A
RF	n_estimators = 7, max_features = sqrt, max_depth = 346, min_sample_split = 8, min_sample_leaf = 4, bootstrap = True
ETC	n_estimators = 30, max_depth = 345, min_sample_split = 0.68, min_sample_leaf = 2
XG-boost	max_depth = 25, subsamples = 0.7, colsample_bytree = 0.4, learning_rate = 0.1, gamma = 0.0, scale_pos_weight = 10, n_estimators = 85
AdaBoost	n_estimators = 429, learning_rate = 3.07, loss = linear
MLP	hidden_layer_size = 92, activation = relu, solver = lbfgs, learning_rate_init = 0.025

Note: 'N/A' denote that the corresponding methods don't include any hyperparameters. (The details about the corresponding hyperparameters of each ML method can be found in [40] and [42])

#### 4.2. Evaluation metrics

The predicted yields from various ML models as well as ensemble approaches are assessed with well-known evaluation metrics such as coefficient of determination ( $R^2$ ), root mean square error (RMSE) and mean absolute error (MAE) [46]. The mean absolute relative errors (MARE) in percentage for each test sample from best-performing models are also reported.

#### 4.3. Simulation results

The optimized ML models with SA are evaluated on 14 test datasets. We performed five runs of each model to report the average results so as to reduce the randomness of taking the result from only a single run (Table 5). Through all indicators, XGBoost has the highest  $R^2$  of 86.43% and minimum errors of RMSE = 0.5598 and MAE = 0.4131. The second ML model is SVR with  $R^2$  of 82.69% and errors of RMSE = 0.6376 and MAE = 0.5453, followed by RF with  $R^2$  of 81.26% and errors of RMSE = 0.6691 and MAE = 0.5347. The other ML models such as DT and MLP have produced the  $R^2$  in the range of 71%-78%. Hence, we would have a well-covered prediction range for peanut yield by only choosing the best three ML models, i.e., XGBoost, SVR and RF, for yield forecasting.

**Table 5.** Performance of eight ML models for yield prediction on the test data.

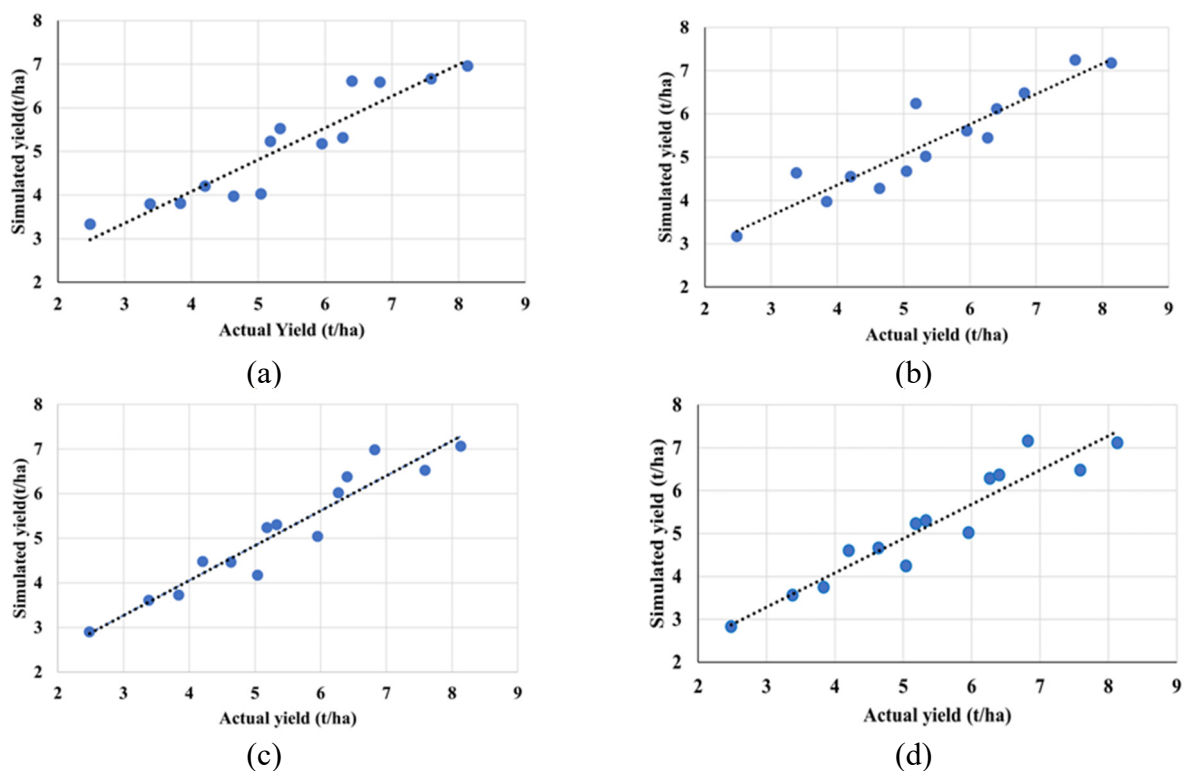
ML Methods	RMSE	MAE	$R^2$ (%)
Decision Tree (DT)	0.8271	0.6580	71.46
<b>SVR</b>	<b>0.6376</b>	<b>0.5453</b>	<b>82.69</b>
MLR	0.7889	0.6775	72.87
<b>RF (RF)</b>	<b>0.6691</b>	<b>0.5347</b>	<b>81.26</b>
ETC	0.7727	0.6638	75.23
<b>XGBoost</b>	<b>0.5598</b>	<b>0.4131</b>	<b>86.43</b>
AdaBoost	0.8014	0.6858	73.21
MLP	0.68755	0.5476	78.09

Note: The reported metrics are taken as an average of five runs of each model.

The prediction results of these three ML models on the test datasets are listed in Table 6, along with the prediction error (the difference between the actual and predicted yields) on each of the 14 test samples. The scatterplots for the three best-performing models are presented in Figure 5. With these individual scatterplots, the divergence between the predicted and actual yield for some points is higher in the case of SVR compared to RF and XGBoost than the ensemble model (Figure 5 (d)).

**Table 6.** The actual yield and predicted yields from the three best ML models and errors.

Test set	Actual yield (t/ha)	Predicted yield (t/ha)			Error (t/ha)		
		RF	SVR	XGBoost	RF	SVR	XGBoost
1	2.4840	3.3264	3.1651	2.8984	-0.8424	-0.6811	-0.4144
2	3.3850	3.7945	4.6319	3.6097	-0.4095	-1.2469	-0.2247
3	3.8390	3.8042	3.9722	3.7263	0.0348	-0.1332	0.1127
4	4.2040	4.1977	4.5490	4.4774	0.0063	-0.345	-0.2734
5	4.6370	3.9719	4.2712	4.4621	0.6651	0.3658	0.1749
6	5.0430	4.0220	4.6703	4.1725	1.021	0.3727	0.8705
7	5.1855	5.2248	6.2340	5.2332	-0.0393	-1.0485	-0.0477
8	5.3333	5.5192	5.0163	5.2981	-0.1859	0.317	0.0352
9	5.9574	5.1734	5.6104	5.0364	0.784	0.347	0.921
10	6.2711	5.3159	5.4402	6.0164	0.9552	0.8309	0.2547
11	6.4050	6.6049	6.1111	6.3740	-0.1999	0.2939	0.031
12	6.8235	6.5846	6.4781	6.9806	0.2389	0.3454	-0.1571
13	7.5905	6.6666	7.2463	6.5162	0.9239	0.3442	1.0743
14	8.1340	6.9546	7.1702	7.0573	1.1794	0.9638	1.0767



**Figure 5.** The scatter plot of simulated yield vs predicted yield with the best three ML models a) RF, b) SVR, c) XGBoost and d) ensemble.

#### 4.4. Ensemble approaches for peanut yield prediction

As the differences between the actual and predicted yields vary with the quantity of actual yield, it is difficult to observe any general trend from these sets of differences. However, more meaningful observations can be drawn from the absolute relative errors (the percental ratio of the absolute difference versus the actual yield) on each test sample shown in Table 7.

The relative errors ( $e_r$ ) from the 14 test sets over the three best prediction models demonstrate the following features. First, over the average of the 14 test sets, XGBoost returned the best performance for yield prediction with an average relative error of 7.5% from the best of less than 0.5% to the worst of 17.3% (Table 7). If we regard a relative error of 20% and above as a fail in the prediction of peanut yield, a relative error greater than or equal to 15% but below 20% as a poor prediction, a relative error greater than or equal to 10% but below 15% as a moderately accurate prediction, a relative error smaller than 10% as a highly accurate prediction, XGBoost would be classified as the most successful method for peanut yield prediction without a single failure (Table 8), the only method among the three best models. Second, both the RF and SVR models have a similar average relative error of around 11%, from less than 0.2% to about 34% for RF and from 3.5 to 37% for SVR. Third, although XGBoost seems the most consistent and successful method among the three models on average, it was not always the best predictor among the three for individual cases. For example, the best predictor for test sets 4, 6 and 11 are RF, SVR and XGBoost, respectively. Furthermore, in 9 out of 14 tests, all three models consistently over-estimated or under-estimated yield but on the other 5 occasions, the three models produced forecasts mixed with over-estimated and under-estimated yields. Hence, in addition to picking up the most consistent and successful performer, another combinative predictor constructed from all the three best models would offer another comparative means in peanut yield forecasting.

**Table 7.** The absolute relative errors in percentage from the three best ML models and the average.

Test set	Actual yield (t/ha)	Absolute relative errors (%)		
		RF	SVR	XGBoost
1	2.4840	33.91	27.42	16.68
2	3.3850	12.10	36.84	6.64
3	3.8390	0.91	3.47	2.94
4	4.2040	0.15	8.21	6.50
5	4.6370	14.34	7.89	3.77
6	5.0430	20.25	7.39	17.26
7	5.1855	0.76	20.22	0.92
8	5.3333	3.49	5.94	0.66
9	5.9574	13.16	5.82	15.46
10	6.2711	15.23	13.25	4.06
11	6.4050	3.12	4.59	0.48
12	6.8235	3.50	5.06	2.30
13	7.5905	12.17	4.53	14.15
14	8.1340	14.50	11.85	13.24
Average		10.54	11.61	7.51

**Table 8.** The performance summary of the best three ML models by relative errors ( $e_r$ ).

Method	$e_r \geq 20\%$ (Fail)	$15\% \leq e_r < 20\%$ (Low accuracy)	$10\% \leq e_r < 15\%$ (Moderate accuracy)	$e_r < 10\%$ (High accuracy)	Percentage of fail (Out of 14)*
RF	2	1	5	6	14.28
SVR	3	0	2	9	21.42
XGBoost	0	3	2	9	0

\* The percental ratio of the number of fails versus the total number (14).

To build a weighted ensemble from the three best ML models, we use their average relative errors and  $R^2$  values to render the weight factors for the three ML models respectively. If the model has an average relative error within the high accuracy class, farmers would be happy to assign a credit of 100 to that model, for instance, XGBoost in Table 7. Similarly, the fail class should be credited with 0 and considered as completely useless. The model that fell in the less credible class of low accuracy would be given the lowest credit of 1. Along the similar line, it would be reasonable to assign a credit of 10 to the model that fell in the moderate accuracy class, like RF and SVR in this study. Using this credit scheme, we can work out the total credit from these three ML models as per Eq (9).

$$Total\ credit = 10 (RF) + 10 (SVR) + 100 (XGB) = 120 \quad (9)$$

Using these individual and total credit scores, a weighted ensemble for the predicted peanut yield can be determined using Eq (10).

$$y_p = \frac{10y_{RF} + 10y_{SVR} + 100y_{XGB}}{120} = \frac{y_{RF} + y_{SVR} + 10y_{XGB}}{12} \quad (10)$$

#### 4.5. Discussion on peanut yield prediction with the ensemble

Using both the XGBoost and ensemble models together, a comparable and relatively consistent forecast on peanut yield could be recommended to the farmers with an accuracy ranging from low to high in all cases (Table 9). The XGBoost model could produce a predicted yield with a high accuracy at a possibility of 64% (9/14), a moderate accuracy at a possibility of 14% (2/14), or a low accuracy at a possibility of 21% (3/14). The ensemble scheme could produce a predicted yield with a high accuracy at the same possibility (9/14) as XGBoost, a moderate accuracy at a possibility of 21% (3/14), or a low accuracy at a possibility of 14% (2/14). The difference between these two appears minor because the ensemble output is largely determined by the output from XGBoost (83%). However, the ensemble provides an alternative that may be able to complement the output from XGBoost in special cases should such cases be encountered. For example, in Case 9 shown in Table 7, both RF and SVR performed better than XGBoost, which resulted in a better yield prediction by the ensemble than the XGBoost alone (Table 9). More noticeably, in most cases, the ensemble model looks a more improved predictor than either RF or SVR alone for peanut yield prediction.

It should be noted that the actual yields used in this study were taken as given. We are unsure if significant errors would exist in some of these records due to various possibilities. Logically, the large errors in the predicted yields associated with both the smallest (Cases 1 and 2) and largest (Cases 13 and 14) actual yields could be explained as a result of a scarcity of data at the two ends of a data sequence which may greatly influence the training of ML models. However, for a model like

XGBoost that performed consistently satisfactorily for most of the middle range, large errors in a couple of predictions (Cases 6 and 9) might be an indication of inaccuracy on the actual yield in the record book.

**Table 9.** The absolute relative errors in percentage from the XGBoost and ensemble model.

Test set	Actual yield (t/ha)	Predicted yield (t/ha)		Absolute relative errors (%)	
		XGBoost	Ensemble	XGBoost	Ensemble
1	2.4840	2.8984	2.9563	16.68	19.01
2	3.3850	3.6097	3.7103	6.64	9.61
3	3.8390	3.7263	3.7533	2.94	2.23
4	4.2040	4.4774	4.4601	6.50	6.09
5	4.6370	4.4621	4.4053	3.77	5.00
6	5.0430	4.1725	4.2014	17.26	16.69
7	5.1855	5.2332	5.3159	0.92	2.51
8	5.3333	5.2981	5.2930	0.66	0.75
9	5.9574	5.0364	5.0957	15.46	14.47
10	6.2711	6.0164	5.9100	4.06	5.76
11	6.4050	6.3740	6.3713	0.48	0.53
12	6.8235	6.9806	6.9057	2.30	1.21
13	7.5905	6.5162	6.5896	14.15	13.19
14	8.1340	7.0573	7.0582	13.24	13.23
Average				7.51	7.88

## 5. Conclusions and future work

UAVs have been very attractive for acquiring high-resolution field images for precision agriculture and plant breeding programs. This study explored ML as well as the ensemble model for peanut yield estimation using UAV multispectral imagery. We analyzed the correlation between the individual spectral bands at various growth stages with peanut yield. The correlation results revealed that the HM stage had a significant correlation with yield. This allowed us to select the best-performing ML models to build ensemble learning for yield prediction. The results showed that the proposed ensemble approach, based on the three best ML models XGBoost, RF and SVR among the eight ML models examined, produced a consistent and comparable peanut yield prediction alongside the best performer XGBoost. Hence, rather than providing only one option to farmers, presenting both the results predicted by the XGBoost model and the ensemble scheme would give the farmers a more reliable estimate for peanut yield as mutual verification.

This work has two limitations. First, only a single-year peanut dataset is considered in this work; hence, the proposed model should be extended to multi-year peanut data to further increase the consistency of the model in future studies. Second, advanced deep learning models such as CNN should be investigated along with more agriculture input data in the future.

## Acknowledgements

The authors would like to acknowledge the Research Training Program (RTP) scholarship funded by the Australian Government and CQUniversity. Also, the authors would like to acknowledge the

support of the Queensland Department of Agriculture and Fisheries, Bundaberg research facility, Queensland, Australia during this study.

### Conflict of interest

The authors declare no conflict of interest.

### References

1. R. Nigam, R. Tripathy, S. Dutta, N. Bhagia, R. Nagori, K. Chandrasekar, et al., Crop type discrimination and health assessment using hyperspectral imaging, *Curr. Sci.*, **116** (2019), 1108–1123. <https://www.jstor.org/stable/27138003>
2. J. ten Harkel, H. Bartholomeus, L. Kooistra, Biomass and crop height estimation of different crops using UAV-based LiDAR, *Remote Sens.*, **12** (2020), 17. <https://doi.org/10.3390/rs12010017>
3. U. S. Panday, N. Shrestha, S. Maharjan, A. K. Pratihast, Shahnawaz, K. L. Shrestha, et al., Correlating the plant height of wheat with above-ground biomass and crop yield using drone imagery and crop surface model, a case study from nepal, *Drones*, **4** (2020), 28. <https://doi.org/10.3390/drones4030028>
4. A. Michez, P. Lejeune, S. Bauwens, A. A. L. Herinaina, Y. Blaise, E. C. Muñoz, et al., Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system, *Remote Sens.*, **11** (2019), 473. <https://doi.org/10.3390/rs11050473>
5. A. I. de Castro, R. Ehsani, R. C. Ploetz, J. H. Crane, S. Buchanon, Detection of laurel wilt disease in avocado using low altitude aerial imaging, *PloS ONE*, **10** (2015), 1–13. <https://doi.org/10.1371/journal.pone.0124642>
6. A. Mahlein, Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping, *Plant Dis.*, **100** (2016), 241–251. <https://doi.org/10.1094/PDIS-03-15-0340-FE>
7. P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka, E. Hernandez, Plant disease detection using hyperspectral imaging, in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, (2017), 1–8. <https://doi.org/10.1109/DICTA.2017.8227476>
8. D. Gómez-Candón, J. Torres-Sanchez, S. Labbé, A. Jolivot, S. Martinez, J. L. Regnard, Water stress assessment at tree scale: high-resolution thermal UAV imagery acquisition and processing, *Acta Hort.*, **1150** (2017), 159–166. <https://doi.org/10.17660/ActaHortic.2017.1150.23>
9. C. A. Reynolds, M. Yitayew, D. C. Slack, C. F. Hutchinson, A. Huete, M. S. Petersen, Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data, *Int. J. Remote Sens.*, **21** (2000), 3487–3508. <https://doi.org/10.1080/014311600750037516>
10. S. S. Panda, D. P. Ames, S. Panigrahi, Application of vegetation indices for agricultural crop yield prediction using neural network techniques, *Remote Sens.*, **2** (2010), 673–696. <https://doi.org/10.3390/rs2030673>
11. Z. Fu, J. Jiang, Y. Gao, B. Krienke, M. Wang, K. Zhong, et al., Wheat growth monitoring and yield estimation based on multi-rotor unmanned aerial vehicle, *Remote Sens.*, **12** (2020), 508. <https://doi.org/10.3390/rs12030508>



12. S. Guan, K. Fukami, H. Matsunaka, M. Okami, R. Tanaka, H. Nakano, et al., Assessing correlation of high-resolution NDVI with fertilizer application level and yield of rice and wheat crops using small UAVs, *Remote Sens.*, **11** (2019), 112. <https://doi.org/10.3390/rs11020112>
13. M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, F. B. Fritschi, Soybean yield prediction from UAV using multimodal data fusion and deep learning, *Remote Sens. Environ.*, **237** (2020), 111599. <https://doi.org/10.1016/j.rse.2019.111599>
14. Y. Guo, S. Chen, X. Li, M. Cunha, S. Jayavelu, D. Cammarano, et al., Machine learning-based approaches for predicting SPAD values of maize using multi-spectral images, *Remote Sens.*, **14** (2022), 1337. <https://doi.org/10.3390/rs14061337>
15. Z. Sun, X. Wang, Z. Wang, L. Yang, Y. Xie, Y. Huang, UAVs as remote sensing platforms in plant ecology: review of applications and challenges, *J. Plant Ecol.*, **14** (2021), 1003–1023. <https://doi.org/10.1093/jpe/rtab089>
16. J. Xue, B. Su, Significant remote sensing vegetation indices: A review of developments and applications, *J. Sens.*, **2017** (2017), 1–17. <https://doi.org/10.1155/2017/1353691>
17. L. Wan, H. Cen, J. Zhu, J. Zhang, Y. Zhu, D. Sun, et al., Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer-a case study of small farmlands in the South of China, *Agric. For. Meteorol.*, **291** (2020), 108096. <https://doi.org/10.1016/j.agrformet.2020.108096>
18. J. Zhou, J. Zhou, H. Ye, M. L. Ali, P. Chen, H. T. Nguyen, Yield estimation of soybean breeding lines under drought stress using unmanned aerial vehicle-based imagery and convolutional neural network, *Biosyst. Eng.*, **204** (2021), 90–103. <https://doi.org/10.1016/j.biosystemseng.2021.01.017>
19. Y. Guo, H. Wang, Z. Wu, S. Wang, H. Sun, J. Senthilnath, et al., Modified red blue vegetation index for chlorophyll estimation and yield prediction of maize from visible images captured by UAV, *Sensors*, **20** (2020), 5055. <https://doi.org/10.3390/s20185055>
20. Y. Guo, Y. Fu, F. Hao, X. Zhang, W. Wu, X. Jin, et al., Integrated phenology and climate in rice yields prediction using machine learning methods, *Ecol. Indic.*, **120** (2021), 106935. <https://doi.org/10.1016/j.ecolind.2020.106935>
21. Peanut company of Australia, How peanuts are grown, 2023. Available from: <https://pca.com.au/pca-profile/how-peanuts-are-grown/>
22. Z. Ji, Y. Pan, X. Zhu, D. Zhang, J. Wang, A generalized model to predict large-scale crop yields integrating satellite-based vegetation index time series and phenology metrics, *Ecol. Indic.*, **137** (2022), 108759. <https://doi.org/10.1016/j.ecolind.2022.108759>
23. H. García-Martínez, H. Flores-Magdaleno, R. Ascencio-Hernández, A. Khalil-Gardezi, L. Tijerina-Chávez, O. R. Mancilla-Villa, et al., Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and RGB images acquired with unmanned aerial vehicles, *Agriculture*, **10** (2020), 277. <https://doi.org/10.3390/agriculture10070277>
24. X. Zhou, H. B. Zheng, X. Q. Xu, J. Y. He, X. K. Ge, X. Yao, et al., Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery, *ISPRS J. Photogramm. Remote Sens.*, **130** (2017), 246–255. <https://doi.org/10.1016/j.isprsjprs.2017.05.003>
25. D. C. Tsouros, S. Bibi, P. G. Sarigiannidis, A review on UAV-based applications for precision agriculture, *Information*, **10** (2019), 349. <https://doi.org/10.3390/info10110349>

26. J. Kim, S. Kim, C. Ju, H. Il Son, Unmanned aerial vehicles in agriculture: A review of perspective of platform, control, and applications, *IEEE Access*, **7** (2019), 105100–105115. <https://doi.org/10.1109/ACCESS.2019.2932119>
27. T. B. Shahi, C. Xu, A. Neupane, W. Guo, Machine learning methods for precision agriculture with UAV imagery: A review, *Electron. Res. Arch.*, **30** (2022), 4277–4317. <https://doi.org/10.3934/era.2022218>
28. A. P. M. Ramos, L. P. Osco, D. E. G. Furuya, W. N. Gonçalves, D. C. Santana, L. P. R. Teodoro, et al., A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices, *Comput. Electron. Agric.*, **178** (2020), 105791. <https://doi.org/10.1016/j.compag.2020.105791>
29. J. Geipel, J. Link, W. Claupein, Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system, *Remote sens.*, **6** (2014), 10335–10355. <https://doi.org/10.3390/rs61110335>
30. A. Ashapure, S. Oh, T. G. Marconi, A. Chang, J. Jung, J. Landivar, et al., Unmanned aerial system based tomato yield estimation using machine learning, in *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*, (2019). <https://doi.org/10.1117/12.2519129>
31. A. Matese, S. F. Di Gennaro, Beyond the traditional NDVI index as a key factor to mainstream the use of UAV in precision viticulture, *Sci. Rep.*, **11** (2021), 1–13. <https://doi.org/10.1038/s41598-021-81652-3>
32. C. Bian, H. Shi, S. Wu, K. Zhang, M. Wei, Y. Zhao, et al., Prediction of field-scale wheat yield using machine learning method and multi-spectral UAV Ddata, *Remote Sens.*, **14** (2022), 1474. <https://doi.org/10.3390/rs14061474>
33. S. Fei, M. A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, et al., UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat, *Precision Agric.*, **24** (2022), 1–26. <https://doi.org/10.1007/s11119-022-09938-8>
34. A. Patrick, S. Pelham, A. Culbreath, C. C. Holbrook, I. J. De Godoy, C. Li, High throughput phenotyping of tomato spot wilt disease in peanuts using unmanned aerial systems and multispectral imaging, *IEEE Instrum. Meas. Mag.*, **20** (2017), 4–12. <https://doi.org/10.1109/MIM.2017.7951684>
35. QGIS development team, QGIS Geographic Information System, 2023. Available from: <https://www.qgis.org>
36. F. I. Matias, M. V. Caraza-Harter, J. B. Endelman, FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials, *Plant Phenom. J.*, **3** (2020), 20005. <https://doi.org/10.1002/ppj2.20005>
37. A. J. Smola, B. Schölkopf, A tutorial on support vector regression, *Statistics Comput.*, **14** (2004), 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
38. X. Zeng, S. Yuan, Y. Li, Q. Zou, Decision tree classification model for popularity forecast of Chinese colleges, *J. Appl. Math.*, (2014), 1–7. <https://doi.org/10.1155/2014/675806>
39. L. Breiman, Random forests, *Mach. Learn.*, **45** (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
40. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, **12** (2011), 2825–2830.

41. T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class adaboost, *Stat. Interface*, **2** (2009), 349–360. <https://doi.org/10.4310/SII.2009.v2.n3.a8>
42. T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), 785–794. <https://doi.org/10.1145/2939672.2939785>
43. M. M. Li, W. Guo, B. Verma, K. Tickle, J. O'Connor, Intelligent methods for solving inverse problems of backscattering spectra with noise: a comparison between neural networks and simulated annealing, *Neural Comput. Appl.*, **18** (2009), 423–430. <https://doi.org/10.1007/s00521-008-0219-x>
44. C. Tsai, C. Hsia, S. Yang, S. Liu, Z. Fang, Optimizing hyperparameters of deep learning in predicting bus passengers based on simulated annealing, *Appl. Soft Comput.*, **88** (2020), 106068. <https://doi.org/10.1016/j.asoc.2020.106068>
45. Q. Yan, J. Chen, L. De Strycker, An outlier detection method based on Mahalanobis distance for source localization, *Sensors*, **18** (2018), 2186. <https://doi.org/10.3390/s18072186>
46. B. Mishra, T. B. Shahi, Deep learning-based framework for spatiotemporal data fusion: an instance of landsat 8 and sentinel 2 NDVI, *J. Appl. Remote Sens.*, **15** (2021), 034520. <https://doi.org/10.1117/1.JRS.15.034520>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)