ORIGINAL RESEARCH **Open access**

Remote Sensing and Artificial Neural Networks for Generalizable Peanut Yield Prediction

Rafael Mendoza[1*], Isabela Cruz[1], Carlos Lopez[2], Laura Gutierrez[3]

Abstract

Estimating crop yield is crucial for agricultural oversight and strategic planning. Remote Sensing (RS) and Artificial Neural Networks (ANN) are gaining traction as viable instruments for forecasting these agronomic metrics. This work aimed to merge RS data with ANN models to forecast peanut yield in advance and remotely. The trial spanned 11 commercial fields, apportioned into 6 in the 2020/21 cycle and 5 in the 2021/22 cycle. The inputs used to construct the models consisted of vegetation indices (EVI, GNDVI, MNLI, NLI, NDVI, SAVI, and SR) derived from high-resolution satellite imagery on five dates, ranging from 1 to 30 days before harvest initiation. Vegetation Index (VI) information from the 20/21 cycle was used to train Multilayer Perceptron (MLP) and Radial Basis Function (RBF) Artificial Neural Networks (ANNs) for parameter adjustment. The generated formulas were later transferred to the fields of the following cycle for generalization, and the models were recalibrated using pooled data from both cycles. Both network architectures demonstrated the ability to generate predictions from VIs as inputs during validation and recalibration, resulting in gains in model precision and accuracy. The validation step revealed a strong aptitude for extending peanut yield variability predictions to unfamiliar fields. The MLP network outperformed the other, yielding a MAPE of 9.3% at 30 days pre-harvest and a coefficient of determination of 0.80. The VIs that proved most influential as inputs were EVI, SAVI, and SR. The fusion of RS with ANN represents a formidable approach for predicting peanut yield and aiding growers in crop management. The outcomes emphasize the worth of building peanut yield forecasting models over successive years, factoring in the interplay between genotype and environment to strengthen model resilience. It is equally critical that these models remain functional in novel locations, as evidenced by this research, which confirmed satisfactory generalization across distinct sites, even under differing management regimes and cultivars.

Keywords Artificial intelligence, Multilayer perceptron, Digital agriculture, Vegetation indices, Model validation, Radial basis function

*Correspondence:
Rafael Mendoza
rafael.mendoza@unam.mx
[1] Department of AI for Avocado and Berry Crop Forecasting, Faculty of Agriculture, National Autonomous University of Mexico, Mexico City, Mexico
[2] Department of Machine Learning for Soil Erosion and Degradation Analysis, Faculty of Agriculture, Monterrey Institute of Technology, Monterrey, Mexico
[3] Department of Smart Fertilization Scheduling, Faculty of Agriculture, University of Guadalajara, Guadalajara, Mexico

Introduction

Projecting agricultural harvests is of paramount significance for cultivation planning and governance, supporting growers, cooperatives, and technical experts in making choices to enhance farming operations [1]. For the peanut crop in particular, characterized by indeterminate growth habits and subterranean fruit formation, deciphering its dynamics and variability becomes problematic. Conventional peanut yield forecasts depend on field inspections by technical staff, rendering the estimation workflow lengthy, strenuous, and non-representative of spatial variability, ultimately delivering

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

averaged figures. Hence, creating trustworthy prediction systems built on satellite data and artificial intelligence can illuminate variability in peanut crop productivity and serve as a compelling alternative to the entire peanut supply chain.

Analytical techniques, such as ANNs coupled with RS, have been successfully applied to estimate yields of diverse crops, including maize, soybean, wheat, and paddy rice [1-4]. Typically, these methods are paired with Vegetation Indices (VIs) serving as predictor variables for yield estimation [5, 6], which hold a strong capacity to remotely account for field-level yield fluctuations [7]. The synergy between ANNs and VIs stems from their capability to capture intricate associations linking VIs to yield outcomes. Nevertheless, several studies advocate including VIs alongside other independent variables, such as climatic data, edaphic properties, and crop biophysical characteristics [8].

Within the spectrum of machine learning approaches and ANN architectures, RBF and MLP are the most frequently utilized. The deployment of these networks for crop yield forecasting has been widespread, with satisfactory findings reported repeatedly [3, 9]. In Brazilian production settings, investigations using these ANN types have already been conducted to develop predictive frameworks for agronomic traits in crops such as coffee [10]. MLP and RBF networks are vital for addressing complex problems; MLPs learn via stacked layers, whereas RBF accommodate data through non-linear fitting, affording adaptability for targeted analyses.

A significant shortcoming in earlier research is the absence of proper testing of the formulated predictive models, particularly regarding the transferability of the derived equations to fresh sites. Consequently, the relevance of these models across heterogeneous contexts, encompassing zones with disparate management practices, seasons with distinct weather regimes, and different cultivars, remains unresolved. Such validation is indispensable for gauging the consistency and efficacy of yield-projection models. Moreover, grasping how the expansion of a genuine field-collected dataset influences predictive models is key. This insight can clarify the need for periodic model updates to sharpen their precision and accuracy.

Against this backdrop, marrying VIs with ANNs offers a powerful avenue for predicting peanut productivity. Additionally, it is anticipated that the models arising from this methodology can demonstrate the capacity to generalize yield variability across varying field scenarios. This would pave the way for more potent decision-support tools, especially tied to pre-harvest and harvest operations, bolstering farmer confidence through refined strategic foresight. Accordingly, the present work sets out to: 1) forecast peanut yield ahead of harvest using orbital-image-derived vegetation indices processed by MLP and RBF algorithms; 2) pinpoint the algorithm with the stronger ability to extend its predictions of peanut yield variability; 3) execute the calibration, validation, and recalibration of the vegetation-index-based models.

Materials and Methods

Experimental area and field data acquisition

The investigations took place across 11 commercial peanut-producing fields in the Alta Mogiana region, São Paulo state, Brazil, spanning two harvest cycles (2020/2021 and 2021/2022). The trial fields were picked to represent areas with contrasting soil profiles (ROSSI, 2017) and planted varieties. The specifics of these areas and their attributes are outlined in **Table 1**.

Table 1. Information about the study areas for the two seasons.

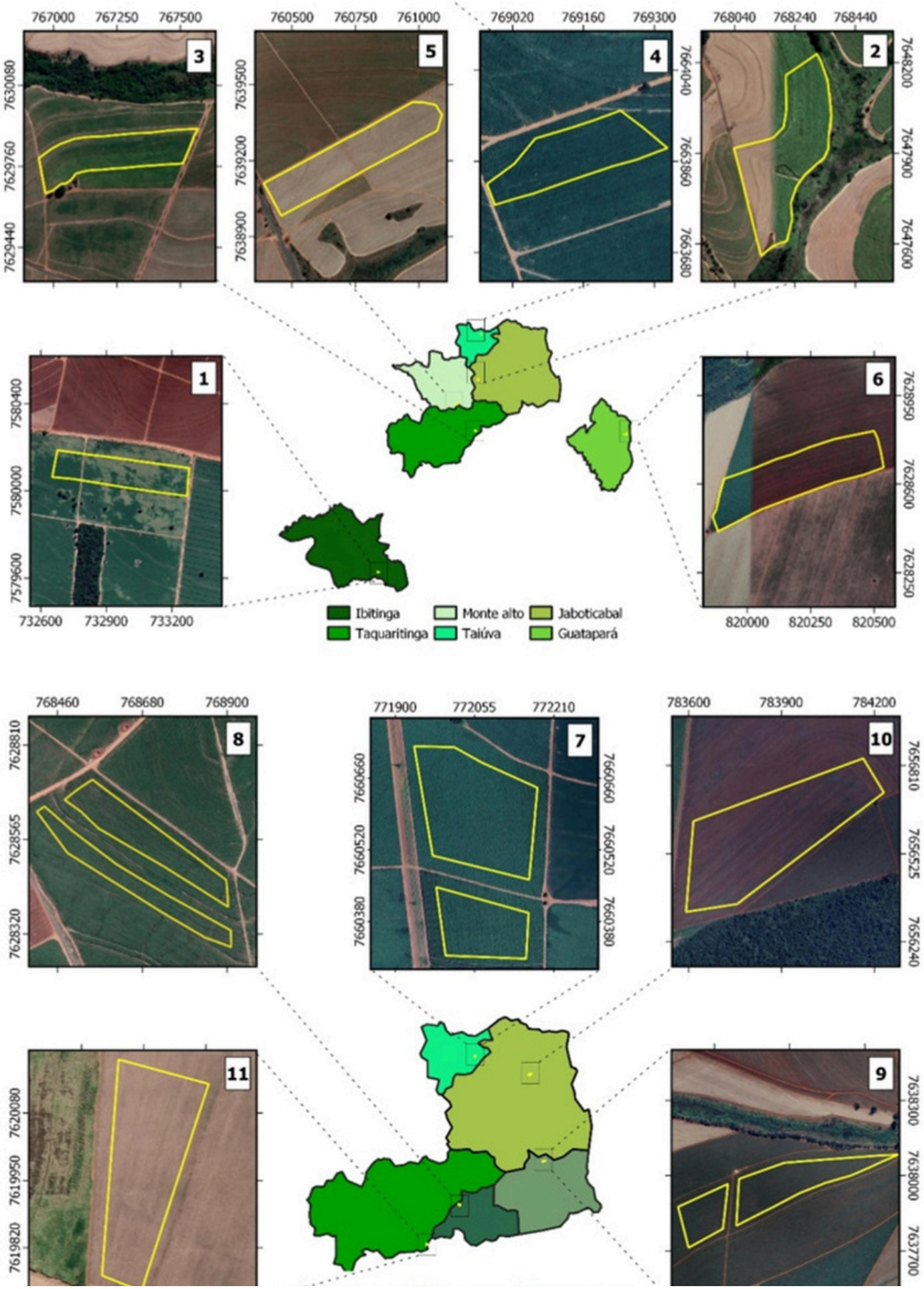
Season 2020/2021					
Area	**Location**	**Cultivar**	**Soil**	**Area (hectare)**	**Cycle (days)**
1	Ibitinga	IAC OL3	Dystrophic Red Latosols	8	120

© 2023 The Author(s). | Published by Agro Intelligence Research Press

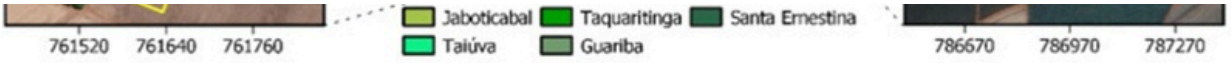
2	Pradópolis	IAC OL3	Dystroferric Red Latosols			11	120
3	Santa Ernestina 1	IAC OL3	Dystrophic Red Latosols			8.1	120
4	Taiuva 1	IAC OL3	Eutrophic Red Argisols			10	120
5	Monte Alto 1	IAC OL3	Eutrophic Red Argisols			10.5	120
6	Monte Alto 2	IAC OL3	Eutrophic Red Argisols			10.2	120
Season 2021/2022							
7	Taiuva 2	IAC OL3	Dystrophic Red Latosols	8	120		
8	Santa Ernestina 2	IAC OL3	Dystrophic Red Latosols	8.8	120		
9	Guariba	IAC 503	Dystrophic Red Latosols	8.7	150		
10	Jaboticabal	IAC 503	Dystroferric Red Latosols	8.1	150		
11	Taquaritinga	IAC 503	Eutrophic Red Argisols	10.2	150		

For the 2020/2021 cycle, the trial was laid out in zones comprising 20 sampling locations, placed 50 meters apart following a strip arrangement. In the cycle that followed, samples were taken at 30 locations, likewise set 50 meters apart **(Figure 1)**. The inaugural season provided a productivity dataset of 120 observations. In contrast, the second season delivered 150 yield data points.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



© 2023 The Author(s). | Published by Agro Intelligence Research Press



The positions gathered in the fields were georeferenced using a DJI P4 Multispectral multirotor Remotely Piloted Aircraft (RPA) (Shenzhen, China), which was operated autonomously via the DJ Ground Station Pro software. This RPA carries a multi-frequency Global Navigation Satellite System (GNSS) receiver that tracks signals from constellations like the Global Positioning System (GPS), Global'naya Navigatsionnaya Sputnikovaya Sistema (GLONASS), Beidou Navigation Satellite System (BeiDou), and Galileo, providing centimeter-grade horizontal (1 cm + 1 ppm) and vertical (2 cm + 1 ppm) positioning accuracy (DJI D-RTK2, Shenzhen, China). Sorties were executed at 90 m above ground, with ground control points demarcated using lime and flags to facilitate their identification in the imagery and during field truthing. The coordinates for these spots were pulled from the constructed orthomosaic and matched to satellite scenes. Inside the QGIS environment, a 3 m buffer zone was generated around each ground control point to extract reflectance values. Vegetation indices were then computed from reflectance values within the buffer zones surrounding the productivity measurement spots.

Peanut yield was assessed 1 day before mechanical inversion. At each sampling location, all plants within a 2 m² quadrat were collected. The harvesting was performed manually within each plot. Afterwards, the material underwent manual pod stripping and cleaning, eliminating foreign matter such as soil aggregates, foliage, and stems. Samples were dried in a forced-air oven at 105 °C for a full day until the mass remained constant to establish the pod dry biomass. This allowed yield to be expressed as kg ha^{-1}, normalized to an 8 % moisture basis **(Figure 2)**.

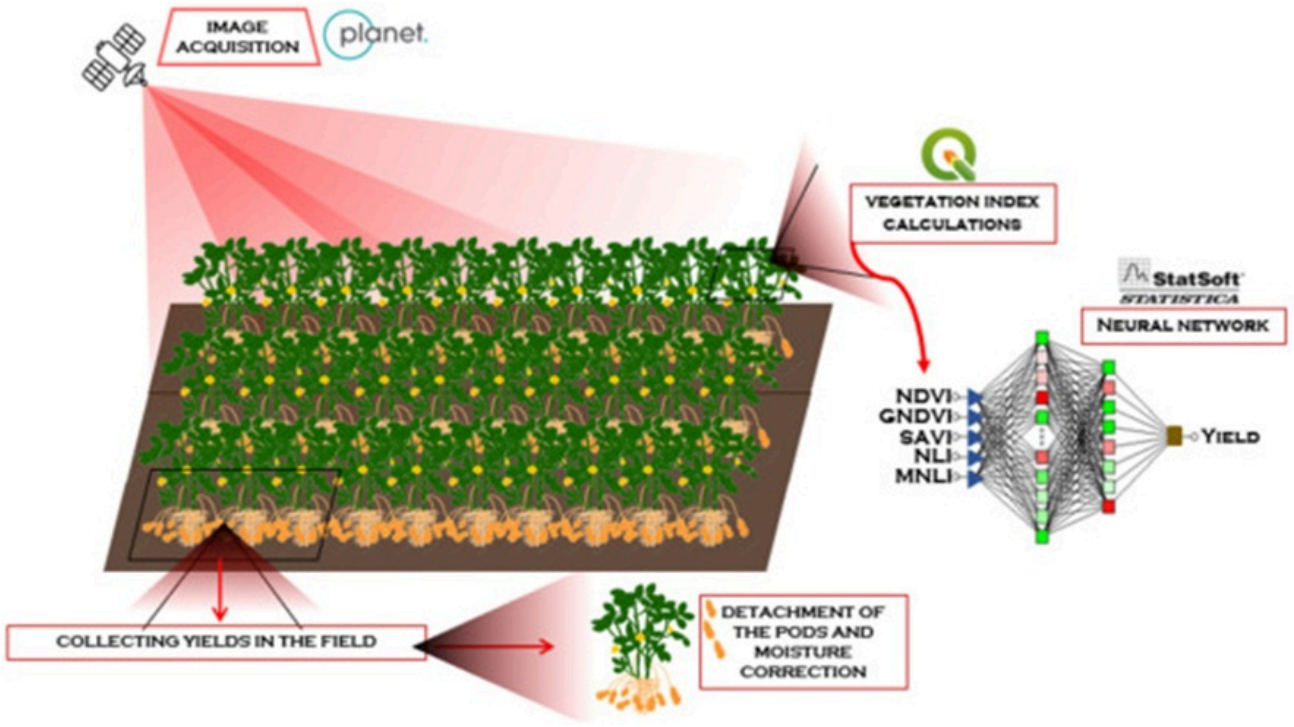


Figure 2. Schematic representation of the methodology.

Satellite image acquisition

Surveillance of the experimental sites from orbit occurred over 30 days before digging, encompassing five time points: 30, 20, 14, 8, and 1 Days Before Harvest (DBH). The orbital data were sourced through the CubeSat PlanetScope (PS) platform sensor. At present, the PS fleet numbers upwards of 180 nanosatellites circling the sun, which together furnish daily revisit capacity at a ground sample distance of 3 meters alongside eight spectral wavebands. For the scope of this

© 2023 The Author(s). | Published by Agro Intelligence Research Press

investigation, only the four-band PS product was drawn upon: Blue (455–515 nm), Green (500–590 nm), Red (590–670 nm), and Near Infrared (780–860 nm).

Atmospheric adjustment of the orbital scenes used the Surface Reflectance product from PlanetScope, which harmonizes observations with prevailing atmospheric states and mitigates inconsistencies in spectral signatures. These adjustments, applied to the full archive of orthorectified scenes, rely on top-of-atmosphere (TOA) reflectance values and coefficients stored in the Planet Radiance product. Bottom of Atmosphere (BOA) Reflectance is resolved pixel by pixel by referencing Look-Up Tables (LUTs) constructed from the 6SV2.1 radiative transfer code, which map the link between TOA reflectance and BOA Reflectance.

Vegetation Indices were cherry-picked, drawing on scholarly reports that substantiated their ability to gauge peanut ripening and productivity across a broad set of cultivated species—wheat, maize, dry bean, sweet potato, rice, and soybean among them [1, 11, 12]. The vast majority of these plants exhibit above-ground architectural traits that mirror those of peanut plants. The morphological resemblance that peanuts bear to other crops implies that vegetation indices are highly transferable for projecting their yield, a fact that underscores the careful curation of this particular VI set in the present work **(Table 2)**.

Table 2. Vegetation index used as input in predictive models.

IV*	Equation	Reference
NDVI	$\frac{\mathrm{NIR}-\mathrm{Red}}{\mathrm{NIR}+\mathrm{Red}}$	Rouse *et al.* [13]
GNDVI	$\frac{\mathrm{NIR}-\mathrm{Green}}{\mathrm{NIR}+\mathrm{Green}}$	Gitelson e Merzlyak [14]
MNLI	$\frac{(\mathrm{NIR2}-\mathrm{Red})(1+\mathrm{L}^{**})}{(\mathrm{NIR2}+\mathrm{Red}+\mathrm{L}^{**})}$	Gong *et al.* [15]
EVI	$\frac{2.5(\mathrm{NIR}-\mathrm{Red})}{(\mathrm{L}^{**}+\mathrm{NIR}+\mathrm{C1Red}-\mathrm{C2Blue})}$	Justice *et al.* [16]
NLI	$\frac{\mathrm{NIR2}-\mathrm{Red}}{\mathrm{NIR2}+\mathrm{Red}}$	Goel and Qin [17]
SAVI	$\frac{(1+\mathrm{L}^{**})(\mathrm{NIR}-\mathrm{Red})}{(L^{**}+NIR+Red)}$	Huete [18]

NDVI: Normalized Difference Vegetation Index; GNDVI: Green Normalized Difference Vegetation Index; MNLI: Modified Non-Linear Index; EVI: Enhanced Vegetation Index; NLI: Non-Linear Index; SAVI: Soil-Adjusted Vegetation Index; SR: Simple Ratio.

∗ VI = Vegetation Index; L = 0.5; C1 = 6; C2 = 7.5.

Data analysis

A boxplot was used to visually inspect the dataset's structure and dispersion. This graphical method reliably flags medians, interquartile intervals, and suspected outliers, allowing the culling of aberrant yield records. Boxplots were constructed for the productivity of each area under study, enabling an uncomplicated side-by-side comparison across distinct fields.

The models aimed at projecting peanut productivity were assembled in Statistica 7 (StatSoft Inc., Tulsa), using two ANN algorithms: Multilayer Perceptron (MLP) and Radial Basis Function (RBF).

The seven VIs introduced earlier served as independent variables fed into the networks to predict the target variable (yield). At the outset, within the 2020/2021 cycle, the pool of 120 yield observations harvested across the different fields was partitioned into training and test sets to train the prediction models (calibration step) for each of the five pre-harvest

© 2023 The Author(s). | Published by Agro Intelligence Research Press

dates. Those calibrated equations were then applied to five untouched fields in the following season for out-of-sample validation. Networks yielding error magnitudes beyond 1000 kg ha^{-1} were discarded, as inaccuracies exceeding that threshold are regarded as excessive. This boundary was set following dialogue with cooperative partners, whose field staff reported that their own ocular estimates routinely fell within comparable error bands. To wrap up, a recalibration round was undertaken, pooling the totality of ground-collected data from both cycles—270 yield entries—to gauge how the accumulation of additional observations modulates model behaviour.

Irrespective of whether the model was calibrated or recalibrated, the assembled data were split into 80% for network training and 20% for independent testing. The respective operations were executed with the Neural Network toolbox bundled in Statistica 7. During the training and testing phase, the ANNs underwent 3000 learning cycles, after which the top 10 candidates were stored—5 representing the MLP architecture and 5 the RBF architecture. The number of neurons oscillated between 1 and 20 across both algorithms, whereas for the MLP topology, the hidden layer count ranged from 1 to 3.

Configuring network topology, fine-tuning hyperparameters, and choosing a training regime can render the workflow drawn out. To circumvent this, the Intelligent Problem Solver (IPS) module streamlines model construction, sidestepping iterative guesswork by adopting a heuristic combinatorial search that accelerates the identification of the most suitable layout. The IPS output also provided sensitivity analyses for the resulting models, illuminating the relative weights of each input variable in determining model output.

For the validation stage, the Run Existing Model (REM) utility was activated, allowing previously trained and archived network equations to be redeployed without re-running the learning procedure, thereby enabling direct application to a fresh data partition and new field locations. The predictive skill of the ANNs was appraised using scatter plots aligned to a 1:1 correspondence line, where accuracy and bias were quantified through the mean absolute error (MAE) and the mean absolute percentage error (MAPE) for the forecasting discrepancy, and the coefficient of determination (R²) for the degree of model fit, following the formulations laid out in **Eqs. 1-3**, respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i \right) \quad (1)$$

$$MAPE = \frac{\sum_{i=1}^{n} \left(\left| \frac{Y_i - \overline{Y}_i}{\overline{Y}_i} \right| * 100 \right)}{n} \quad (2)$$

Where n denotes the number of data points, Yi denotes the variable value actually measured in the field, and Y¯i denotes the value estimated by the model.

$$R^2 = \frac{SQR}{SQT} \quad (3)$$

Where SQR corresponds to the sum of squared residuals, and SQT equals the sum of total squares.

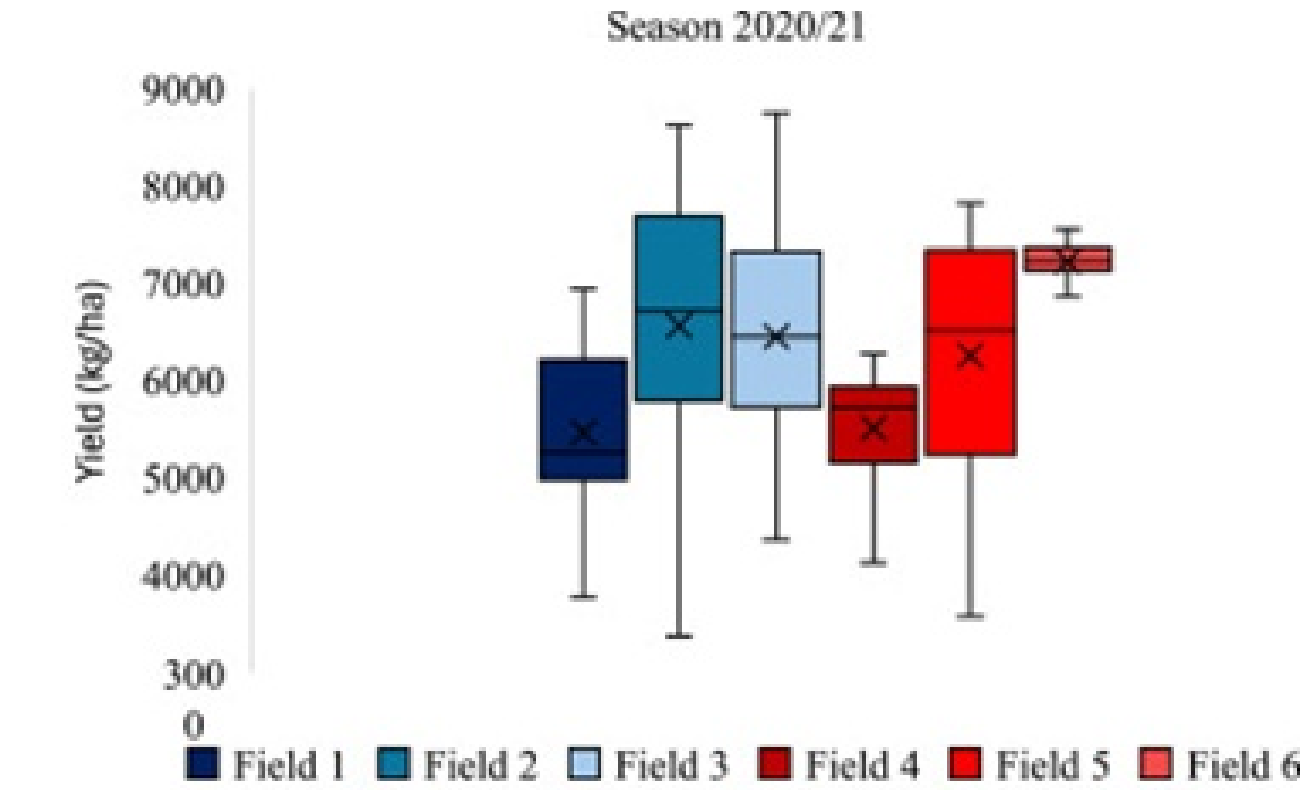
Results and Discussion

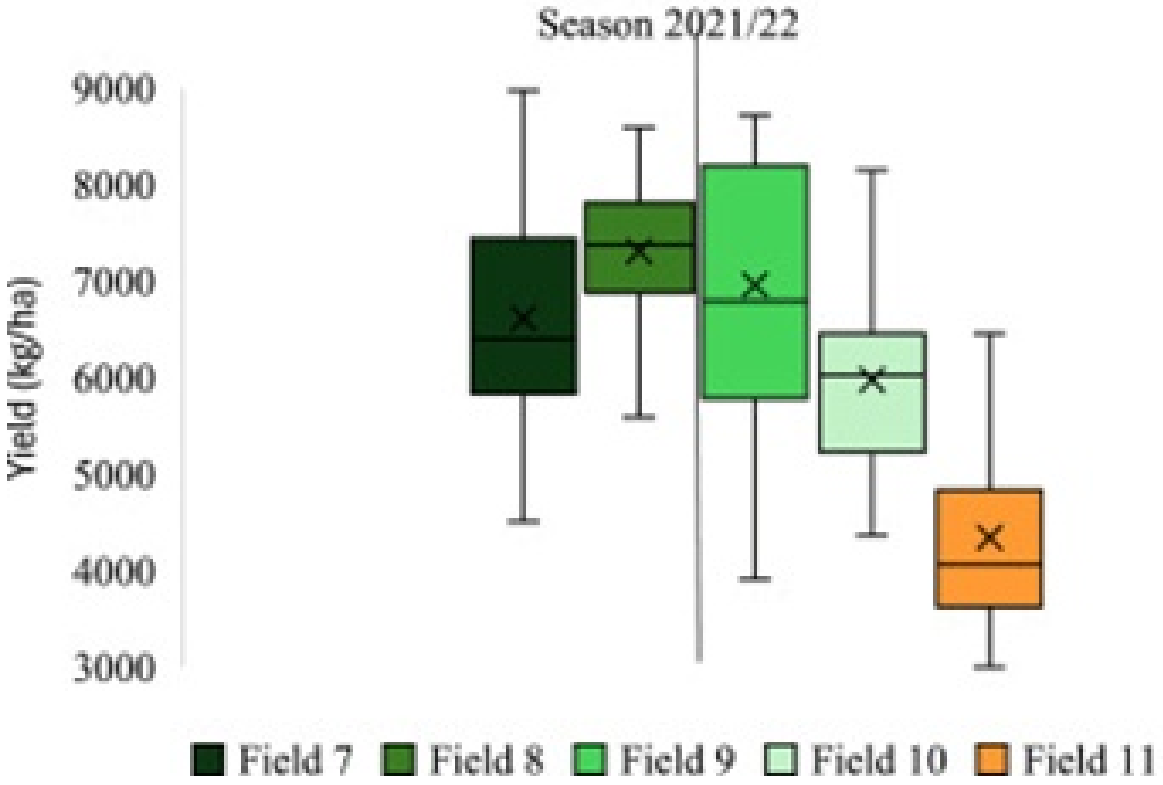
© 2023 The Author(s). | Published by Agro Intelligence Research Press

Exploratory analysis of yield data

Inspection of the exploratory data summaries shows fields 2 and 3 returning nearly equivalent mean yields **(Figure 3)**. By contrast, field 1 posted a notably depressed average within the same soil grouping, trailing by close to 1000 kg ha^{-1} **(Table 1)**. Yield dispersion throughout the 2020/21 cycle was amplified on plots situated on Latosol, as reflected in elongated interquartile ranges (IQRs) and extended whisker extremities (upper and lower thresholds), a phenomenon most conspicuous in field 2, whose variance exceeded that of all others.

© 2023 The Author(s). | Published by Agro Intelligence Research Press





© 2023 The Author(s). | Published by Agro Intelligence Research Press

Figure 3. Exploratory analysis of peanut yield in the 20/21 (A) and 21/22 (B) seasons.

Within the Argisol-managed fields, field 6 deserves particular mention for attaining both the highest average (7229.34 kg ha^{-1}) and median (7234.75 kg ha^{-1}) across the entirety of the first growing cycle. Yield fluctuation is more pronounced in field 5, as evidenced by broader IQRs, with ceiling observations reaching 7826.70 kg ha^{-1} and floor values at 3577.44 kg ha^{-1}.

Panel 3B contains five boxplots for fields 7 through 11, stratified by planted variety (IAC OL3 and IAC 503) and soil type, and contrasting soil types. The leftmost pair of green boxes belongs to IAC OL3, whereas the three on the right correspond to IAC 503; the solitary orange box designates the clay-textured soil (Field 11). Regarding the IAC OL3 material, the uppermost mean and median both originated from field 8, which had an average of 7313.98 kg ha^{-1}. The whisker span of field 9 indicates the greatest yield dispersion within this cultivar.

Considering the IAC 503 entries, fields 10 and 11 shared broadly comparable IQRs and variance magnitude, yet their mean yields fell below that of field 9. Field 11 produced the lowest average and median in the entire dataset—4344.68 kg ha^{-1} and 4062.40 kg ha^{-1}, respectively. This depressed performance was attributed to uncontrolled sugarcane volunteer regrowth on the site, which triggered interspecific competition and ultimately reduced pod production. The situation may also help explain why Latosols yielded a higher average for the IAC 503 cultivar. Field 9 combined the second-strongest average for the season with the largest spread of values.

Validation of prediction models

The accompanying graphical outputs illustrate the operationalization of ANNs—specifically MLP and RBF variants—for estimating groundnut productivity during the 2020/2021 cycle. Results are presented through two complementary visual formats: bivariate scatterplots linking field-observed yield against algorithm-generated estimates, and sequential line traces charting observed versus predicted values point by point for the test subset **(Figure 4)**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

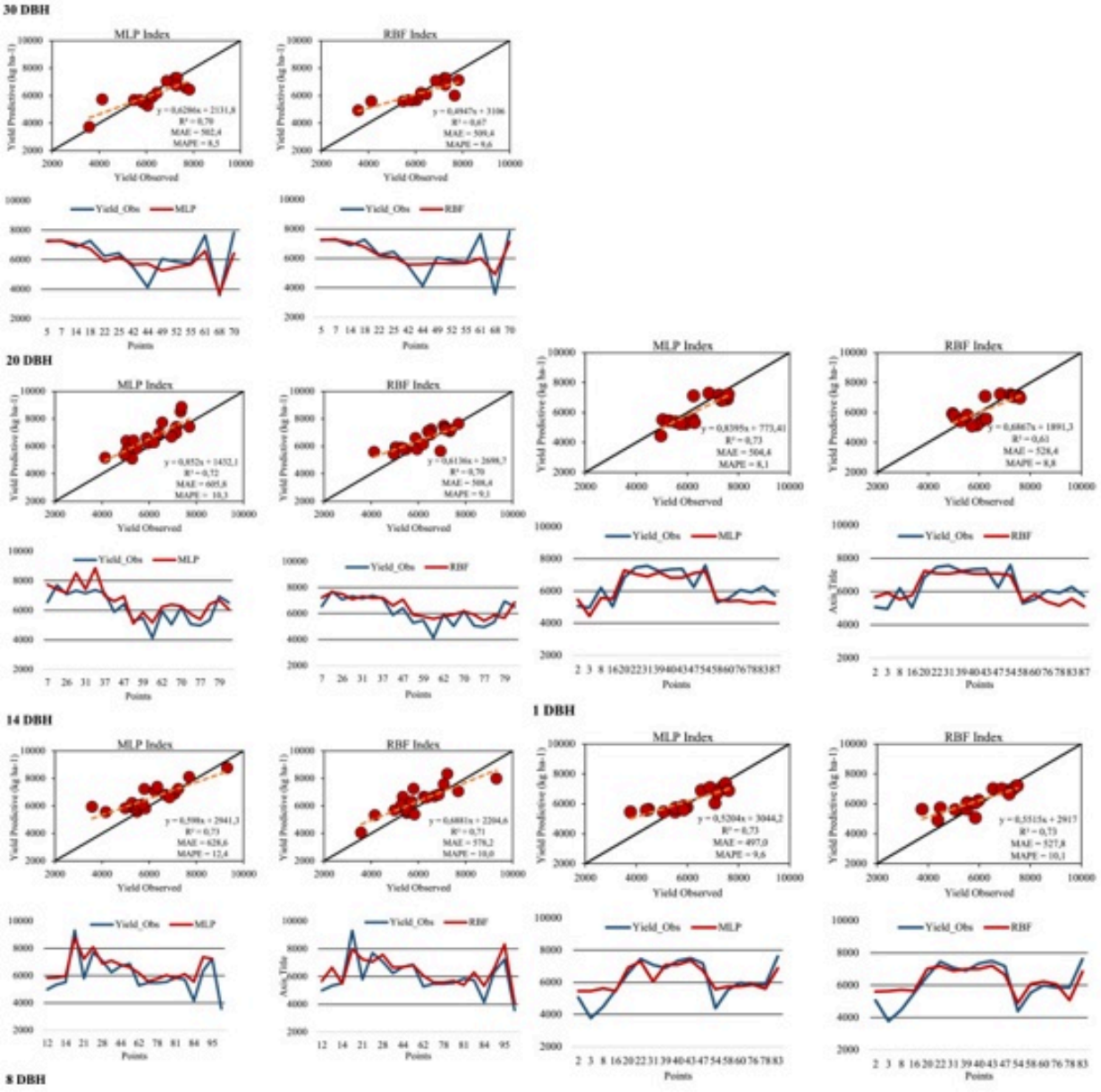


Figure 4. Calibration of Peanut Yield Prediction Models with Data from the 2020/21 Season.

A closer examination of the compilations that contrast MLP- and RBF-based groundnut yield forecasts across successive pre-digging moments (30, 20, 14, 8, and 1 DBH) reveals that both model structures consistently uphold R^2 readings beyond the 0.60 mark throughout all scenarios, indicative of a stable capacity to reproduce the variation inherent in the empirically recorded yield figures. MLP formulations tended to register fractionally stronger R^2 scores relative to RBF, implying that a marginally larger proportion of the actual yield variance was being captured. The precision band (R^2) for MLP-generated models extended from 0.70 to 0.73, whilst that of RBF-derived counterparts stretched from 0.61 to 0.73.

MAPE figures for the MLP architecture were slightly higher than those for RBF, except for the 20- and 14-DBH intervals. Conversely, RBF yielded marginally inflated MAE values in those same two assessment windows, indicating somewhat coarser absolute deviations. The MAE window for MLP spans 497 kg ha^{-1} to 628 kg ha^{-1}, whilst for RBF, the oscillation in error stayed beneath 70 kg ha^{-1}. MAPE tracked the same behavioural pattern as MAE (MLP = 8.1 to 12.4 %; RBF = 8.8 to 10.1 %).

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The rhythmic undulations displayed in the line plots do not suggest any meaningful tightening of estimate fidelity as the harvest date approaches. To summarize, both MLP and RBF formulations demonstrated robustness for the yield projection assignment, with MLP exhibiting a slight precision advantage. Selection between the two ought to weigh ancillary considerations—processing overhead, convergence time, and implementation practicality—acknowledging that the precision differential is comparatively minor.

Importance of input variables in validation models

Across the 2020/2021 cycle, sensitivity analyses were applied to the constructed MLP and RBF models to identify which of the seven vegetation indices (VIs) exerted the greatest influence on groundnut yield estimation **(Table 1)**. In the MLP configuration, NLI was the most responsive predictor at 30 days before digging. By the 20-day milestone, SAVI had the greatest relative weight, and MNLI ranked first at 14 DBH. SR ranked as the most sensitive factor on two distinct occasions—the 8- and 1-DBH evaluations. Spanning all five forecasting dates, EVI, GNDVI, and MNLI featured among the influential inputs on every single occasion.

Shifting to the RBF configuration, SAVI proved dominant at 30 DBH, while at 20 days before lifting, MNLI occupied the primary position. At both the 14- and 8-day thresholds, GNDVI emerged as the most critical variable, and with 1 DBH remaining, EVI rose to prominence. For the RBF algorithm, the indices EVI and SAVI demonstrated consistently elevated sensitivity throughout virtually the entire sequence of dates, the lone departure being the 8 DBH point **(Table 3)**.

Table 3. Sensitivity analysis for determining important variables within the generated models for MLP and RBF algorithms in the 2020/21 season.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Algorithm MLP

VI	Days to Harvest (20/21)				
	30 DBH	20 DBH	14 DBH	8 DBH	1 DBH
EVI	3	5	2	4	2
GNDVI	5	4	3	6	5
MNLI	4	2	1	5	3
NDVI			4	2	4
NLI	1				
SAVI	2	1		3	
SR		3		1	1

Algorithm RBF

VI	Days to Harvest (20/21)				
	30 DBH	20 DBH	14 DBH	8 DBH	1 DBH
EVI	3	4	2		1
GNDVI			1	1	7
MNLI	2	1	3		6
NDVI			6		3
NLI		3	5		4
SAVI	1	2	4		5
SR			7	2	2

Transfer learning for prediction models

The bubble plots provide a comparative portrayal of how forecasting equations trained on 2020/2021 observations, all tied to the OL3 cultivar, behaved when extrapolated **(Figure 5)**. It is worth underscoring that only those projections whose validation inaccuracy fell below the 1000 kg ha^{-1} ceiling are displayed. The equations were exported to previously unseen locations in the 2021/2022 cycle—specifically, fields 10 and 11 cultivated with IAC 503, alongside fields 7 and 8 planted to IAC OL3—to gauge their transferability. Field 9 did not produce any prediction meeting the sub-1000 kg ha^{-1} criterion. Model performance was examined relative to pre-harvest intervals (30, 20, 14, 8, and 1 DBH) within each target zone.

Concentrating first on the IAC OL3 variety, the MLP calibration equations yielded error magnitudes of 502.4, 628.3, and 497.0 kg ha^{-1} for the 30, 14, and 1 DBH dates, respectively. Upon transfer to field 8, the corresponding validation inaccuracies amounted to 550.5, 576.3, and 539.5 kg ha^{-1} for those identical assessment moments. In field 7, MAE values from the validation step were 900.0 and 733.7 at 30 and 1 DBH, respectively. Remaining with the same variety but switching the algorithmic approach to RBF, calibration inaccuracies settled at 528.4 and 509.4 kg ha^{-1} for 30 and 8 DBH, respectively; for those same dates, the field 8 predictions delivered MAEs of 708.9 and 648.4 kg ha^{-1}, respectively. Generating a forecast for field 7 was achievable solely at 20 DBH, where the calibration equation carried an MAE of 508.4, and the ensuing validation rose to 869.6 kg ha^{-1}.

Turning now to the IAC 503 variety, the MLP approach showed calibration inaccuracies of 504.4 and 605.8 kg ha^{-1} at 20 and 8 DBH, respectively. When the calibration model was applied to field 10, validation yielded an MAE of 920.5 kg ha^{-1} at 8 DBH. For field 11, validation inaccuracies were 940.1 and 648.3 kg ha^{-1} for the 20- and 8-DBH windows, respectively. Using the RBF approach for IAC 503, the calibration step yielded errors of 578.2, 528.4, and 527.8 kg ha^{-1}

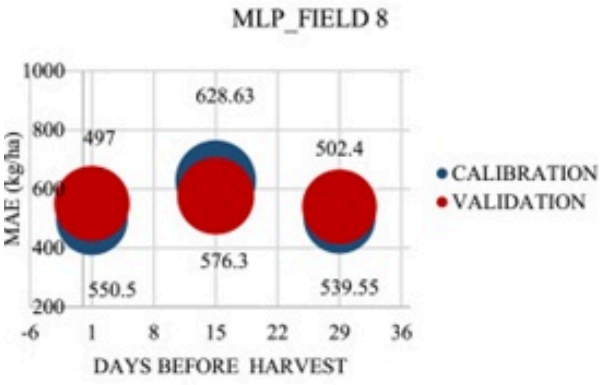
© 2023 The Author(s). | Published by Agro Intelligence Research Press

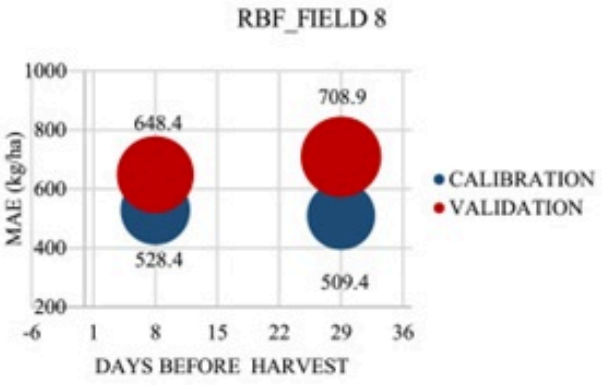
for 14, 8, and 1 DBH, respectively. Within field 10, the validation error climbed substantially, reaching 889.4 kg ha^{-1} at 14 DBH, whereas at 8 and 1 DBH the MAE readings were 697.7 and 774.3 kg ha^{-1}, respectively. Notably, the RBF approach failed to generate even a single prediction beneath the 1000 kg ha^{-1} benchmark for field 11 on any assessment date.

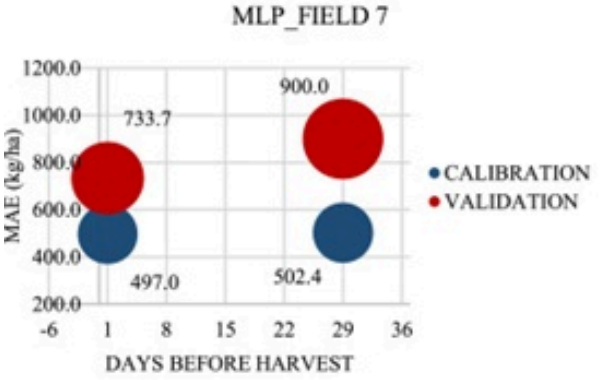
To summarise, the equations performed markedly better on the calibration dataset (2020/2021) than on new deployment sites (2021/2022). The MLP approach was marginally more accurate than RBF in field 8, yet both algorithmic families struggled in fields 7 and 10. Furthermore, neither approach managed to deliver yield projections across every available date. These observations underscore the imperative to refine the model to better align with the idiosyncrasies of the novel target environments. Addressing this may require either architectural fine-tuning or the injection of additional, more representative observations from the localities in question, steps that could reduce MAE and enhance the durability and broader utility of the forecasting frameworks across a wider spectrum of agronomic settings **(Figure 5)**.

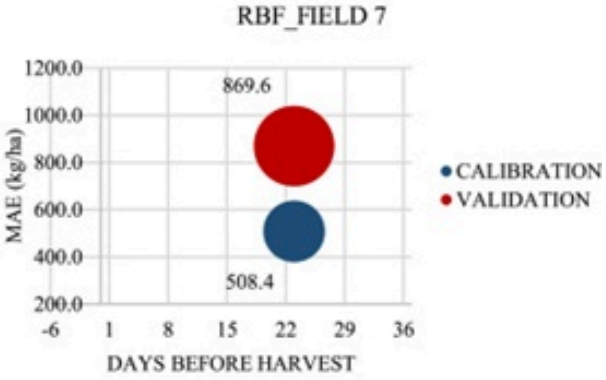
© 2023 The Author(s). | Published by Agro Intelligence Research Press

A) IAC OL3

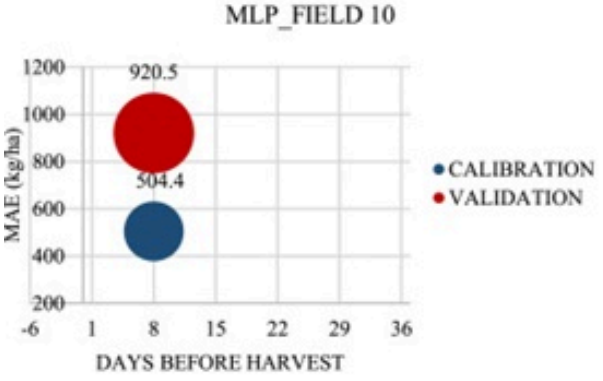


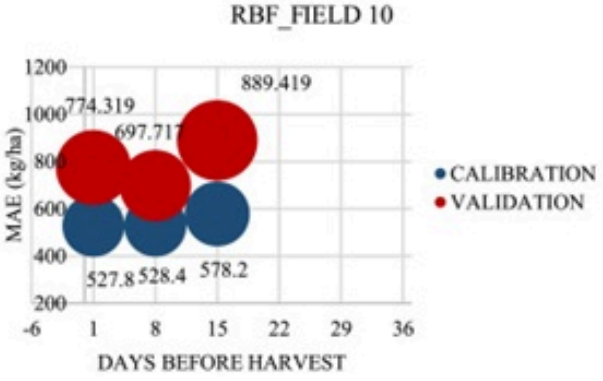


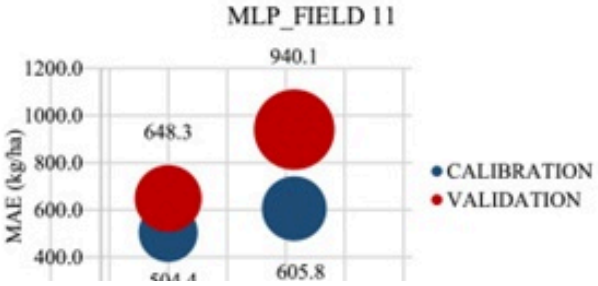




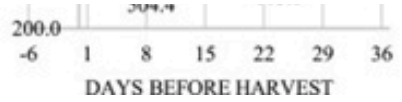
B) IAC 503







© 2023 The Author(s). | Published by Agro Intelligence Research Press



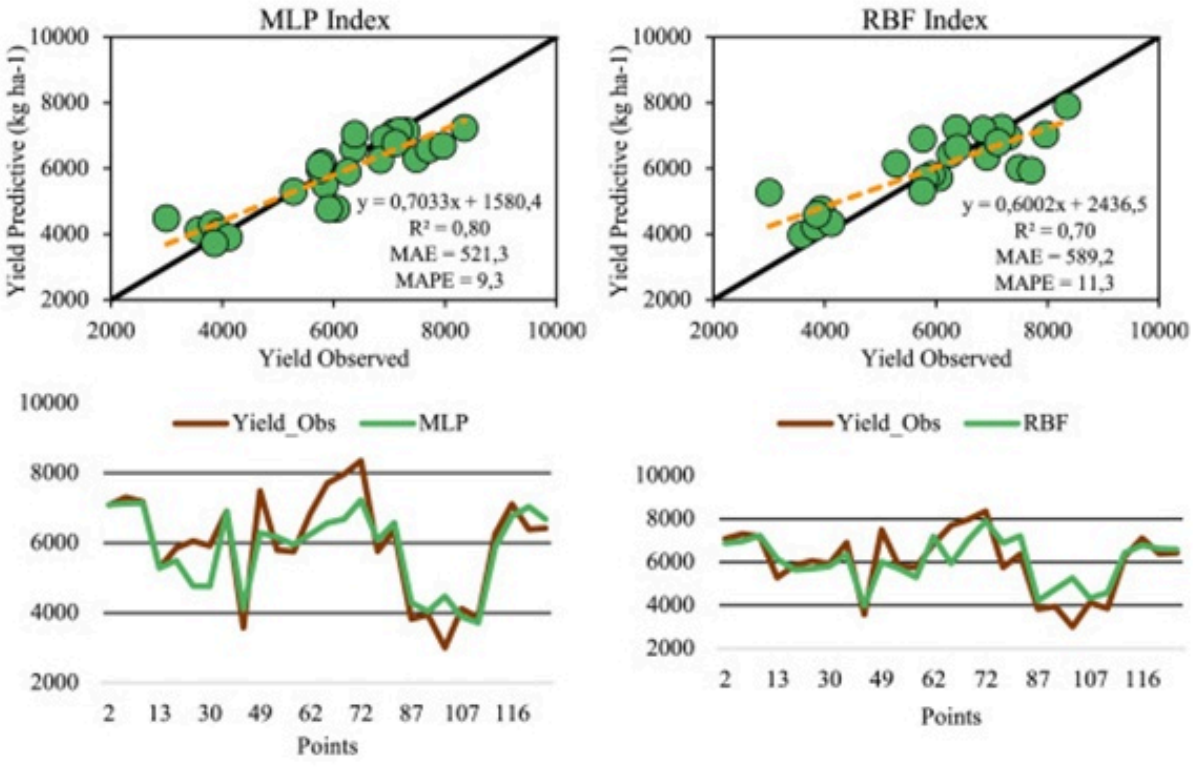
Recalibration of prediction models

What follows are the outputs stemming from retrained models that drew upon a consolidated pool of observations harvested across two separate cropping cycles (2020/21 and 2021/22), encompassing a collective of 11 field sites **(Figure 6)**. The underlying logic motivating this data enrichment exercise was the assumption that expanding the training corpus would produce formulations of greater sturdiness and fidelity. A noticeably tighter alignment emerges from the dual-season models compared with their single-season predecessors, as shown in **Figure 4**. The MLP architecture delivered R^2 values consistently above 0.80, except at 20 DBH, where the figure settled at 0.72. Corroborating this precision were MAE readings compressed within the 418.1–523.4 kg ha^{-1} range and MAPE values constrained between 6.3% and 9.7%—a considerable leap forward relative to the preliminary formulations depicted in **Figure 2**. By comparison, the RBF variant, while still exhibiting a defensible predictive grasp with R^2 spanning 0.53 to 0.74, was surpassed by MLP at every single forecasting juncture without exception. Its associated MAE band ranged from 544.2 to 636.0 kg ha^{-1}, with MAPE values ranging from 8.4% to 11.7%, indicating greater scatter in its estimates.

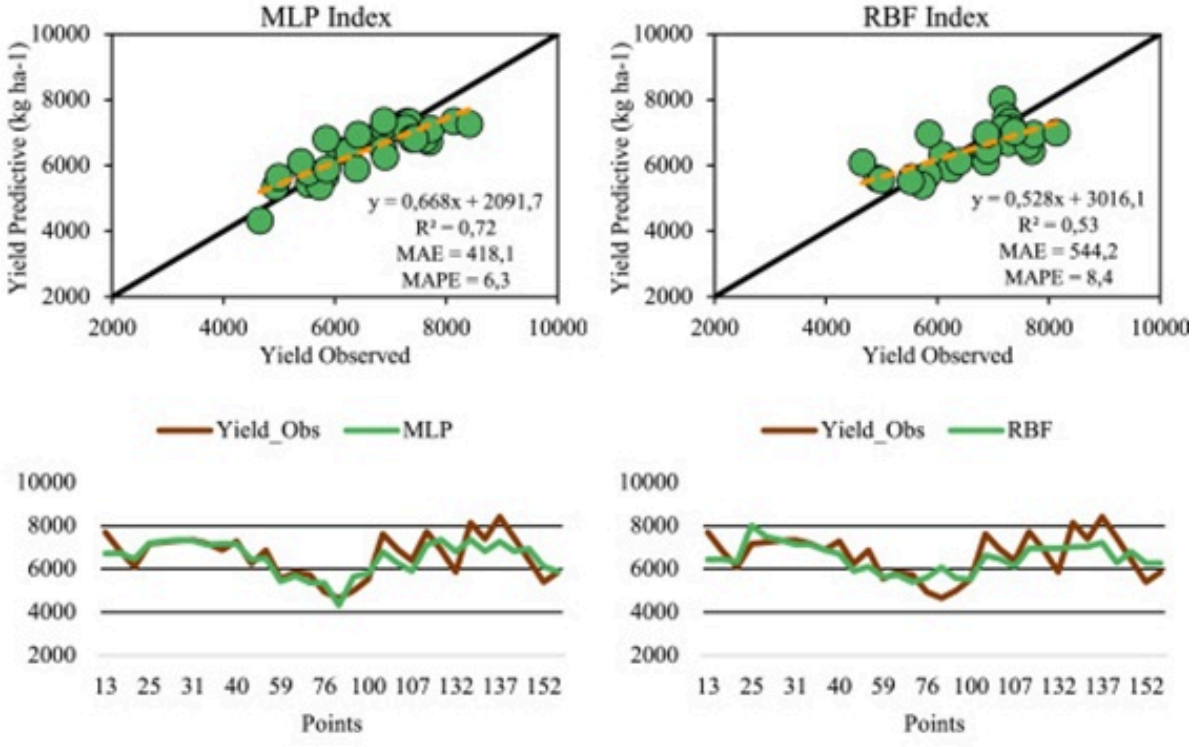
Stepping back to survey the full picture, the weight of evidence indicates that the MLP variant is the more resilient and operationally suited candidate for groundnut yield projection. This verdict holds regardless of whether the underlying training set comprised a single season or a two-season amalgamated dataset. RBF, though not without merit, has inherent constraints that diminish its appeal compared to MLP. These patterns underscore the pivotal role of judiciously selecting a modelling paradigm in achieving an optimal balance between predictive exactness and practical utility in precision farming **(Figure 6)**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

30 DBH

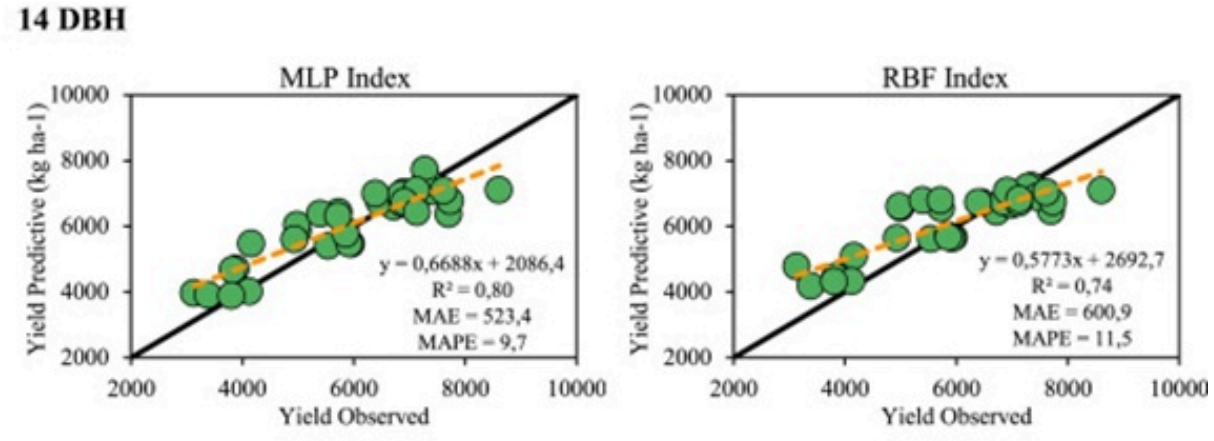


20 DBH



324

© 2023 The Author(s). | Published by Agro Intelligence Research Press



configuration: SR commanded the highest sensitivity ranking 30 days before digging, only for EVI to gain influence at the 20-DBH mark. Nearing harvest, NDVI proved most consequential at 14 DBH; MNLI took precedence at 8 DBH; and EVI once more claimed the top sensitivity spot, a single day before lifting commenced.

What this dissection reveals is a landscape of shifting index relevance—far from static, the importance attached to each VI oscillates markedly between the 30-DBH window and harvest itself, and diverges further depending on which algorithmic framework is employed, mirroring tendencies already witnessed in the single-season models built for 2020/21 **(Figure 3)**. Notably, a subset of indices—NLI, SAVI, NDVI, MNLI, and SR—were recurrently flagged as consequential across both MLP and RBF implementations aimed at groundnut yield estimation. The accumulated evidence also positions EVI and SR as particularly durable, maintaining elevated sensitivity across nearly the full span of pre-harvest assessment windows. That the infusion of a second season's worth of observations strengthens the case for periodic model recalibration as a pathway to more trustworthy forecasts is a conclusion the data firmly support **(Table 4)**.

Table 4. Sensitivity analysis for determining important variables within the models generated for MLP and RBF algorithms from 2020 to 2022.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Algorithm MLP

VI	Days to harvest				
	30 DBH	20 DBH	14 DBH	8 DBH	1 DBH
EVI	1	4	4	3	4
GNDVI	4	5	7	5	
MNLI	3		6	2	3
NDVI		2	2	1	
NLI		3	5		1
SAVI			3	4	
SR	2	1	1		2

Algorithm RBF

Bands e VI	Days to harvest				
	30 DBH	20 DBH	14 DBH	8 DBH	1 DBH
EVI	4	1	7	3	1
GNDVI	7	2	5	4	7
MNLI	3	5	3	1	2
NDVI	6	4	1	7	4
NLI	5	6	2	5	6
SAVI	2	7	4	2	3
SR	1	3	6	6	5

Influence of soil and cultivar on yield prediction

With specific reference to groundnut production, the nature of the soil profile exerts a first-order control over yield dispersion, a circumstance largely rooted in the crop's peculiar growth habit. Groundnut plants generally flourish in substrates endowed with a greater proportion of sand owing to the superior drainage such soils afford—an advantage clay-rich profiles cannot match, often resulting in restricted pod set and consequent yield depression. Yet this very textural property cuts both ways: the moisture storage capacity of lighter soils can become disadvantageous when rainfall becomes scarce, suppressing groundnut productivity [19].

Evidence from the present inquiry indicates discernible differences in both yield spread and central tendency across soil categories, a signal that this factor should not be overlooked. Any effort to upgrade model performance might therefore profit from the inclusion of predictor variables encoding soil attributes. This enrichment could equip the algorithm to resolve yield differences traceable to edaphic properties [20].

Among the fields evaluated here, IAC OL3 consistently outperformed IAC 503. These two materials diverge on several traits, with growth cycle duration standing out as a key point of contrast. The OL3 variety is known to deliver higher yields than genotypes with protracted development [21]. Zooming out, the genotype-by-environment interplay adds layers of intricacy to the yield forecasting enterprise, a complexity compounded by the patchwork of within-field variability driven by an assortment of internal and external influences on the cropping system. Isolating which precise drivers bear most directly on groundnut yield heterogeneity, therefore, becomes an essential prerequisite for further sharpening predictions and associated management prescriptions [22].

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Influence of vegetation indices on predictive models

The preponderance of scholarly efforts to develop yield projection tools for cereals, root and tuber crops, and pulse species relies on remotely sensed inputs, with meteorological models the next most prevalent approach [11]. The present dataset confirms that each index introduced into the predictive workflow was selected at some point. The sensitivity profiling underscores the fluid nature of vegetation index relevance—their importance ebbs and flows with the prediction horizon and shifts depending on whether an MLP or RBF framework is adopted. Such temporal and model-dependent variation argues for a deliberate, stage-aware approach to variable selection when constructing predictive pipelines.

Consider EVI, which was purpose-built to advance vegetation tracking by tempering sensitivity to dense canopies and curbing atmospheric artefacts through its blue band component [23, 24]. Set beside the ubiquitous NDVI, EVI shows greater reactivity to canopy architectural features—leaf area index, stand structure, and background soil effects—whereas NDVI tracks chlorophyll abundance [25]. This property makes EVI less susceptible to saturation. In the current body of work, EVI eclipsed NDVI in the sensitivity hierarchy, reaffirming its suitability for crops that develop dense foliar profiles, groundnut being a case in point.

GNDVI constitutes another VI endowed with distinctive behaviour, displaying amplified responsiveness to chlorophyll dynamics relative to indices anchored on the red spectral region. Those vegetation indices that leverage the green waveband tend to correlate tightly with plant nitrogen economy, thereby offering a window into nutritional stress [26-29].

The coordinated deployment of these vegetation indices within yield forecasting frameworks carries substantial potential, resting on the premise that all the variables in question are, through one pathway or another, entangled with crop productivity. Oliveira *et al.* [30] provided evidence that vegetation indices (VIs), including NLI, NDVI, SAVI, GNDVI, and MNLI, harbour the capacity to track not only groundnut maturation but also the fluctuation of pod biomass across the growing season, underscoring their suitability for mapping spatiotemporal variation in pod yield. The findings reported here lend further weight to the connection that may exist between VIs and groundnut crop output.

Effectiveness of ANN models in predicting agricultural yield

The ANN algorithms used in this investigation demonstrated their utility for groundnut yield estimation, with notable aptitude for reproducing the variation inherent in ground-referenced data. Previous research has identified both MLP and RBF networks as capable predictors of groundnut ripening [12]. With respect to algorithmic prowess specifically directed at yield forecasting, a sizable body of published work indicates that MLPs tend to produce favourable outcomes and enjoy broad application for agricultural output prediction [2-4, 31].

The present inquiry reveals that RBF-derived formulations lagged somewhat in predictive performance; however, the same ANN class has, under different circumstances, surpassed MLPs in predicting groundnut maturity and soybean productivity, as documented in earlier studies [12, 32]. Those prior accounts, considered alongside the evidence assembled here, collectively illuminate the promise of both algorithmic families for forecasting the yields of various agricultural commodities. Beyond this, when pitted against a suite of alternative machine learning approaches (KNN, SVR, RF, and AB), ANN-based prediction of aubergine yield achieved superior results, with vegetation indices as the independent variables [6]. This finding speaks to the robustness of artificial neural networks in predictive contexts.

Insofar as reliability and steadiness of yield forecasts are concerned, MLP outperformed RBF at every assessment window. However, it must be noted that no meaningful improvement in accuracy materialised as the harvest date approached—a pattern that echoes the behaviour of the single-season models built for 2020/21. One plausible interpretation of this observation is the progressive narrowing of uncertainty as the lifting date nears, suggesting that forces beyond the mere countdown shape the outcomes to be harvested.

Transfer learning and applicability of prediction models

© 2023 The Author(s). | Published by Agro Intelligence Research Press

For any yield-forecasting tool to be deemed effective, precise, and practically usable, both the formulation of the predictive equation and its subsequent testing on agricultural parcels not involved in the developmental phase must be undertaken. A major hurdle lies in the ability of such forecasting approaches to adjust to year-to-year weather shifts, contrasting soil profiles, and differing on-farm practices, which collectively render the portability of models a non-trivial pursuit.

Only a handful of contributions within the academic record go the extra step of not merely documenting the fitted functions but also verifying their performance in previously unseen settings, a step that confirms the resilience of the methodological proposition. The work of Becker-Reshef *et al.* [33] drew on empirically calibrated regression expressions. It tested the resulting functions beyond their original political boundaries—spanning countries and states—for the forecasting of wheat and maize output using VIs. A caveat must be inserted here, nonetheless: formulations rooted in straightforward empirical linkages suffer from restricted predictive power when confronted with conditions unlike those under which they were tuned; in simpler terms, their reliability evaporates once moved outside the calibration envelope [34].

Impact of recalibrating models with expanded data

A pronounced increase in both predictive exactness and accuracy was observed when the models were trained on an enlarged dataset, as our findings show. With every additional wave of field observations ingested, the parameter values and trait representations that condition yield undergo revision, and this cyclical retraining, driven by incoming data, does not merely boost model accuracy—it also enriches the grasp that ANNs acquire of the complex interplay among a host of agronomic and environmental variables.

The question of what constitutes an adequate volume of training observations for predictive modelling continues to attract vigorous debate across the research community. In agronomic science, the scarcity of large datasets—the so-called “small data” bottleneck—primarily stems from the practical difficulty of assembling ground-reference measurements. This process often hinges on substantial human labour [1] and, in some instances, can make the acquisition of truly large databases physically unfeasible.

That it is now possible to engineer peanut yield forecasting models that are at once robust and readily portable to unfamiliar locations, while generating projections a full 30 days before the scheduled harvest, constitutes a significant step forward for the body of scientific knowledge surrounding this crop. The central thrust of this paper is the design of predictive models defined by their hardiness, straightforward transferability, and capacity to supply early yield intelligence. To that end, the objective was to craft models that confer reliability and precision without requiring either an exhaustive set of predictor variables or an impractically large training corpus.

Conclusion

Pairing RS-derived information with artificial intelligence techniques demonstrated substantial power to anticipate and to extrapolate ground-reference measurements of peanut productivity across space. Among the algorithms evaluated, MLP emerged as the top performer, providing the strongest predictive fits. At the same time, EVI was identified as the most influential VI driving these projections, with SAVI and SR following in importance.

The data show that peanut yield can be predicted up to 30 days in advance through the synergistic use of artificial neural networks and vegetation indices. When the calibrated functions were exported to other production sites, a meaningful ability to extend predictions to new fields was documented; that said, a gap was detected between the outputs generated by the first-season fits and the estimates delivered in those previously unseen areas, a discrepancy particularly pronounced in fields sown with the IAC 503 variety.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The act of retraining the models on pooled multi-season data proved effective, bringing about a meaningful gain in both exactness and accuracy. It follows that inflating the data foundation is a necessary condition for rendering the model progressively more resilient, thereby assisting ANNs in disentangling the effects of management regimes, agronomic influences, and site-specific environmental signatures. Even so, the door should remain open to the integration of further independent variables with a known bearing on crop output—climatic data and soil descriptors chief among them—a direction that can be pursued in subsequent studies to probe how such factors affect the accuracy of peanut yield forecasting.

Abbreviations and nomenclature

RS, Remote Sensing; ANN, Artificial Neural Networks; VI, Vegetation Indices; RBF, Radial Basis Function; MLP, Multilayer Perceptron; RPA, Remotely Piloted Aircraft; GNSS, Global Navigation Satellite System; GPS, Global Positioning System; GLONASS, Global'naya Navigatsionnaya Sputnikovaya Sistema; BeiDou, Beidou Navigation Satellite System; DPC, Days to Harvest; PS, PlanetScope; TOA, Top of Atmosphere Reflectance; BOA, Bottom of Atmosphere Reflectance; NDVI, Normalized Difference Vegetation Index; GNDVI, Green Normalized Difference Vegetation Index; MNLI, Modified Non-Linear Index; EVI, Enhanced Vegetation Index; NLI, Non-Linear Index; SAVI, Soil-Adjusted Vegetation Index; SR, Simple Ratio; IPS, Intelligent Problem Solver; REM, Run Existing Model; MAE, Mean Absolute Error; MAPE, Mean Absolute Percentage Error; R^2, Coefficient of Determination; IQRs, Interquartile Ranges

Acknowledgements

This Project was funded by the São Paulo Research Foundation (FAPESP - grants no. 2023/14041-2, 2022/16084-8, and 2021/06029-7, and the National Council for Scientific and Technological Development (CNPq) for their financial support.

Conflict of interest

Jarlyson Brunno Costa Souza's report was provided by São Paulo State University Faculty of Agricultural and Veterinary Sciences. Jarlyson Brunno Costa Souza reports a relationship with the São Paulo State University Faculty of Agricultural and Veterinary Sciences, including funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Financial support

None

Ethics statement

None

Received: 15 Dec 2021 Revised: 18 Feb 2022 Accepted: 07 Mar 2022
Published online: 25 May 2023

Rights and permissions

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Safonova A, Ghazaryan G, Stiller S, Main-Knorn M, Nendel C, Ryo M. Ten deep learning techniques to address small data problems with remote sensing. Int J Appl Earth Obs Geoinf. 2023;125:e103569.

2. Farjam A, Omid M, Akram A, Fazel Niari Z. A neural network based modeling and sensitivity analysis of energy inputs for predicting seed and grain corn yields. J Agric Sci Technol. 2014;16(4):767-78.

3. Khalifani S, Darvishzadeh R, Azad N, Seyed Rahmani R. Prediction of sunflower grain yield under normal and salinity stress by RBF, MLP and CNN models. Ind Crops Prod. 2022;189:e115762.

4. Niedbała G. Simple model based on artificial neural network for early prediction and simulation winter rapeseed yield. J Integr Agric. 2019;18(1):54-61.

5. Son NT, Chen CF, Chen CR, Minh VQ, Trung NH. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. Agric For Meteorol. 2014;197:52-64. https://doi.org/10.1016/j.agrformet.2014.06.007.

6. Taşan S, Cemek B, Taşan M, Cantürk A. Estimation of eggplant yield with machine learning methods using spectral vegetation indices. Comput Electron Agric. 2022;202:e107367. https://doi.org/10.1016/j.compag.2022.107367.

7. Shammi SA, Meng Q. Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. Ecol Indic. 2021;121:e107124.

8. Desloires J, Ienco D, Botrel A. Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. Comput Electron Agric. 2023;209:e107807. https://doi.org/10.1016/j.compag.2023.107807.

9. Taki M, Rohani A, Soheili-Fard F, Abdeshahi A. Assessment of energy consumption and modeling of output energy for wheat production by neural network (MLP and RBF) and Gaussian process regression (GPR) models. J Clean Prod. 2018;172:3028-41. https://doi.org/10.1016/j.jclepro.2017.11.107.

10. Oliveira MFde, Santos AFdos, Kazama EH, Rolim GdeS, Silva RPda. Determination of application volume for coffee plantations using artificial neural networks and remote sensing. Comput Electron Agric. 2021;184:e106096. https://doi.org/10.1016/j.compag.2021.106096.

11. Basso B, Liu L. Seasonal crop yield forecast: methods, applications, and accuracies. Adv Agron. 2019;154:201-55. https://doi.org/10.1016/bs.agron.2018.11.002.

12. Souza JBC, Almeida SLHde, Oliveira MFde, Santos AFdos, Filho ALdeB, Meneses MD, Silva RPda. Integrating satellite and UAV data to predict peanut maturity upon artificial neural networks. Agronomy. 2022;12(7):e1512. https://doi.org/10.3390/agronomy12071512.

13. Rouse JW Jr, Haas RH, Schell JA, Deering DW. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec Publ. 1974;351:309.

14. Gitelson AA, Merzlyak MN. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. J Plant Physiol. 1996;148:494-500.

15. Gong P, Pu R, Biging GS, Larrieu MR. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. IEEE Trans Geosci Remote Sens. 2003;41:1355-62.

16. Justice CO, Vermote E, Townshend JRG, Defries R, Roy DP, Hall DK, et al. The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. IEEE Trans Geosci Remote Sens. 1998;36:1228-49. https://doi.org/10.1109/36.701075.

17. Goel NS, Qin W. Influences of canopy architecture on relationships between various vegetation indices and LAI and fPAR: a computer simulation. Remote Sens Rev. 1994;10:309-47.

18. Huete AR. A soil-adjusted vegetation index (SAVI). Remote Sens Environ. 1988;25:295-309.

19. Woli P, Paz JO, Hoogenboom G, Garcia y Garcia A, Fraisse CW. The ENSO effect on peanut yield as influenced by planting date and soil type. Agric Syst. 2013;121:1-8.

20. Kross A, Znoj E, Callegari D, Kaur G, Sunohara M, Lapen DR, et al. Using artificial neural networks and remotely sensed data to evaluate the relative importance of variables for prediction of within-field corn and soybean yields. Remote Sens. 2020;12(14):e2230. https://doi.org/10.3390/rs12142230.

21. De Godoy IJ, Dos Santos JF, De Carvalho CRL, Michelotto MD, Bolonhezi D, De Freitas RS, et al. IAC OL 3 and IAC OL 4: new Brazilian peanut cultivars with the high oleic trait. Crop Breed Appl Biotechnol. 2014;14(3):200-3. https://doi.org/10.1590/1984-70332014V14N3A30.

22. Hara P, Piekutowska M, Niedbała G. Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data. Land. 2021;10(6):e609. https://doi.org/10.3390/land10060609.

23. Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ. 2002;83(1-2):195-213. https://doi.org/10.1016/S0034-4257(02)00096-2.

24. Kaufman YJ, Tanré D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Trans Geosci Remote Sens. 1992;30(2):261-70. https://doi.org/10.1109/36.134076.

25. Gao X, Huete AR, Ni W, Miura T. Optical-biophysical relationships of vegetation spectra without background contamination. Remote Sens Environ. 2000;74(3):609-20. https://doi.org/10.1016/S0034-4257(00)00150-4.

26. Chang Q, Xiao X, Jiao W, Wu X, Doughty R, Wang J, et al. Assessing consistency of spring phenology of snow-covered forests as estimated by vegetation indices, gross primary production, and solar-induced chlorophyll fluorescence. Agric For Meteorol. 2019;279:e107707. https://doi.org/10.1016/j.agrformet.2019.06.002.

27. Tong A, He Y. Estimating and mapping chlorophyll content for a heterogeneous grassland: Comparing prediction power of a suite of vegetation indices across scales between years. ISPRS J Photogramm Remote Sens. 2017;126:146-67.

28. Viña A, Gitelson AA, Nguy-Robertson AL, Peng Y. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. Remote Sens Environ. 2011;115:3468-78. https://doi.org/10.1016/j.rse.2011.08.010.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

29. Zhao B, Duan A, Ata-Ul-Karim ST, Liu Z, Chen Z, Gong Z, et al. Exploring new spectral bands and vegetation indices for estimating nitrogen nutrition index of summer maize. Eur J Agron. 2018;93:113-25.

30. Oliveira MF, Carneiro FM, Ortiz BV, Thurmond M, Oliveira LP, Bao Y, et al. Predicting below and above-ground peanut biomass and maturity using multi-target regression. Comput Electron Agric. 2024;218:108647.

31. Niedbała G, Nowakowski K, Rudowicz-Nawrocka J, Piekutowska M, Weres J, Tomczak RJ, et al. Multicriteria prediction and simulation of winter wheat yield using extended qualitative and quantitative data based on artificial neural networks. Appl Sci. 2019;9(14):2773.

32. Yoosefzadeh-Najafabadi M, Tulpan D, Eskandari M. Application of machine learning and genetic optimization algorithms for modeling and optimizing soybean yield using its component traits. PLoS One. 2021;16(4):e0250665.

33. Becker-Reshef I, Vermote E, Lindeman M, Justice C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens Environ. 2010;114(6):1312-23.

34. Doraiswamy PC, Moulin S, Cook PW, Stern A. Crop yield assessment from remote sensing. Photogramm Eng Remote Sens. 2003;69:665-74.
https://doi.org/10.14358/PERS.69.6.665.

© 2023 The Author(s). | Published by Agro Intelligence Research Press