Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Developing a farm-level productivity indicator for coffee and cocoa via NDVI time series and National Agricultural Survey data in Peru

Robinson López-Monzon
Ministry of Agriculture of Peru
José Antonio Otoya-Barrenechea
Ministry of Agriculture of Peru
Juan Leonardi Romero-Vasquez
Ministry of Agriculture of Peru
Diana Carolina Quintero-Giraldo
diana.quintero29@gmail.com

Ministry of Agriculture of Peru

Research Article

Keywords: remote sensing, yield mapping, multispectral time series, ensemble models, official statistics, tropical perennial crops

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9464917/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Context: Farm-level productivity indicators are still limited in Peruvian agricultural statistics, particularly for smallholder perennial crops such as coffee and cocoa, where strong within-region heterogeneity and complex management conditions make productivity difficult to measure consistently. Moreover, increasing access to satellite imagery creates new opportunities to complement survey-based agricultural information with spatially explicit indicators.

Aims: This study aimed to develop and validate a farm-level indicator of productive capacity for smallholder coffee and cocoa systems in Peru by integrating National Agricultural Survey (ENA) microdata with Sentinel-2 normalized difference vegetation index (NDVI) time series.

Methods: ENA 2023–2024 microdata were linked to georeferenced farm polygons and cloud-filtered 10 m Sentinel-2 normalized difference vegetation index (NDVI) imagery aggregated into 12 bimonthly composites. NDVI snapshots and temporal features were derived and used first in ridge and random forest regression models for continuous yield prediction and then in gradient boosting ensemble models for ordinal classification into low-, medium-, and high-productivity classes. A simplified decision-tree indicator was subsequently derived for operational use.

Key Results: Continuous-yield regression performed poorly ($R^2 \approx 0$; RMSE ≈ 340 kg ha^{-1}), indicating that the NDVI alone was insufficient for reliable parcel-level yield estimation. In contrast, ordinal classification performed well, reaching 0.73 validation accuracy, 0.85 evaluation accuracy, and Cohen's $\kappa = 0.75$. The simplified indicator achieved lower but still useful performance (56.7% accuracy; $\kappa = 0.31$).

Conclusion: Survey microdata and NDVI time series can be combined to generate an operational, farm-level indicator of productive capacity for smallholder perennial systems, even when continuous yield prediction remains weak.

Implications and Impacts: The proposed indicator provides a practical way to translate satellite time series and survey microdata into an operational tool for official statistics. In coffee and cocoa systems, it can support stratified sampling, improve the review of anomalous survey records, and guide more spatially explicit interventions, with potential relevance for other contexts after local calibration.

KEYPOINTS/HIGHLIGHTS

- This study develops and validates a farm-level indicator of productive capacity for smallholder coffee and cocoa systems in Peru by integrating National Agricultural Survey (ENA) microdata with Sentinel-2 normalized difference vegetation index (NDVI) time series data.
- The results show that NDVI-based models are insufficient for reliable, continuous yield prediction at the parcel level, as regression performance remained weak ($R^2 \approx 0$; RMSE ≈ 340 kg ha^{-1}).
- Reframing the problem as an ordinal classification into low-, medium-, and high-productivity categories substantially improved model performance, reaching 0.73 validation accuracy, 0.85 evaluation accuracy, and Cohen's $\kappa = 0.75$.
- Although less accurate than ensemble models, the simplified decision tree indicator offers a more transparent and operationally feasible tool for implementation in agricultural statistical systems.
- The proposed indicator is a practical tool for supporting agricultural decision-making, particularly in improving sampling strategies, strengthening data quality control, and helping target extension and technical assistance more effectively.

IMPACT: This study demonstrates that Sentinel-2 NDVI time series and official agricultural survey microdata can be combined to produce an operational indicator of productive capacity for smallholder coffee and cocoa systems in Peru. While developed with the microdata of the National Agricultural Survey (ENA) of Peru, the methodology also offers a potentially useful framework for other contexts seeking to make spatial variability actionable for sampling design, record review, and targeted support, subject to local calibration and validation.

INTRODUCTION

Agriculture remains a strategic sector for Peru, both for its contribution to the gross domestic product (GDP) and for its role in employment, territorial cohesion and rural livelihoods (World Bank, 2023, 2024). However, the lack of objective and comparable indicators of productivity at the farm level limits the capacity to target public policies, technical assistance and financial services to the areas and producers where they are most needed (World Bank, 2017; Gil et al., 2019). In practice, most decisions are still based on averages calculated at the regional or domain level, which masks the strong microterritorial heterogeneity observed within the same crop and year.

Peru has made significant progress in agricultural statistics through the National Agricultural Survey (ENA), which provides key information on area, production, yields and inputs for major crops (MIDAGRI, 2022). Nevertheless, estimating productive capacity at the farm level remains challenging, especially in perennial and smallholder-dominated systems such as coffee and cocoa, where plant density, age structure, management and environmental conditions vary widely even within small areas.

In parallel, the availability of freely accessible satellite imagery and cloud-computing platforms has opened new possibilities for the use of remote sensing to monitor crop vigor and yield with high spatial and temporal resolutions (Flores-Anderson et al., 2023; Wuepper et al., 2025). Vegetation indices derived from multispectral images, particularly the normalized difference vegetation index (NDVI), have been consistently associated with crop development and yield in a wide range of systems (Groten, 1993; Labus et al., 2002; Wang et al., 2005; Fuentes et al., 2025). In coffee and cocoa, recent studies have shown that the NDVI and related indices can help characterize yield variability, stress patterns and climatic gradients (Bernardes et al., 2012; Nogueira et al., 2018; Martello et al., 2022; Atalaya-Marín et al., 2024; Íñiguez Freites et al., 2025).

However, the application of NDVI-based models in official operational statistics poses specific challenges. First, yield information in surveys such as ENA is collected through farmer-reported production and area, which may introduce measurement error and bias (FAO, 2017). Second, perennial crops exhibit complex phenological patterns and alternate bearing and management cycles that are not easily summarized by single-date NDVI snapshots (Anyimah et al., 2021; Manoel et al., 2024). Third, statistical offices need simple and interpretable indicators that can be integrated into existing workflows rather than black-box models that are difficult to communicate and maintain (Wuepper et al., 2025).

The objective of this study is to develop and validate an annual farm-level indicator of productive capacity for smallholder coffee and cocoa systems in Peru through the integration of ENA microdata and Sentinel-2 NDVI time series. To achieve this goal, a harmonized database linking individual farms with satellite-derived NDVI metrics is first constructed; the capacity of these time series to predict continuous yield is then evaluated; then, ordinal classification models are developed to group farms into low-, medium-, and high-productivity categories; and finally, a simple and interpretable decision-rule indicator is derived for potential use within the Peruvian agricultural statistical system.

MATERIALS AND METHODS

Study area and data sources

This study focused on smallholder coffee and cocoa systems located in the eastern Andean slopes of Peru, particularly in the regions of San Martín, Amazonas and Ucayali, which constitute a large share of national production.

Three main data sources were integrated. First, *microdata from the National Agricultural Survey* (ENA 2023–2024) was used, as this survey provides farm-level information on cultivated area, production, management practices, and structural characteristics (INEI, 2025). For coffee and cocoa, plot-level data on harvested area (ha) and production (kg) were extracted, and yields (kg ha^{-1}) were calculated as the ratio of production to area. Records with missing or inconsistent values for area or production were excluded in accordance with ENA quality-control criteria. Second, *georeferenced polygons* corresponding to each sampled farm were obtained from ENA and used to aggregate satellite-derived NDVI data at the farm level. Third, *NDVI time series* were generated from Sentinel-2 Level-2A surface reflectance products at a 10 m spatial resolution, using bands 4 (red) and 8 (near-infrared). These data were processed for the 2023–2024 agricultural season over a 12-month period and aligned with ENA yield reporting. Cloud- and cloud shadow-contaminated pixels were removed via quality flags and conservative thresholds, and the remaining observations were summarized into 12 bimonthly normalized difference vegetation index (NDVI) composites by calculating the median number of cloud-free pixels (Flores-Anderson et al., 2023).

Data integration and analytical sample

For each sampled farm, the farm polygon was overlaid on the series of NDVI composites, and zonal statistics were calculated via the median NDVI value at each bimonthly interval. Farms with very small coffee or cocoa cultivation areas relative to the total polygon, as well as those with insufficient cloud-free observations, were excluded from the analysis. After applying all the consistency checks, the final analytical sample included several hundred coffee and cocoa farms. In the case of coffee, the yield distribution was strongly skewed, with most observations concentrated below 500 kg ha^{-1} and relatively few above 1,000 kg ha^{-1}, which is consistent with previous evidence on smallholder productivity in the Peruvian Amazon (World Bank, 2017; Morales et al., 2022). For cocoa, yields were more concentrated at approximately 600 kg ha^{-1}, with very few observations exceeding 1,500 kg ha^{-1}.

These distributions are illustrated by histograms of coffee (Fig. 1) and cocoa yields (Fig. 2), which show a marked right-skewness and a small number of high-yield outliers.

The planting density distributions derived from the ENA confirmed substantial deviations from the theoretical or recommended densities; thus, the plant density was excluded as a direct predictor.

Feature Engineering

Two families of NDVI-based features were constructed. The first consisted of *snapshot metrics*, defined as the NDVI values observed during specific bimonthly periods corresponding to key phenological stages of coffee and cocoa, such as vegetative growth, flowering, and grain filling. For each crop, these periods were identified on the basis of agronomic calendars and local expert knowledge, and the median NDVI value of each farm polygon was retained for the selected dates.

The second family consisted of *derived temporal features* calculated from the full 12-step NDVI series. These included the annual maximum and minimum NDVI, the maximum minus minimum (NDVI) amplitude, the coefficient of variation, the interquartile range, the selected percentiles, and the cumulative NDVI, which were used as approximations of the area under the NDVI curve over the agricultural season.

The full set of features was standardized and screened to avoid redundant predictors and reduce collinearity. A correlation matrix of NDVI-derived features was used to guide feature selection and aggregation, and a feature importance ranking from preliminary models helped identify the most informative predictors.

Table 1 provides a detailed description of all the dependent, explanatory and control variables used in the analysis, including their roles, sources (ENA vs. Sentinel-2) and units.

Table 1
Definitions, roles and sources of variables used in the models.

Variable	Role	Operational definition/Source	ENA Fields/Sentinel	Unit	Scale	Procedure
REND	Dependent	Total production/area	P117_SUP_ha; totales/destinos Cap. 200AB (P220_*, en kg)	kg/ha	Ratio	Regression and binning, classification
RENDPLANT	Dependent	REND/PLANTHA	P126 (plantas), P117_SUP_ha	kg/plant	Ratio	Regression; sensitivity to biases
PLANTHA	Control/Explanatory	P126/P117_SUP_ha	ENA	plants/ha	Ratio	Control; quality filters
REGION	Control	Geographical Region	REGION/NOMBREDD	—	Nominal	Fixed effects/blocks
STARTMES	Control	Month when harvest starts	P124_MES	month	Ordinal	Phenological alignment
ndvi_cv_all	Explanatory (global)	std(NDVI)/mean(NDVI) 2023–2024	Sentinel-2 (zonal stats)	adim.	Ratio	Key feature; permutation importance
ndvi_mean_all	Explanatory(global)	Mean NDVI 2023–2024	Sentinel-2	adim.	Interval	Feature; control for baseline status
ndvi_amp_all	Explanatory(global)	max – min NDVI 2023–2024	Sentinel-2	adim.	Interval	Seasonality signal
AUC_main_23/24	Explanatory(annual)	Area under the curve (main window)	Sentinel-2	NDVI· bimonthly	Interval	Within-year phenology
auc_ratio_main_minor_23/24	Explanatory(annual)	AUC_main/AUC_minor	Sentinel-2	adim.	Ratio	Balance of peaks
rise_pre_23/24, fall_harv_23/24, rebound_23/24	Explanatory(annual)	Slopes by window	Sentinel-2	ΔNDVI/ Δbimonthly	Interval	Vigor and recovery
skew_23/24, kurt_23/24	Explanatory(annual)	Shape of NDVI distribution	Sentinel-2	adim.	Interval	Cycle stability
delta_<métrica>	Explanatory(interannual)	2024 – 2023 for each feature	Sentinel-2	according to metric	—	Interannual change
NDVI* (selectas)	Explanatory (snapshot)	NDVI in critical bimonthly periods	Sentinel-2	adim.	Interval	Only if phenologically justified

To document collinearity patterns and guide feature selection, a correlation heatmap of all NDVI-derived variables was computed, which is presented in Fig. 3.

Continuous yield modeling

NDVI time series were evaluated to validate whether continuous yields (kg ha^{-1}) can be predicted at the holding level. Two supervised regression models were trained: (i) *ridge regression*, a linear model with L2 regularization that handles correlated predictors while preserving interpretability; and (ii) *random forest regression*, an ensemble of decision trees that are able to capture nonlinear relationships and interactions.

The dataset was randomly split into training (60%), validation (20%) and evaluation (20%) subsets, which were stratified by yield quantiles. The hyperparameters (e.g., the regularization strength for ridges and the number and depth of trees for the random forest) were tuned via cross-validation on the training and validation sets. Model performance was evaluated on the held-out evaluation set via the root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2).

Table 2 summarizes the performance of the continuous regression models, reporting the RMSE, MAE and R^2 for both ridge and random forest.

Table 2
Performance metrics (RMSE, MAE, R^2) of the ridge and random forest regression models for predicting continuous yields

Método	RMSE	MAE	R2
Ridge	339.4397	243.262	-0.0524
RandomForest	346.477	235.234	-0.0965

Ordinal classification of productive capacity

Given the weak performance of the continuous models, the problem was reformulated as an *ordinal classification task* with three mutually exclusive categories of productive capacity: low, medium, and high, as illustrated in Fig. 4. Cutoff points were established via ENA yield data and agronomic criteria so that each category would represent meaningful productivity thresholds for policy and extension purposes, including very low yields well below regional averages, intermediate yields, and top-performing farms.

Gradient boosting ensemble models were then trained via CatBoost and LightGBM, both of which are well suited to structured tabular data. Multiclass and ordinal classification approaches were tested. The same training, validation, and evaluation strategies were applied, and model performance was assessed via overall accuracy, class-specific precision and recall, and Cohen's κ to account for agreement occurring by chance. Confusion matrices were also examined to identify misclassification patterns, particularly between the medium- and high-productivity categories.

Table 3 presents the main performance indicators of the best ensemble models and the simplified indicator, including overall accuracy, balanced accuracy and κ.

Table 3
Performance of the ensemble ordinal classifiers and simplified indicator (overall accuracy, balanced accuracy and Cohen's κ).

#	model	score_val	eval_metric
0	CatBoost_r177_BAG_L2	0.727	accuracy
1	WeightedEnsemble_L3	0.723	accuracy
2	LightGBM_r131_BAG_L2	0.723	accuracy
3	CatBoost_r9_BAG_L2	0.722	accuracy
4	LightGBM_BAG_L2	0.722	accuracy
5	CatBoost_BAG_L2	0.771	accuracy
6	LightGBMXT_BAG_L2	0.712	accuracy
7	LightGBMLarge_BAG_L2	0.707	accuracy

The confusion matrix (Fig. 4) for the selected three-class ensemble classifier illustrates how often low, medium and high categories are correctly or incorrectly predicted.

Feature dimensionality reduction.

While ensemble models achieve high predictive performance, their complexity poses challenges for operational use within national statistical offices. Therefore, feature correlations were analyzed to select variables with low mutual dependencies, as illustrated in Fig. 5. This procedure reduced the original 54-variable feature space to the 20 most informative predictors.

Gradient boosting ensemble models were trained via CatBoost and LightGBM, both of which are well suited to structured tabular data. Both multiclass and ordinal classification settings were tested. The same training, validation, and evaluation strategies were applied throughout, and model performance was assessed via overall accuracy, class-specific precision and recall, and Cohen's κ to account for agreement by chance. In addition, confusion matrices were examined to identify misclassification patterns, particularly between the medium- and high-productivity categories.

RESULTS

Yield distributions of coffee and cocoa

Coffee yields derived from the ENAI sample showed a markedly right-skewed distribution, with most smallholder farms concentrated below 500 kg ha^{-1} and relatively few observations above 1,000 kg ha^{-1}. Cocoa yields displayed a similar shape but were more concentrated at approximately 600 kg ha^{-1}, with very few farms exceeding 1,500 kg ha^{-1} (Figs. 1–2). These patterns confirm large productivity gaps within the same regions and crops and motivate the construction of an indicator capable of stratifying farms by productive capacity.

Relationship between the NDVI and continuous yield

NDVI dynamics were compared across the agricultural season for two contrasting groups of farms: those with the lowest yields and those with the highest yields in the sample. As shown in Fig. 6, both groups displayed the typical seasonal pattern of coffee and cocoa systems, characterized by an early-season decline followed by recovery; however, differences were observed in both average NDVI levels and temporal stability. Farms with higher yields maintained consistently higher NDVI values, experienced less pronounced troughs, and showed faster recovery, whereas lower-yield farms exhibited deeper declines and slower rebounds.

Exploratory analysis revealed no strong linear relationship between the NDVI metrics and continuous yields at the farm level. Scatter plots of the annual cumulative NDVI and maximum NDVI against yield (Fig. 7) revealed large dispersion and negligible linear trends, suggesting that multiple unobserved factors (management, age, shade, pests) modulate the NDVI–yield relationship.

The ridge regression model achieved RMSE values of 339–340 kg ha^{-1} and a MAE of approximately 240 kg ha^{-1}, with a slightly negative R^2, indicating that the model did not improve upon a naive mean-based prediction. The random forest model yielded similar RMSE and MAE values and likewise failed to capture substantial variance in yields. These results suggest that, in this context, NDVI time series alone are not sufficient to provide reliable continuous yield estimates at the scale of individual smallholder farms.

Performance of ordinal classification models

When yields were grouped into three categories of productive capacity (low, medium and high), model performance improved substantially. Gradient boosting ensemble classifiers trained on NDVI-derived features achieved validation accuracies of approximately 0.72–0.73 across several CatBoost and LightGBM configurations, with the best CatBoost model achieving approximately 0.77. On the held-out evaluation set, the selected three-class ensemble classifier attained overall accuracies of approximately 0.73 (validation) and 0.85 (evaluation), with Cohen's $\kappa \approx 0.75$, indicating substantial agreement beyond chance.

Table 3 presents the main performance indicators of the best ensemble models and the simplified indicator, including overall accuracy, balanced accuracy and κ. Misclassifications occurred mainly between adjacent categories (low vs. medium, medium vs. high), whereas confusion between the lowest and highest classes was rare. Importantly, the models were particularly effective at identifying the lowest-productivity farms, which is a critical segment for targeting public support and extension.

The confusion matrix (Fig. 8) for the selected three-class ensemble classifier illustrates how often low, medium and high categories are correctly or incorrectly predicted.

Performance of the simplified decision-cut indicator

The simplified decision-cut tree, built from a subset of NDVI features and constrained to a shallow depth, achieved an overall accuracy of 56.7% and $\kappa = 0.31$ on the evaluation set (Table 3). While clearly inferior to the ensemble models, this indicator still performed substantially better than random assignment and retained reasonable sensitivity for identifying low-yield farms.

The final tree contains a small number of rules that combine thresholds on the NDVI in one or two critical windows with the maximum NDVI or NDVI amplitude. For example, farms with persistently low NDVI throughout the season and low maximum NDVI are classified as "low capacity", whereas those that reach high NDVI peaks and maintain moderate variability are classified as "high capacity". Intermediate combinations are assigned to the "medium" category.

From an operational standpoint, this indicator can be implemented using preprocessed NDVI data stored in a central database, without the need for complex machine learning pipelines. It can be recalibrated periodically as new ENA campaigns and satellite data become available, adjusting cutoff points to reflect changes in management, varieties or climate.

DISCUSSION

This study explored the potential of the Sentinel-2 NDVI time series to support the construction of a parcel-level indicator of productive capacity for smallholder coffee and cocoa systems in Peru in combination with official agricultural survey data. The main findings can be summarized in three points. First, the ability of NDVI-based models to predict continuous yields at the individual holding scale was very limited. Second, when yields were grouped into broad categories of productive capacity, ensemble classifiers achieved substantially better performance. Third, a simplified rule-based indicator derived from a small set of NDVI features provided a transparent and operationally feasible approximation, albeit with lower accuracy than the full ensemble models.

Our results confirm that the NDVI time series is a useful but incomplete proxy for productive capacity at the farm level in perennial smallholder systems. The lack of strong predictive power for continuous yields is consistent with recent meta-analyses showing that the NDVI–yield relationship weakens at fine spatial scales and in complex systems with strong management and environmental heterogeneity (Fuentes et al., 2025; Wuepper et al., 2025). In coffee and cocoa, factors such as tree age structure, pruning cycles, shade composition and pest and disease incidence can significantly affect yields without necessarily being fully captured by the NDVI.

Nonetheless, the capacity of NDVI-derived features to support ordinal classification into meaningful categories of productivity suggests that they encode enough information to distinguish broad performance strata. This approach fits with the way in which satellite data have been used to monitor stress,

canopy vigor and broad-scale yield patterns in coffee and cocoa landscapes (Anyimah et al., 2021; Ashiagbor et al., 2020; Atalaya-Marín et al., 2024; Martello et al., 2022).

For national statistical systems, the value of the proposed indicator lies not in replacing survey-based yield estimation but in complementing it. By combining ENA microdata with NDVI time series, statistical agencies can construct an external reference that helps design stratified sampling frames that explicitly incorporate productivity gradients, identify extreme or inconsistent yield reports that merit further review, and characterize the distribution of productive capacity within and across regions at a finer spatial resolution than is possible with survey data alone.

This aligns with international recommendations to integrate remote sensing into agricultural statistics as a way to improve efficiency, reduce respondent burden and increase spatial resolution (FAO, 2017; Wuepper et al., 2025). In addition, the indicator can inform the design of targeted policies and programmes in priority value chains such as coffee and cocoa, where productivity gaps, climate risks, and deforestation concerns intersect.

Limitations and future research

This analysis has certain limitations. First, it is based on one or a few agricultural seasons; extending the series to multiple campaigns would allow a better characterization of interannual variability, alternate bearing, and climate shocks (García et al., 2022; Manoel et al., 2024). Second, the models rely solely on the NDVI and do not incorporate other potentially informative layers, such as radar backscatter (Sentinel-1), climatic variables, soil properties or topography (Lu & Weng, 2007; Flores-Anderson et al., 2023). Third, yields are derived from farmer-reported data in the ENA, which may contain measurement errors; combining objective plot-level measurements with survey data in subsamples could help calibrate and validate the indicator more robustly. Finally, the study focuses on coffee and cocoa; assessing the transferability of the approach to other crops and agroecological contexts within Peru would be valuable for scaling. Building on these constraints, future work should incorporate multiyear time series and enrich the feature set with complementary sensors and environmental data.

CONCLUSION

This study shows that integrating official agricultural survey microdata with Sentinel-2 normalized difference vegetation index (NDVI) time series enables the construction of a farmer-level indicator of productive capacity for smallholder coffee and cocoa systems in Peru. While NDVI-based models exhibit very limited performance for continuous yield prediction at the farm level, they support robust ordinal classification into low-, medium-, and high-productivity categories when combined with appropriate modeling strategies.

Gradient boosting ensembles achieve high classification performance but are difficult to implement operationally within statistical offices. A simplified decision-cut indicator derived from a small set of NDVI features offers a practical alternative that balances interpretability and predictive capacity. It can be embedded in the national agricultural statistical system to support sampling design, quality control and the targeting of extension and support services.

The proposed indicator represents a first step toward a more integrated, spatially explicit, and data-rich approach to measuring productive capacity in Peruvian agriculture. Over time, expanding and refining this approach could help close productivity gaps, strengthen the resilience of smallholder systems and support more sustainable coffee and cocoa value chains.

Declarations

Competing Interest

The authors declare that they have no competing interests.

Funding

The research leading to these results was funded by the Ministerio de Desarrollo Agrario y Riego del Perú (MIDAGRI) through the project Mejoramiento del Sistema de Información Estadística Agraria y del Servicio de Información Agraria para el Desarrollo Rural del Perú (PIADER), implemented by the Unidad Ejecutora de Gestión de Proyectos Sectoriales. Funding was awarded through the public call Concurso de Investigación sobre la Encuesta Nacional Agraria (Orden de Servicio No. 2672-2025).

Data Availability

ENA data are managed by the National Institute of Statistics and Informatics (INEI) of Peru. ENA public-use microdata can be downloaded from its official portal. Geospatial polygons of sampled farms are not publicly released and can only be accessed under the institution's standard procedures. Sentinel-2 imagery is available through the Copernicus Open Access Hub.

ACKNOWLEDGMENTS

The authors thank the teams of UEGPS (Ministerio de Desarrollo Agrario y Riego del Peru) and INEI for facilitating access to ENA microdata and geospatial information, as well as the PIADER project team for their financial and technical support in the construction of the analytical database. Any remaining errors are the sole responsibility of the authors.

Author Contribution

R.L.-M. contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, resources, software, supervision, validation, visualization, and writing the original draft and reviewing and editing the manuscript. J.A.O.-B. contributed to funding acquisition, supervision, and review and editing. J.L.R.-V. contributed to funding acquisition, supervision, and review and editing. D.Q.-G. contributed to conceptualization, data curation, funding acquisition, investigation, project administration, resources, and writing the original draft and reviewing and editing the manuscript. All authors reviewed the manuscript.

References

1. Anyimah FO, Osei Jnr EM, Nyamekye C, 2021. Detection of stress areas on cocoa farms using GIS and remote sensing: A case study of Offinso Municipal and Offinso North District, Ghana. Environ Challenges 4: 100087.
2. Ashiagbor G, Forkuo EK, Asante WA, Acheampong E, Quaye-Ballard JA, Boamah P, Mohammed Y, Foli E, 2020. Pixel-based and object-oriented approaches in segregating cocoa from forest in the Juabeso-Bia landscape of Ghana. Remote Sens Appl Soc Environ 19: 100349.
3. Atalaya-Marín N, Goñas M, Tineo D, Chuquibala-Checan B, Arce-Inga M, Tarrillo C, Alvarez-Robledo YA, Tafur-Culqui J, Cabrera-Hoyos H, Gómez-Fernández D, et al., 2024. Integrating remote sensing and in situ data to determine climate diversity and variability in cocoa systems in the provinces of Jaén and San Ignacio, Cajamarca (NW Peru). Trees People 19: 100749.
4. Bernardes T, Moreira MA, Adami M, Giarolla A, Rudorff BFT, 2012. Monitoring biennials bearing effect on coffee yield via MODIS vegetation indices. Remote Sens 4(9): 2492-2509.
5. FAO, 2017. Handbook on remote sensing for agricultural statistics. Global Strategy to improve Agricultural and Rural Statistics (GSARS). FAO, Rome, Italy.
6. Flores-Anderson AI, Groom G, Brumby SP, et al., 2023. Spatial and temporal availability of cloud-free optical observations over the tropics. Sci Data 10: 574.
7. Fuentes I, Al-Shammari D, Al-Nasrawi AKM, Wang Y, Wang J, Lebrini Y, Chen Y, Jones BG, Bishop TFA, 2025. The normalized difference vegetation index as an analytic tool for wheat crop yield prediction: A review and meta-analysis. Precis Agric 26: 55.
8. García GM, Perez-Esteban J, Tarquis AM, 2022. Reproductive trade-offs in a perennial crop: Exploring the causes of alternate bearing in coffee. Agric For Meteorol 323: 109049.
9. Gil JDB, Reidsma P, Giller KE, Todman LC, Whitmore AP, van Ittersum MK, 2019.
10. Sustainable development goal 2: Improved targets and indicators for agriculture and food security. Ambio 48(7): 685-698.
11. Groten SME, 1993. NDVI-crop monitoring and early yield assessment of Burkina Faso. Int J Remote Sens 14(8): 1495-1515.
12. Íñiguez Freites EA, Echevarría Zamora E, Ramírez Sánchez A, 2025. Geomorphological, biophysical and photogrammetric analysis of agroforestry systems associated with cacao (Theobroma cacao L.) cultivation in the Dominican Republic. Open Access Libr J 12(7): 1-16.
13. INEI, 2025. Encuesta Nacional Agropecuaria 2023–2024: Metodología y principales resultados. Instituto Nacional de Estadística e Informática, Lima, Peru.
14. Labus MP, Nielsen GA, Lawrence RL, Engel R, Long DS, 2002. Wheat yield estimates using multitemporal NDVI satellite imagery. Precis Agric 3(1): 15-26.
15. Lu D, Weng Q, 2007. A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens 28(5): 823-870.
16. Manoel MC, Rosa MR, Queiroz AP, 2024. Analysis of the biennial productivity of Arabica coffee with Google Earth Engine in the northeast region of São Paulo, Brazil. Remote Sens 16(20): 3833.
17. Martello M, Molin JP, Wei MCF, Canal Filho R, Nicoletti JV, 2022. Coffee-yield estimation using high-resolution time-series satellite images and machine learning. AgriEngineering 4(4): 888-902.
18. Ministerio de Desarrollo Agrario y Riego del Perú (MIDAGRI), 2022. Plan Estadístico Agrario Nacional 2022–2026 del Sistema Integrado de Estadística Agraria. Ministerio de Desarrollo Agrario y Riego, Lima, Peru.
19. Morales LV, Robiglio V, Baca M, Bunn C, Reyes M, 2022. Planning for adaptation: A system approach to understand the value chain's role in supporting smallholder coffee farmers' adaptive capacity in Peru. Front Clim 4: 788369.
20. Nogueira SMC, Moreira MA, Volpato MML, 2018. Relationships between coffee crops productivity and vegetation indices derived from OLI/Landsat-8 sensor data with and without topographic correction. Eng Agric 38(3): 387-394.
21. Wang J, Rich PM, Kettle WD, Price KP, 2005. Relationships between the NDVI and grassland production and crop yield in the Central Great Plains. Geocarto Int 20(3): 5-11.
22. World Bank, 2017. Gaining momentum in Peruvian agriculture: Opportunities to increase productivity and enhance competitiveness. World Bank Group, Washington, DC.
23. World Bank, 2023. World Development Indicators: Agriculture, forestry and fishing, value added (% of GDP) – Peru. https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=PE [10 October 2025].
24. World Bank, 2024. World Development Indicators: Agriculture, forestry and fishing, value added (% of GDP) – Peru. https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS [10 October 2025].
25. Wuepper D, Oluoch WA, Hadi H, 2025. Satellite data in agricultural and environmental economics: Theory and practice. Agric Econ 56(3): 493-511.

Figures

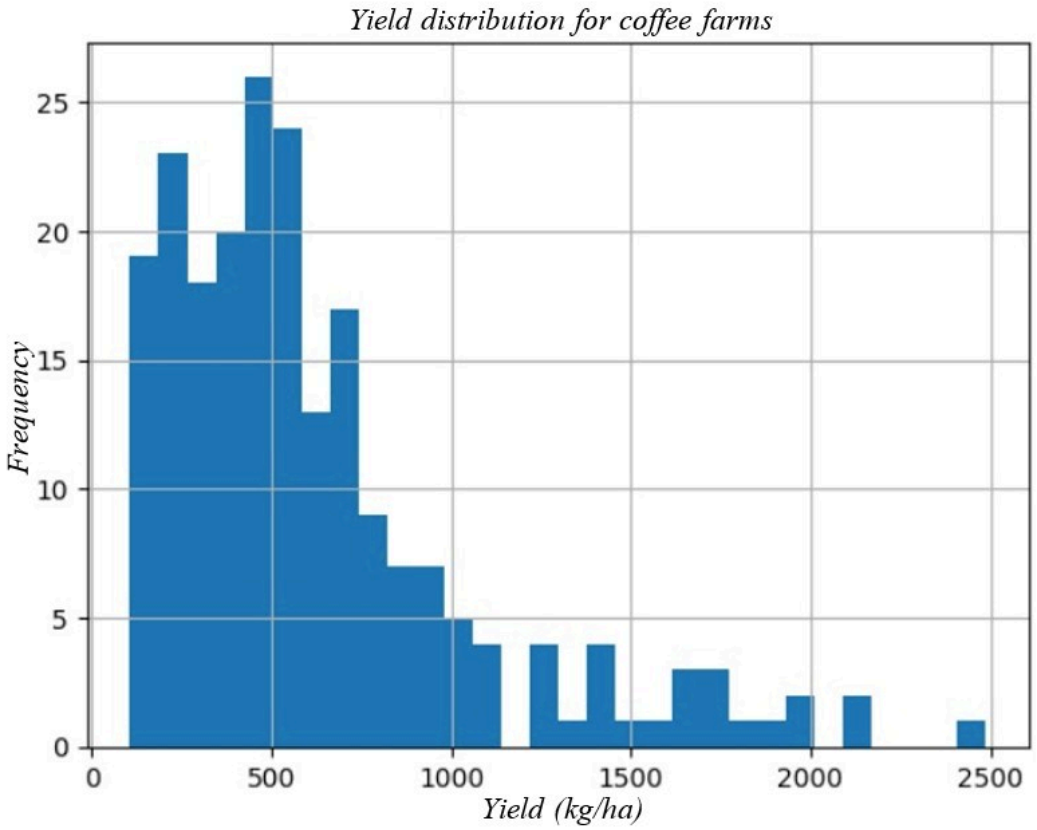


Figure 1

Histogram of coffee yields (ENA sample)

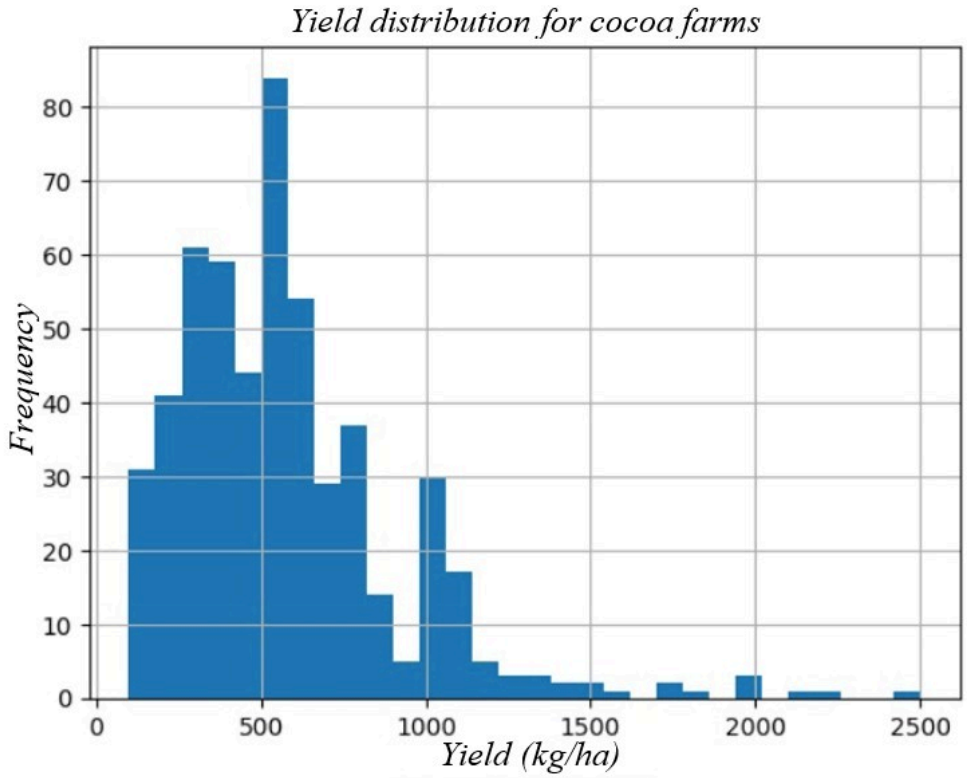


Figure 2

Histogram of cocoa yields (ENA sample)

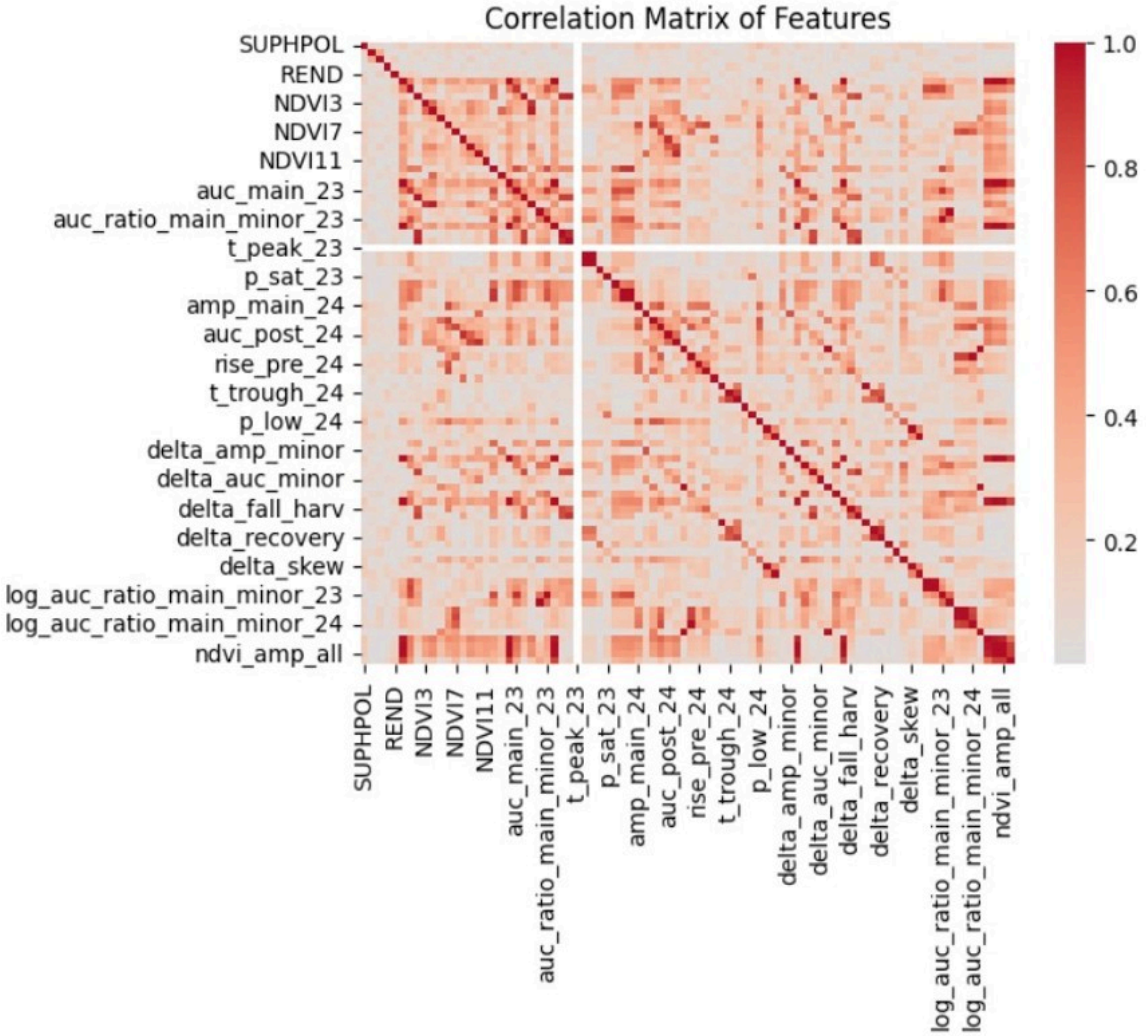


Figure 3

Correlation matrix of the NDVI-derived features

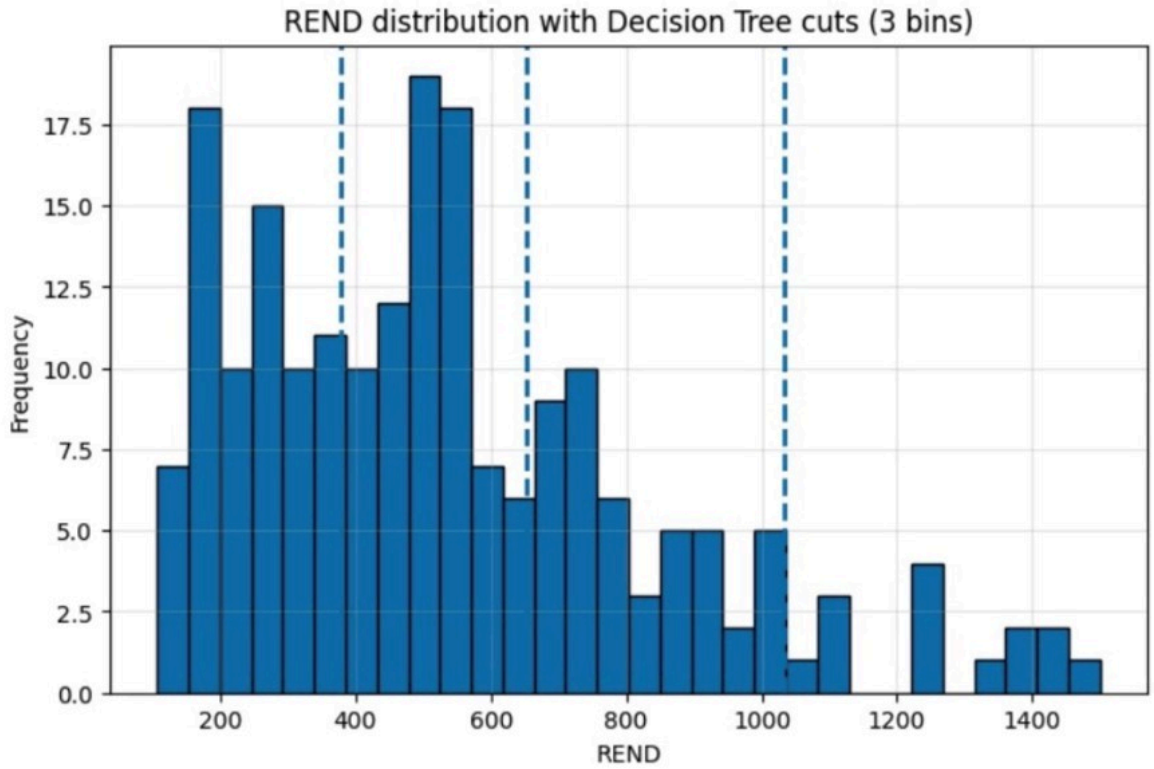


Figure 4

Decision tree cuts with dotted lines

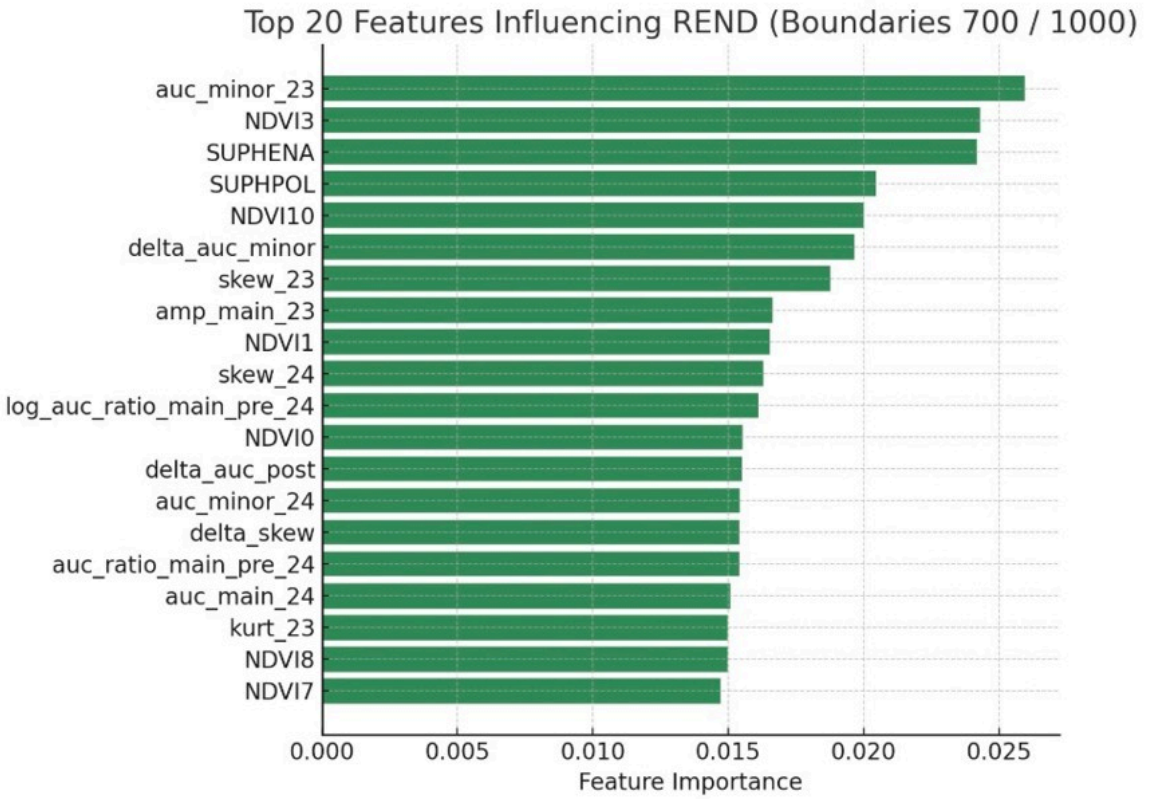


Figure 5

Ranking of the most important NDVI-derived features for the ensemble classifier

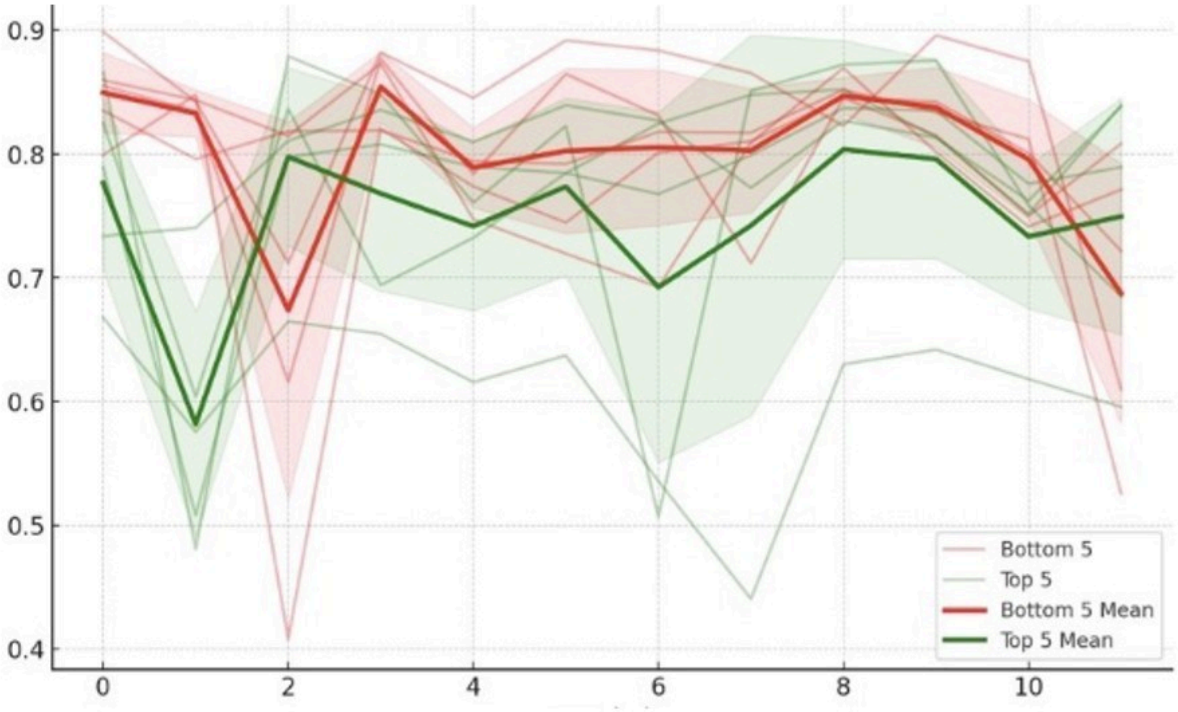


Figure 6

NDVI time series for the Top 5 vs. the Bottom 5 farms over 12 bimonthly periods

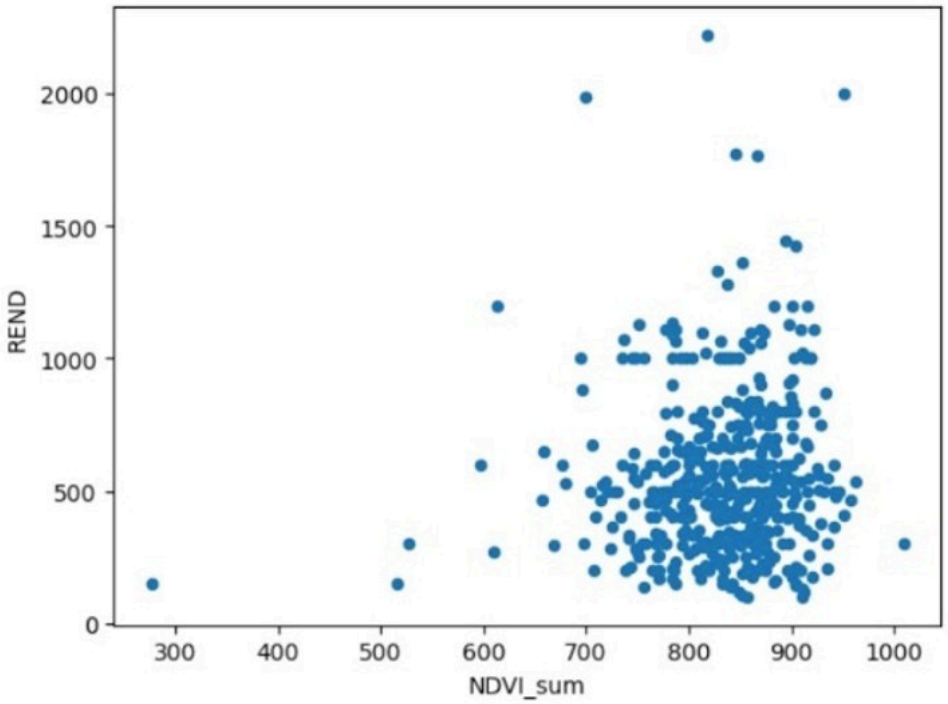


Figure 7

Scatter plot(s) of a key NDVI metric (e.g., cumulative NDVI or annual maximum) versus yield

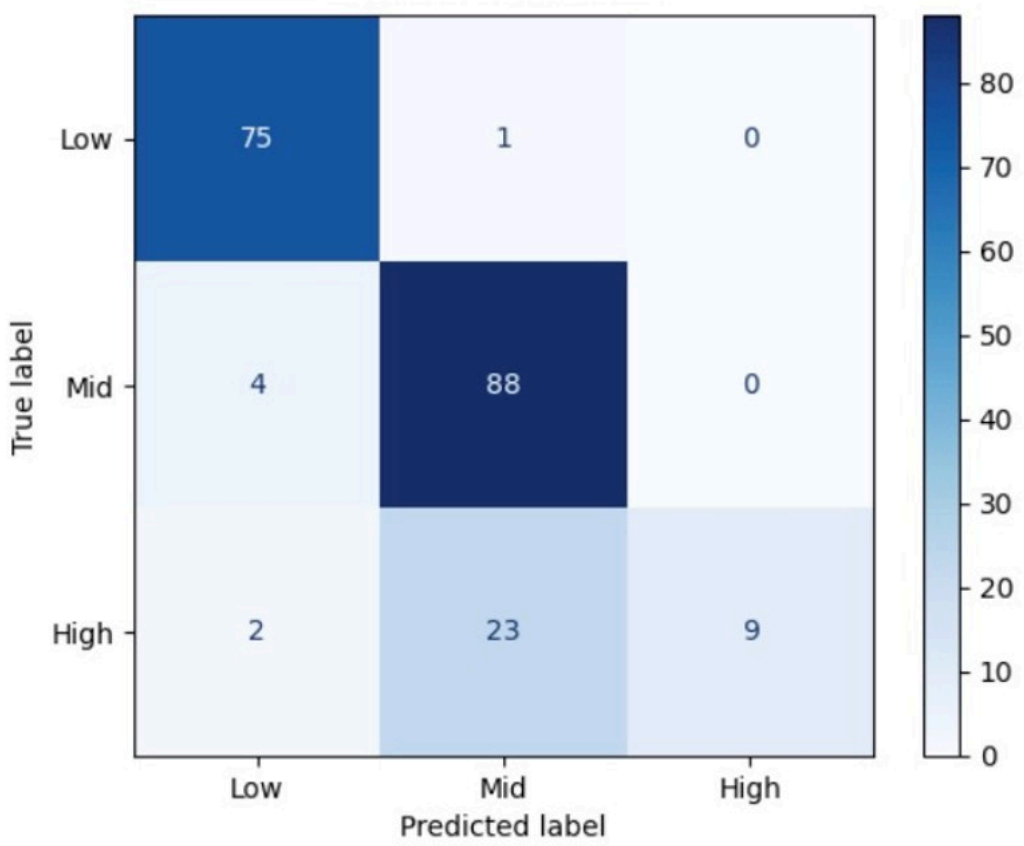


Figure 8

Confusion matrix for the selected three-category ensemble classifier (Low, Mid, High)