

Article



AI Based Prediction of Crop Loss Risk in Tropical Agricultural Systems of Panama with Implications for Food Security Planning

Paul R. Peterson ^{1,2,*} and Jay Molino ^{2,*}

¹ Food and Nutrition Security, Faculty of Biosciences and Public Health, Universidad Especializada de las Américas (UDELAS), Los Santos, Panama

² Environmental Analysis Laboratory, Center for Biotechnology, Green Energies and Climate Change, Faculty of Biosciences and Public Health, Universidad Especializada de las Américas (UDELAS), Panama City, Panama

* Correspondence: paul.peterson.2@udelas.ac.pa (P.R.P.); jay.molino@udelas.ac.pa (J.M.)

How To Cite: Peterson, P. R., & Molino, J. (2026). AI Based Prediction of Crop Loss Risk in Tropical Agricultural Systems of Panama with Implications for Food Security Planning. *Annals of Agri-bio Research*, 31(2), 158–167. <https://doi.org/10.53941/agrbio.2026.100018>

Received: 4 August 2026

Revised: 11 October 2026

Accepted: 15 October 2026

Published: 29 October 2026

Abstract: This study evaluated a Random Forest model for estimating crop-loss risk in small- and medium-scale maize and rice production systems in Los Santos, Guararé, and Tonosí, in the Azuero region of Panama. The dataset comprised approximately 720 repeated monthly records from 20 production units across three agricultural cycles, combining climatic, agronomic, phytosanitary, and production information. Crop loss was expressed relative to a common yield reference of 6.0 t/ha, with operational categories of low (<10%), medium (10–25%), and high (>25%) risk. The analysis used an 80/20 training-validation partition and five-fold cross-validation. Internal validation yielded $R^2 = 0.78$, while the fold-specific R^2 values shown in the cross-validation output ranged from approximately 0.75 to 0.80. Accumulated precipitation (28.6%) and average temperature (22.1%) had the greatest model importance, followed by irrigation frequency (15.4%) and pest incidence (13.7%). A matrix-based classification analysis of 81 observations yielded 64 correct risk assignments (79.0%), with an 85.7% high-risk recall. These findings support localized crop-loss risk assessment within the participating production units and identify potential applications in irrigation management, pest surveillance, and planting decisions relevant to food-security planning.

Keywords: Random Forest; crop-loss risk; tropical agriculture; climate variability; food security; Panama

1. Introduction

Agricultural production is central to food availability and rural livelihoods, particularly in tropical regions where climatic variability can amplify production risk. The Azuero region of Panama includes small- and medium-scale maize and rice production systems exposed to seasonal rainfall variation, high temperatures, pest pressure, and differences in field management. These conditions make localized crop-loss assessment relevant to agricultural planning. This study's food-security relevance concerns the availability and stability of staple-crop production.

At the global scale, climate change measurably affects agricultural productivity. Anthropogenic climate change has slowed global agricultural productivity growth since 1961, with particularly important implications for warmer regions (Ortiz-Bobea et al., 2021). Synchronized low-yield events across major agricultural regions represent a systemic risk that current climate-crop modeling frameworks may



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

underestimate (Kornhuber et al., 2023). Crop yields can respond nonlinearly to climatic stressors, and drought timing, duration, and severity consistently reduce yields (Santini et al., 2022; Leng, 2021; Liu et al., 2022; Tran et al., 2025). These findings support the development of predictive approaches to anticipate crop-loss risk and support early warning (Anderson et al., 2024).

Artificial intelligence, particularly machine learning, has emerged as a powerful approach for modeling agricultural production systems. Machine learning models can integrate heterogeneous datasets, including climatic, agronomic, and environmental variables, while capturing nonlinear interactions that traditional statistical approaches struggle to represent. A systematic review by van Klompenburg et al. (2020) identified Random Forest, Gradient Boosting, and neural networks as the most commonly used models for crop yield prediction, with temperature, precipitation, and soil properties as key predictive variables. More recent studies have demonstrated that integrating remote sensing data, such as Sentinel-2 time series, significantly improves predictive accuracy (Amin et al., 2024), while comparative analyses show that ensemble methods often achieve competitive performance while improving interpretability (Huber et al., 2022). Despite these advances, challenges related to model validation, generalization, and causal interpretation remain significant (Paudel et al., 2021; Hertel & de Lima, 2020).

Among machine learning approaches, Random Forest models offer distinct advantages for agricultural applications, particularly in data-constrained environments typical of tropical regions. Their robustness to multicollinearity, ability to capture nonlinear relationships, and resistance to overfitting make them suitable for modeling complex agro-environmental systems (Breiman, 2001). Recent developments, such as federated Random Forest models, have expanded their applicability in decentralized data contexts (Zhang et al., 2023). Additionally, studies incorporating multisource data, including climatic variables, soil characteristics, and evapotranspiration, have shown improved predictive performance, with Random Forest models serving as strong baselines (Khosravani Shariati & Abbasi, 2025). However, understanding model behavior under extreme climatic conditions and improving interpretability remain ongoing challenges (Paudel et al., 2023).

Despite these advances, important gaps persist in applying artificial intelligence to agricultural systems in tropical regions. Many existing studies rely on large-scale or global datasets that do not adequately capture local variability in agronomic practices and microclimatic conditions. Moreover, predictive models still insufficiently represent interactions between climatic variables, such as temperature and precipitation, despite evidence that these interactions significantly influence crop outcomes (Yin et al., 2022). Empirical analyses further indicate that current modeling approaches often underestimate the effects of extreme events such as droughts and heatwaves (Heinicke et al., 2022; Nóia Júnior et al., 2025).

Localized risk assessment is particularly relevant to Azuero because climatic variability and heterogeneous management conditions may influence productivity at the production-unit level. Integrating these variables in a predictive model can help identify combinations associated with crop loss and organize information for agricultural decision-making. Potential applications include irrigation planning, pest surveillance, and planting decisions (van Klompenburg et al., 2020; Paudel et al., 2021).

This study aimed to develop and internally evaluate a Random Forest model to estimate crop-loss risk in Panama's Azuero region. The analysis integrates climatic, agronomic, and phytosanitary variables, describes their predictive importance, and examines operational risk categories. The study considers food-security implications as potential applications of risk assessment to production planning.

2. Materials and Methods

2.1. Study Design and Data Sources

The study employed a non-experimental, quantitative, observational, and predictive design. It included 20 small- and medium-scale agricultural production units in Los Santos, Guararé, and Tonosí, primarily engaged in maize and rice cultivation. Records covered three consecutive agricultural cycles, with repeated monthly measurements generating approximately 720 observations. The 20 production units constitute the field sampling base; repeated records provide temporal information within these units rather than additional independent farms.

The database combined historical production records (65%), structured field survey data (25%), and climatic records from local meteorological stations (10%). Data collection used the same variables, instruments, and recording criteria across the three agricultural cycles. The dataset represents a local observational sample of the participating production units.

2.2. Outcome and Predictors

We defined crop loss as the percentage reduction in observed yield per hectare relative to a common reference yield of 6.0 t/ha. We calculated it as $100 \times (6.0 - \text{observed yield}) / 6.0$, with yield expressed in t/ha. The regional reference ranges described for the study were 5.8–6.2 t/ha for maize and 6.5–7.0 t/ha for rice. The common 6.0 t/ha reference provides a consistent denominator for the operational loss index. We grouped predicted losses into low (<10%), medium (10–25%), and high (>25%) risk categories. These thresholds interpret continuous predictions.

An algebraic sensitivity comparison relates the reference yield to the category boundaries. For a positive reference yield B , the 10% and 25% loss cutpoints correspond to yields of $0.90B$ and $0.75B$. At $B = 6.0$ t/ha, low risk corresponds to yield >5.40 t/ha, medium risk to 4.50 – 5.40 t/ha, and high risk to yield <4.50 t/ha. Using the maize range described above moves the two cutpoints to 5.22 – 5.58 and 4.35 – 4.65 t/ha; using the rice range moves them to 5.85 – 6.30 and 4.875 – 5.25 t/ha. For an illustrative yield of 5.0 t/ha, calculated loss is 16.7% at $B = 6.0$ and 28.6% at $B = 7.0$, corresponding to medium and high risk, respectively.

Predictors included accumulated precipitation, average temperature, relative humidity, pest incidence, irrigation information, fertilization, and agronomic management practices. Climatic variables were aggregated monthly. We classified pest incidence as low, medium, or high using producer reports and field observations. We summarized irrigation type at the production-unit level as supplementary irrigation or rainfed management; the model importance output identified irrigation frequency as the irrigation predictor. We described fertilization as organic, chemical, or mixed.

2.3. Data preprocessing, Model Development and Validation

We organized and processed the data in Python using Pandas and NumPy. We removed records with missing values in critical variables; these variables had missingness in less than 5% of the dataset. Missing values in secondary variables represented less than 8% and were imputed using the mean for numerical variables and the mode for categorical variables. We screened potential outliers using the interquartile range (IQR): values below $Q1 - 1.5 \times \text{IQR}$ or above $Q3 + 1.5 \times \text{IQR}$ were reviewed for agronomic plausibility. This procedure excluded only records considered agronomically implausible, accounting for approximately 3.2% of records. We normalized numerical variables and encoded categorical variables before model fitting.

We implemented a supervised Random Forest Regressor with scikit-learn to capture nonlinear relationships among the predictors. Model parameters were adjusted using cross-validation. The final configuration was `n_estimators = 150`, `max_depth = 10`, `min_samples_split = 5`, `min_samples_leaf = 2`, `max_features = "sqrt"`, and `random_state = 42`. Monthly aggregation and categorical encoding formed part of predictor preparation. We then assigned continuous model predictions to the operational loss-risk categories defined in Section 2.2.

The analysis used an 80/20 training-validation partition and five-fold cross-validation to quantify predictive performance within the study dataset. We evaluated regression performance using the coefficient of determination (R^2), and fold-specific R^2 values captured variation across the internal partitions. The documented results are reported as record-level internal validation; grouping by production unit or time was not documented. We summarized continuous variables using the mean, median, standard deviation, minimum, maximum, and IQR, while summarizing categorical management characteristics as frequencies and percentages across the 20 production units. We evaluated yield differences among predicted risk categories using one-way ANOVA followed by Tukey post hoc comparisons, along with category-specific means and standard deviations. ANOVA, Tukey, and Pearson p -values were unadjusted for production-unit dependence. We estimated predictor importance using the Random Forest mean-decrease-in-impurity (MDI) criterion to describe relative split-based importance in the fitted model.

The confusion matrix summarizes categorical agreement for 81 observations. Rows denote observed categories and columns denote predicted categories. Accuracy is the diagonal total divided by the matrix total; precision is the diagonal count divided by the corresponding column total; recall is the diagonal count divided by the row total; and F1 is the harmonic mean of precision and recall. These matrix-based metrics complement regression validation by quantifying agreement within the categorized observations. We calculated two-sided 95% Wilson score intervals for accuracy and each class-specific precision and recall from the matrix counts (Wilson, 1927). These nominal intervals use an independent-binomial assumption and are conditional on the matrix sample.

2.4. Food Security Interpretation

We considered food-security implications through potential decision-support applications of model-generated risk information. We examined how risk identification could inform irrigation adjustments, pest surveillance, and planting decisions, with emphasis on the availability and stability of staple-crop production.

3. Results

3.1. Production and Management Characteristics

Mean observed yield was 5.21 ± 0.88 t/ha. Precipitation, temperature, and relative humidity varied across the study records (Table 1), reflecting heterogeneous environmental conditions. Supplementary irrigation was reported for 13 of the 20 production units, while seven were rainfed. Chemical fertilization was the most frequent fertilization category, and pest or disease levels ranged from low to high (Table 2).

Table 1. Descriptive statistics of continuous agricultural variables. Rows, in order, describe yield (t/ha), monthly accumulated precipitation (mm), mean temperature (°C), and relative humidity (%). SD denotes standard deviation; IQR denotes interquartile range.

Mean	Median	SD	Min	Max	IQR
5.21	5.18	0.88	3.40	6.90	1.12
182	175	64	70	325	95
29.8	29.6	2.1	26.1	34.2	3.4
78.5	79.0	5.8	65	89	7.2

Table 2. Agricultural management characteristics of the 20 production units.

Variable	Category	Frequency (n)	Percentage (%)
Irrigation type	Supplementary irrigation	13	65
	Rainfed only	7	35
Fertilization	Organic	6	30
	Chemical	9	45
	Mixed	5	25
Pest/disease level	Low	8	40
	Medium	7	35
	High	5	25

Temperature showed a negative association with crop yield, indicating that higher average temperatures were linked to reduced productivity. The Pearson correlation indicated a moderate negative relationship between average temperature and yield ($r = -0.48$; nominal $p < 0.01$), consistent with a possible thermal association with crop performance in the analyzed production units (Figure 1).

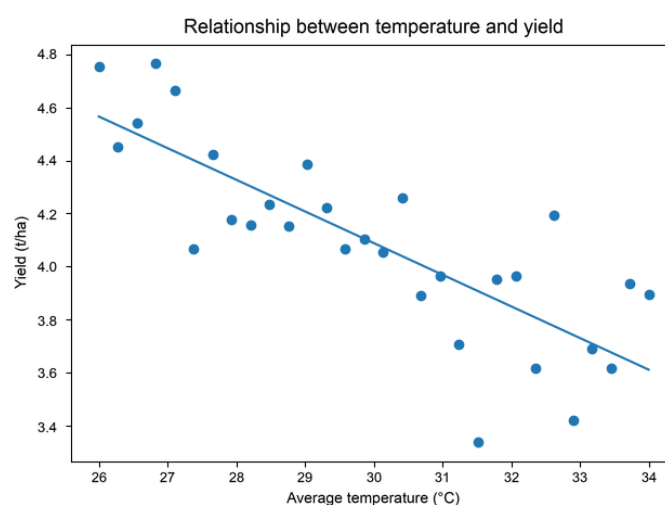


Figure 1. Relationship between average temperature (°C; horizontal axis) and agricultural yield (t/ha; vertical axis). Blue points show the plotted temperature–yield pairs, and the blue line shows the linear fit. The Spanish labels Temperatura promedio and Rendimiento denote average temperature and yield, respectively.

The Random Forest model achieved an internal validation R^2 of 0.78. The five fold-specific R^2 values displayed in the cross-validation output ranged from approximately 0.75 to 0.80 (Figure 2), indicating broadly similar performance across the internal partitions.

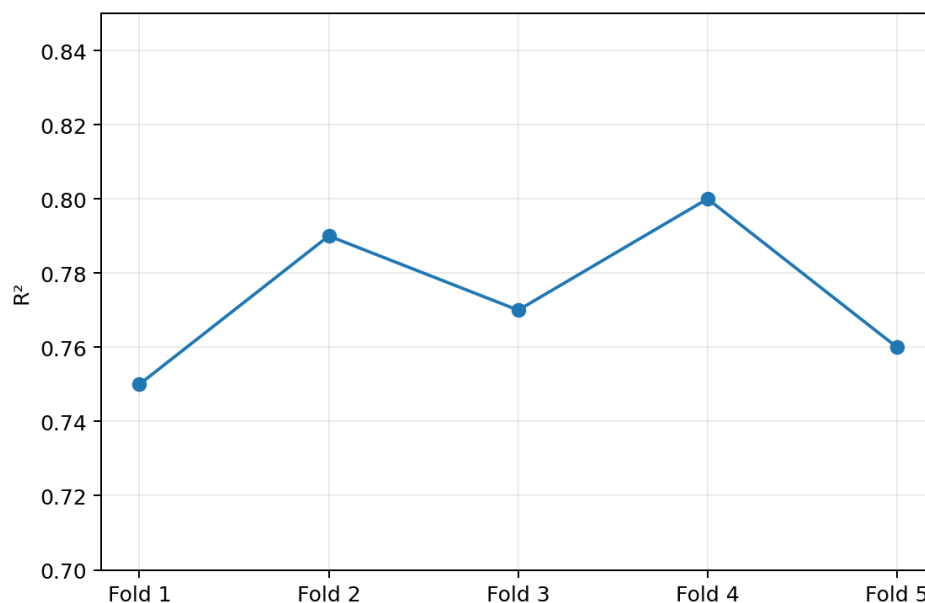


Figure 2. Five-fold internal cross-validation performance of the Random Forest model. The horizontal axis identifies folds 1–5; the vertical axis shows R^2 (dimensionless), with displayed values of approximately 0.75, 0.79, 0.77, 0.80, and 0.76. Connecting lines aid visual comparison.

Accumulated precipitation had the highest mean decrease in impurity importance (28.6%), followed by average temperature (22.1%), irrigation frequency (15.4%), and pest incidence (13.7%). Relative humidity (8.2%), fertilization (6.8%), and agronomic management practices (5.2%) accounted for the remaining importance (Figure 3). Together, precipitation and temperature represented 50.7% of the reported importance, indicating their prominence in the fitted model.

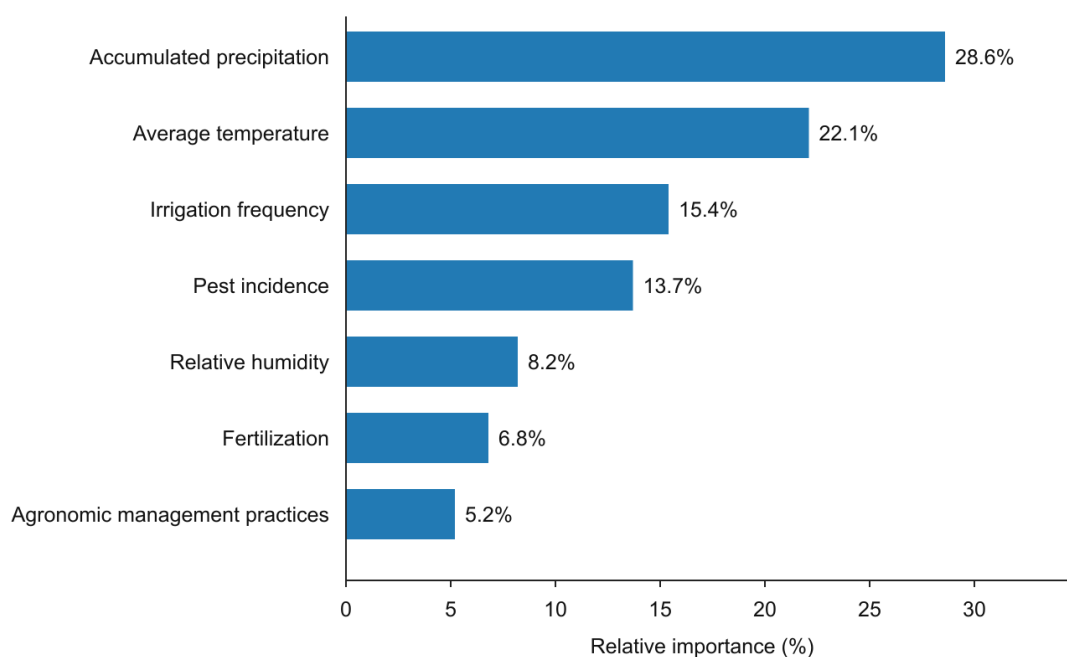


Figure 3. Relative importance of predictors in the Random Forest model, expressed as percentages of mean-decrease-in-impurity importance. Bars identify the seven predictors; the horizontal axis gives normalized impurity importance (%), summing to 100%. Precipitation is accumulated in mm, temperature in °C, and relative humidity in %. The percentages describe relative split-based importance in the fitted model.

3.3. Operational Risk Categories

The confusion matrix contains 26 low-risk, 27 medium-risk, and 28 high-risk observations, for a total of 81 (Figure 4). The diagonal contains 64 observations, giving 79.0% agreement between observed and predicted categories. High-risk recall was 85.7% and high-risk F1 was 0.828. Errors were concentrated between neighboring categories, with lower recall for the medium-risk category (Table 3). Overall accuracy had a nominal 95% Wilson interval of 68.9–86.5%, and high-risk recall had an interval of 68.5–94.3%.

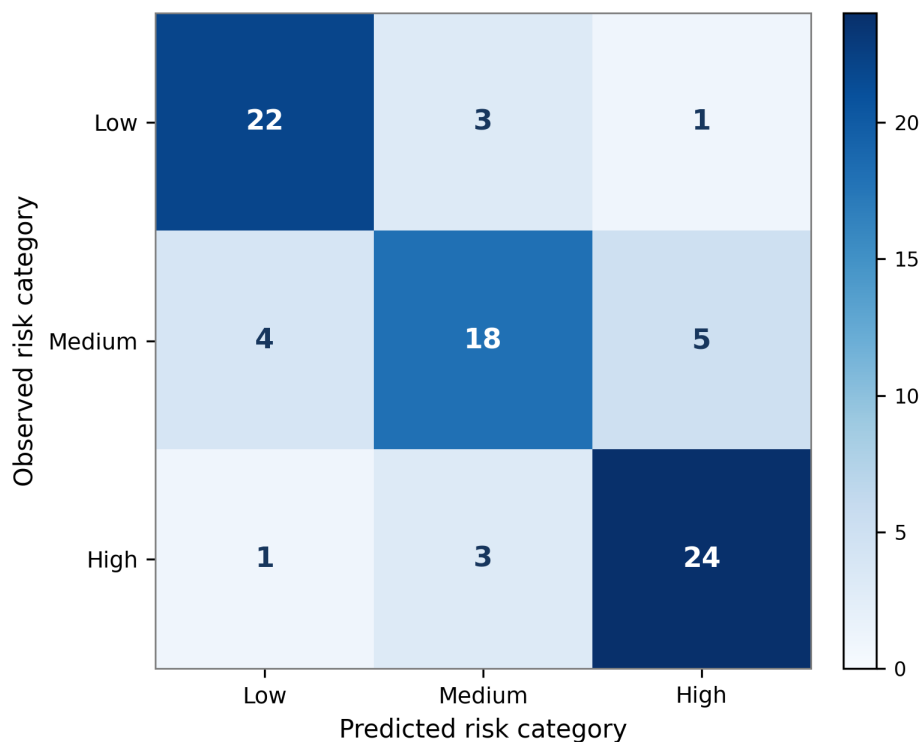


Figure 4. Observed and predicted crop-loss risk categories in the confusion matrix ($n = 81$). Rows indicate observed categories and columns indicate predicted categories. Cells and the blue color scale show observation counts, with darker shading indicating larger counts; diagonal cells denote correct assignments. Categories are low ($<10\%$), medium ($10\text{--}25\%$), and high ($>25\%$) loss relative to 6.0 t/ha. Row totals are 26, 27, and 28; column totals are 27, 24, and 30.

Table 3. Classification metrics derived from the confusion matrix.

Risk Category	Precision (95% CI)	Recall (95% CI)	F1	Support
Low ($<10\%$)	0.815 (0.633–0.918)	0.846 (0.665–0.938)	0.830	26
Medium ($10\text{--}25\%$)	0.750 (0.551–0.880)	0.667 (0.478–0.814)	0.706	27
High ($>25\%$)	0.800 (0.627–0.905)	0.857 (0.685–0.943)	0.828	28

Note: Metrics calculated from the cell counts in Figure 4. Overall accuracy = $64/81 = 79.0\%$ (nominal 95% Wilson CI: 68.9–86.5%). Support denotes the row total for each observed category. Parentheses give nominal 95% Wilson intervals under an independent-binomial assumption. Precision denominators are the column totals (27, 24, 30); recall denominators are the row totals (26, 27, 28).

Yield values decreased progressively across predicted risk categories as crop loss risk increased. The mean yield was 5.88 ± 0.64 t/ha in the low-risk group, 5.22 ± 0.71 t/ha in the medium-risk group, and 4.74 ± 0.83 t/ha in the high-risk group. One-way ANOVA reported differences among risk categories ($F(2,717) = 38.7$; nominal $p < 0.001$), and Tukey post hoc comparisons indicated that the high-risk group had a mean yield 1.14 t/ha lower than the low-risk group (nominal $p < 0.001$). Under the standard one-way ANOVA formulation, the degrees of freedom correspond to three categories and 720 observations. The yield gradient describes consistency between observed yield and predicted risk categories (Table 4 and Figure 5).

Table 4. Agricultural yield by predicted crop-loss risk category.

Predicted Risk Category	Mean Yield (t/ha)	SD (t/ha)
Low (<10%)	5.88	0.64
Medium (10–25%)	5.22	0.71
High (>25%)	4.74	0.83

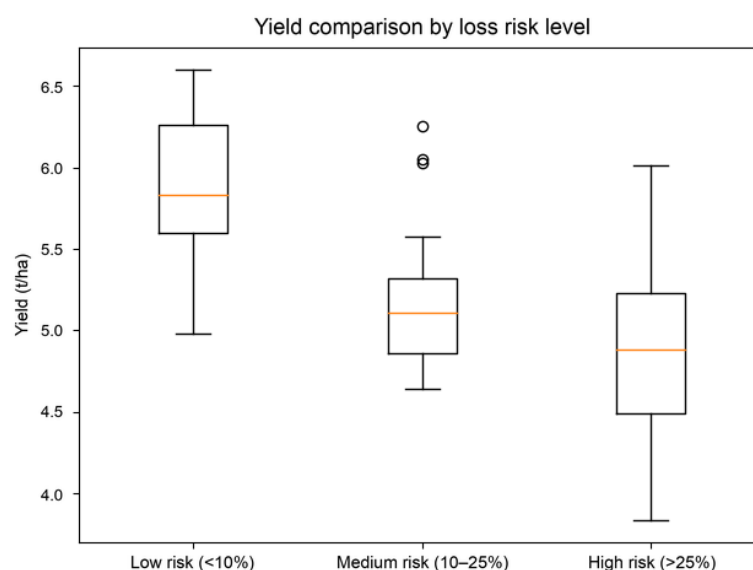


Figure 5. Agricultural yield (t/ha; vertical axis) by predicted crop-loss risk category (horizontal axis). Riesgo Bajo, Medio, and Alto denote low (<10%), medium (10–25%), and high (>25%) loss relative to 6.0 t/ha. Boxes show the interquartile ranges and orange lines indicate medians; whiskers and open circles display the spread and plotted outliers. Observed yield is shown within predicted categories, so individual yields need not satisfy the predicted-category cutpoints.

4. Discussion

The internal validation R^2 of 0.78 indicates that a Random Forest model can capture predictive relationships among climatic, agronomic, and phytosanitary information in the Azuero study dataset. The fold-specific cross-validation results show broadly similar performance across internal partitions. These findings are consistent with the suitability of ensemble tree methods for heterogeneous agricultural predictors and nonlinear relationships (Breiman, 2001; van Klompenburg et al., 2020). Their contribution is a localized assessment of crop-loss risk using information relevant to small- and medium-scale production systems.

Precipitation and temperature together accounted for 50.7% of total model importance, indicating their prominence under the fitted model's impurity criterion. This pattern is consistent with agricultural production's sensitivity to water availability and thermal conditions. Previous research has linked climatic stress to productivity changes and yield instability, including drought and temperature-precipitation interactions (Ortiz-Bobea et al., 2021; Santini et al., 2022; Yin et al., 2022; Kornhuber et al., 2023).

Irrigation frequency and pest incidence also contributed to prediction, supporting the value of combining environmental information with field-management and phytosanitary observations. The combined predictor set enabled the model to identify production units experiencing conditions associated with elevated risk. This multidimensional structure strengthens the model's utility for localized risk assessment and targeted management planning.

Operational categories provide a practical way to communicate continuous predictions. Within the 81 categorized observations, overall accuracy was 79.0%, with recall of 84.6% for low risk, 66.7% for medium risk, and 85.7% for high risk. The lower recall for the medium-risk category indicates greater overlap at intermediate conditions. Yield also showed a descriptive decrease across categories, and the unadjusted Tukey comparisons reported a 1.14 t/ha difference between the high- and low-risk groups. These results describe the internal risk stratification and the overlap between neighboring categories.

The food-security relevance lies in using model-generated risk information to support the availability and stability of staple-crop production. For example, an elevated-risk estimate during a dry period could prompt inspection of water availability before an irrigation decision, while elevated risk accompanied by

pest reports could prioritize field scouting before treatment. Aggregated alerts could help extension services prioritize visits and prepare contingency planning for local staple supply.

Limitations and Scope of Application

The 20 production units provide the independent field sampling base, while approximately 720 monthly records describe variation within those units. Repeated records share management and temporal conditions, which can inflate internal performance when the same unit occurs in training and validation (Roberts et al., 2017). Fold membership and shuffle settings were not retained, so the extent of overlap and potential bias cannot be quantified. $R^2 = 0.78$ should therefore be interpreted as internal performance within the study dataset. Production-unit-grouped and forward temporal validation remain necessary to assess transferability to unseen farms and future cycles. The observation-level database and partition logs are unavailable for these reanalyses and for auditing preprocessing within folds.

ANOVA, Tukey comparisons, and Pearson correlation were not adjusted for within-unit dependence; their p-values are nominal and may overstate statistical evidence. Group counts after preprocessing cannot be verified from the aggregate outputs. Because the loss outcome is derived from yield, differences in yield across risk categories describe internal consistency rather than independent model validation. The 81-case confusion matrix is a separate categorization output whose selection criteria, crop composition, and production-unit membership are unavailable; it cannot be identified as the complete 20% validation partition. Its nominal Wilson intervals quantify uncertainty conditional on the matrix counts and do not account for clustering, case selection, model fitting, or threshold choice. Point-level selection in the scatter plot and group-specific counts in the boxplot also cannot be reconstructed.

The common 6.0 t/ha reference standardizes the operational loss index. The algebraic comparison in Section 2.2 shows that reference choice can change categories, particularly for rice, whose stated reference range is higher. These ranges provide a numerical comparison but do not establish calibrated attainable yields or equal agronomic loss severity across crops. Classification robustness across crop-specific references requires evaluation using the underlying records. Further model development could incorporate soil properties, crop phenology, remotely sensed vegetation indices, socioeconomic conditions, and input intensity.

MDI importance can favor continuous or high-cardinality predictors, while correlated predictors may share or redistribute importance (Strobl et al., 2007, 2008). The reported percentages describe the fitted model and should not be interpreted as causal contributions or unique predictor effects. Permutation importance on appropriately held-out data, with attention to predictor correlation and production-unit grouping, would provide a complementary assessment.

The proposed decision pathways may support production availability and stability. Intervention effects, household food access, dietary adequacy, and food-security indicators were not measured. Monthly predictor aggregation does not establish a reliable pre-harvest warning lead time. Prospective evaluation should use information available at the forecast issue date and assess performance at prespecified horizons, alongside production loss, supply stability, and household outcomes.

5. Conclusions

The study demonstrates the feasibility of localized Random Forest-based crop-loss assessment using climatic, agronomic, and phytosanitary information from production units in Panama's Azuero region. The model achieved an internal validation R^2 of 0.78, with broadly similar fold-specific cross-validation performance. Precipitation and temperature had the greatest reported MDI importance, followed by irrigation frequency and pest incidence. The operational risk categories provided an interpretable framework for communicating model predictions.

The findings support further development of agricultural decision support within the conditions represented by the study. Potential applications include prioritizing irrigation management, pest surveillance, and extension activities relevant to food-security planning. Validation across independent production units and future agricultural cycles will help establish the scope of these applications.

Author Contributions

P.R.P. led the conceptualization, methodology, investigation, formal analysis, and preparation of the original manuscript. P.R.P. also contributed to data curation, funding acquisition, project administration, resource provision, software development, supervision, validation, visualization, and the review and editing

of the final manuscript. J.M. contributed to the conceptualization, methodology, investigation, data curation, formal analysis, project administration, resource provision, supervision, validation, visualization, and critical review and editing of the manuscript. All authors have read and approved the final version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable. This study did not involve human participants.

Data Availability Statement

The results supporting the findings are reported in this article. Requests for the underlying in situ data should be directed to the corresponding author.

Conflicts of Interest

Given their roles as Guest Editors, Paul R. Peterson and Jay Molino had no involvement in the peer review of this paper and had no access to information regarding its peer-review process. Full responsibility for the editorial process of this paper was delegated other editors of the journal.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used Grammarly and ChatGPT (OpenAI) to assist with language editing. The authors reviewed and edited the content and take full responsibility for the published article.

References

- Amin, E., Pipia, L., Belda, S., Perich, G., Graf, L. V., Aasen, H., Van Wittenberghe, S., Moreno, J., & Verrelst, J. (2024). In-season forecasting of within-field grain yield from Sentinel-2 time series data. *International Journal of Applied Earth Observation and Geoinformation*, 126, 103636. <https://doi.org/10.1016/j.jag.2023.103636>.
- Anderson, W., Shukla, S., Verdin, J., Hoell, A., Justice, C., Barker, B., Slinski, K., Lenssen, N., Lou, J., Cook, B. I., & McNally, A. (2024). Preseason maize and wheat yield forecasts for early warning of crop failure. *Nature Communications*, 15(1), 7262. <https://doi.org/10.1038/s41467-024-51555-8>.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>.
- Heinicke, S., Frieler, K., Jägermeyr, J., & Mengel, M. (2022). Global gridded crop models underestimate yield responses to droughts and heatwaves. *Environmental Research Letters*, 17(4), 044026. <https://doi.org/10.1088/1748-9326/ac592e>.
- Hertel, T. W., & de Lima, C. Z. (2020). Viewpoint: Climate impacts on agriculture: Searching for keys under the streetlight. *Food Policy*, 95, 101954. <https://doi.org/10.1016/j.foodpol.2020.101954>.
- Huber, F., Yushchenko, A., Stratmann, B., & Steinhage, V. (2022). Extreme Gradient Boosting for yield estimation compared with Deep Learning approaches. *Computers and Electronics in Agriculture*, 202, 107346. <https://doi.org/10.1016/j.compag.2022.107346>.
- Khosravani Shariati, S. A., & Abbasi, A. (2025). Machine learning-based winter wheat yield prediction using multisource data. *Agricultural Water Management*, 322, 109951. <https://doi.org/10.1016/j.agwat.2025.109951>.
- Kornhuber, K., Lesk, C., Schleussner, C. F., Jägermeyr, J., Pfleiderer, P., & Horton, R. M. (2023). Risks of synchronized low yields are underestimated in climate and crop model projections. *Nature Communications*, 14(1), 3528. <https://doi.org/10.1038/s41467-023-38906-7>.
- Leng, G. (2021). Maize yield loss risk under droughts in observations and crop models in the United States. *Environmental Research Letters*, 16(2), 024016. <https://doi.org/10.1088/1748-9326/abd500>.

- Liu, S., Xiao, L., Sun, J., Yang, P., Yang, X., & Wu, W. (2022). Probability of maize yield failure increases with drought occurrence but partially depends on local conditions in China. *European Journal of Agronomy*, 139, 126552. <https://doi.org/10.1016/j.eja.2022.126552>.
- Nóia Júnior, R. D. S., Asseng, S., Müller, C., Deswarte, J. C., Cohan, J. P., & Martre, P. (2025). Negative impacts of climate change on crop yields are underestimated. *Trends in Plant Science*, 30(11), 1262–1273. <https://doi.org/10.1016/j.tplants.2025.05.002>.
- Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G., & Lobell, D. B. (2021). Anthropogenic climate change has slowed global agricultural productivity growth. *Nature Climate Change*, 11(4), 306–312. <https://doi.org/10.1038/s41558-021-01000-1>.
- Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187, 103016. <https://doi.org/10.1016/j.agsy.2020.103016>.
- Paudel, D., de Wit, A., Boogaard, H., Marcos, D., Osinga, S., & Athanasiadis, I. N. (2023). Interpretability of deep learning models for crop yield forecasting. *Computers and Electronics in Agriculture*, 206, 107663. <https://doi.org/10.1016/j.compag.2023.107663>.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. <https://doi.org/10.1111/ecog.02881>.
- Santini, M., Noce, S., Antonelli, M., & Caporaso, L. (2022). Complex drought patterns robustly explain global yield loss for major crops. *Scientific Reports*, 12(1), 5792. <https://doi.org/10.1038/s41598-022-09611-0>.
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>.
- Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307. <https://doi.org/10.1186/1471-2105-9-307>.
- Tran, B. L., Tseng, W. C., & Chen, C. C. (2025). Climate change impacts on crop yields across temperature rise thresholds and climate zones. *Scientific Reports*, 15(1), 23424. <https://doi.org/10.1038/s41598-025-07405-8>.
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. <https://doi.org/10.1016/j.compag.2020.105709>.
- Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, 22(158), 209–212. <https://doi.org/10.1080/01621459.1927.10502953>.
- Yin, X., Leng, G., & Yu, L. (2022). Disentangling the separate and confounding effects of temperature and precipitation on global maize yield using machine learning, statistical and process crop models. *Environmental Research Letters*, 17(4), 044036. <https://doi.org/10.1088/1748-9326/ac5716>.
- Zhang, Q., Zhao, X., Han, Y., Yang, F., Pan, S., Liu, Z., Wang, K., & Zhao, C. (2023). Maize yield prediction using federated random forest. *Computers and Electronics in Agriculture*, 210, 107930. <https://doi.org/10.1016/j.compag.2023.107930>.