Toward Robust and Interpretable Rice Yield Prediction: A Season-Specific Machine Learning Framework

Yuxin Luo Eurus Chang M. Nur Alam

Authors and Affiliations:
Yuxin Luo, Analytics, University College London (UCL), `y.luo.24@ucl.ac.uk`
Eurus Chang, Analytics, University College London (UCL), `yang.chang.24@ucl.ac.uk`
M. Nur Alam, Applied Research, GreentecAI Ltd, `nur.alam@greentecai.com`

Abstract

Rice sustains over three billion people, yet its yield formation is highly vulnerable to climatic extremes, input price volatility, and heterogeneous management by smallholder farmers. Anticipating yield outcomes under these uncertainties is both a scientific challenge and a policy necessity. We present a *season-specific machine learning (ML) framework* that formalizes rice yield prediction as supervised learning under distributional shift across ecotypes: Boro (irrigated dry-season), T.Aus (pre-monsoon), and T.Aman (monsoon).

Our contributions are fourfold. We fuse agronomic inputs (fertilizer applications of N, P, K, S; irrigation; biological nitrogen fixation) with high-frequency NASA POWER weather variables (temperature, precipitation, relative humidity), aggregated to phenology-aware windows. We benchmark eight regression families under leakage-controlled pipelines and identify ecotype-specific optima: Linear Regression for Boro (R^2=0.878), Support Vector Regression for T.Aus (R^2=0.862), and K-Nearest Neighbors for T.Aman (R^2=0.857). We expand interpretability through bias–variance decomposition, mutual information analysis, and causal structural equation modeling, providing theory-grounded explanations of why different inductive biases dominate per season. Finally, we quantify predictive uncertainty using bootstrap ensembles and deploy the pipelines in a lightweight Streamlit tool that supports real-time "what-if" scenario testing (e.g., rainfall shocks, fertilizer adjustments). The results illustrate how interpretable and uncertainty-aware ML can provide a transferable framework for data-driven, climate-resilient agriculture in smallholder contexts.

1 Introduction

Rice sustains over 3.5 billion people worldwide, providing more than 20% of global caloric intake [1]. Its centrality is particularly pronounced in South and Southeast Asia, where smallholder farmers dominate production and where climatic variability, rising input costs, and fragile infrastructure combine to create systemic risks. Anticipating rice yield under uncertainty is therefore both a scientific and policy priority: accurate forecasts inform food security planning, insurance, and subsidy allocation.

Traditional process-based crop growth models such as DSSAT [2], APSIM [3], and ORYZA [4] provide mechanistic insight by simulating crop–soil–climate interactions. However, they require fine-grained calibration (soil chemistry, cultivar genetics) rarely available in smallholder systems and often struggle to capture sub-seasonal weather extremes or heterogeneous farmer practices, limiting predictive fidelity.

Machine learning (ML) offers a complementary paradigm: by directly learning from observational data, ML can capture nonlinear dependencies and high-order interactions between management practices and weather events. Recent advances have demonstrated success in predicting yield from remote sensing and weather data [5–7], yet much of this work either aggregates rice into a single category, ignoring ecotypic differences, or focuses narrowly on retrospective accuracy without building operational pipelines.

Research gap

In South Asia, rice spans three ecotypes with distinct hydrological regimes and climate sensitivities. Boro is a dry-season crop that is fully irrigated and therefore relatively controlled but energy-intensive. T.Aus is a short-duration pre-monsoon crop that is highly vulnerable to heat and rainfall shocks. T.Aman is a long-duration monsoon crop, largely rain-fed and exposed to floods and nutrient leaching. Treating these ecotypes as a single distribution $p(y \mid X)$ mis-specifies the problem; each season induces a distinct conditional distribution $p_s(y \mid X)$ shaped by local hydrology, farmer decisions, and climatic variability. This mismatch introduces systematic bias and undermines interpretability.

Contributions

We propose a season-specific ML framework that integrates district-level management data with high-frequency NASA POWER weather series. Our contributions are fourfold.

First, we design ecotype-specific predictors and show that the inductive bias must match ecotype complexity: linear models excel in Boro, kernel methods in T.Aus, and local instance-based methods in T.Aman.

Second, we provide a theoretical lens—bias–variance decomposition, mutual information, and structural causal models—that explains *why* these choices work and connects ML with agronomic theory.

Third, we quantify uncertainty using bootstrap-based prediction intervals and Bayesian baselines, which is essential for risk-aware agricultural planning.

Finally, we operationalize the best pipelines in a Streamlit tool that enables real-time scenario testing for agronomists and policymakers.

2 Related Work

Yield prediction spans process-based models, statistical approaches, and modern ML. Early works relied on regressions using aggregated climate indices [10]; process-based models such as DSSAT and ORYZA offered physiological realism but required heavy calibration. With the rise of ML, studies demonstrated that satellite vegetation indices and weather features could explain yield variability at scale [5], and deep models trained on weather and soil achieved state-of-the-art results in large row crops [6]. Reviews highlight opportunities and pitfalls for deep learning in agriculture [7]. Within South Asia, prior work has explored rice yield and price prediction [9] and climate impacts on

Bangladesh rice yields [8]. However, season-specific modeling and operational deployment remain under-explored; our work addresses both.

3 Problem Formulation

We formalize rice yield prediction as supervised regression with distributional shift across ecotypes. Let $\mathcal{D}_s = \{(X_i^{(s)}, y_i^{(s)})\}_{i=1}^{n_s}$ for $s \in \{\text{Boro}, \text{T.Aus}, \text{T.Aman}\}$, where $X_i^{(s)} \in \mathbb{R}^d$ are feature vectors and $y_i^{(s)} \in \mathbb{R}$ are yields (t/ha). Each season defines a distinct conditional distribution $p_s(y \mid X)$.

Learning objective. For each season s, we learn $f^{(s)} : \mathbb{R}^d \to \mathbb{R}$ that minimizes

$$\mathcal{L}^{(s)}(f) = \mathbb{E}_{(X,y)\sim p_s}\big[(y - f(X))^2\big], \qquad \hat{\mathcal{L}}^{(s)}(f) = \frac{1}{n_s}\sum_{i=1}^{n_s}\Big(y_i^{(s)} - f(X_i^{(s)})\Big)^2. \tag{1}$$

Evaluation metrics. We report

$$\text{RMSE}^{(s)} = \sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s}(y_i^{(s)} - \hat{y}_i^{(s)})^2}, \quad \text{MAE}^{(s)} = \frac{1}{n_s}\sum_{i=1}^{n_s}|y_i^{(s)} - \hat{y}_i^{(s)}|, \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_s}(y_i^{(s)} - \hat{y}_i^{(s)})^2}{\sum_{i=1}^{n_s}(y_i^{(s)} - \bar{y}^{(s)})^2}. \tag{3}$$

Bias–variance perspective. Generalization error decomposes as

$$\mathbb{E}[(y - \hat{f}(X))^2] = \underbrace{\sigma^2}_{\text{irreducible}} + \underbrace{\text{Bias}^2[\hat{f}(X)]}_{\text{systematic}} + \underbrace{\text{Var}[\hat{f}(X)]}_{\text{instability}}. \tag{4}$$

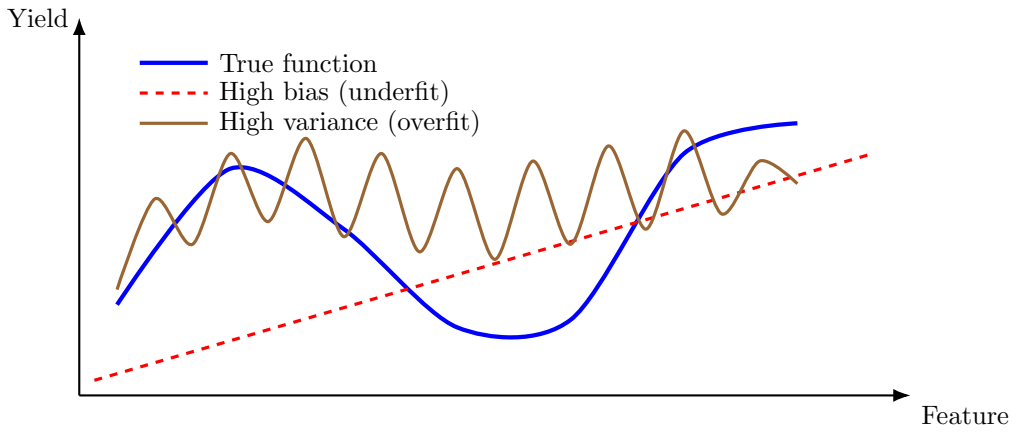
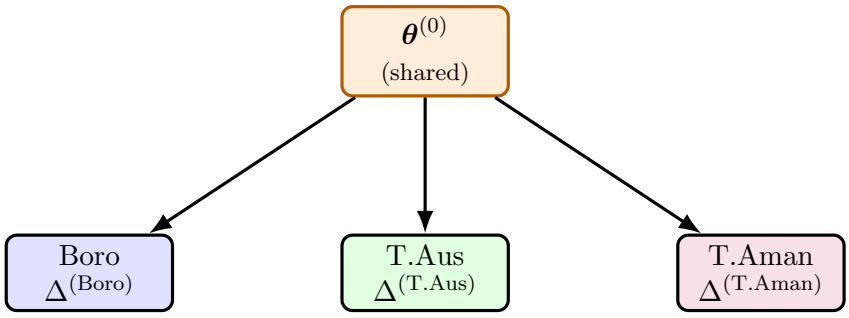


Figure 1: Bias–variance schematic. The dashed red line illustrates a high-bias underfit, the brown wiggly curve a high-variance overfit, and the solid blue curve the underlying true function.

This motivates different inductive biases across seasons: linear for Boro (low variance), kernel for T.Aus (nonlinear climate sensitivity), and local instance-based for T.Aman (heterogeneity).

Multi-task sharing. We couple seasons via $\theta^{(s)} = \theta^{(0)} + \Delta^{(s)}$ and estimate

$$\min_{\theta^{(0)},\{\Delta^{(s)}\}} \sum_s \sum_i \Big(y_i^{(s)} - (X_i^{(s)})^\top(\theta^{(0)} + \Delta^{(s)})\Big)^2 + \lambda \sum_s \|\Delta^{(s)}\|_1, \tag{5}$$



$\theta^{(s)} = \theta^{(0)} + \Delta^{(s)}$ with group-sparse penalty $\sum_s \|\Delta^{(s)}\|_1$.

Figure 2: Multi-task sharing of parameters. The global parameter $\theta^{(0)}$ captures universal relationships (e.g., baseline nutrient response), while season-specific deviations $\Delta^{(s)}$ capture ecotype heterogeneity (e.g., rainfall sensitivity in T.Aus, irrigation dominance in Boro). This encourages statistical efficiency through shared structure while preserving critical seasonal adaptations.

Uncertainty modeling. Point estimates alone are insufficient for agricultural decision-making under climate variability. We therefore complement predictions with uncertainty quantification via a Bayesian linear regression baseline:

$$y \mid X, \theta \sim \mathcal{N}(X^\top\theta, \sigma^2), \quad \theta \sim \mathcal{N}(0, \tau^2 I),$$

yielding the posterior predictive distribution

$$p(y_* \mid X_*, \mathcal{D}) = \mathcal{N}\big(X_*^\top\hat{\theta},\, X_*^\top\Sigma X_* + \sigma^2\big).$$

This formulation provides not only a point forecast but also calibrated confidence intervals. In practice, narrower bands (e.g., in Boro under stable irrigation) indicate greater reliability, while wider bands (e.g., in T.Aman under volatile monsoon conditions) reveal high environmental uncertainty.

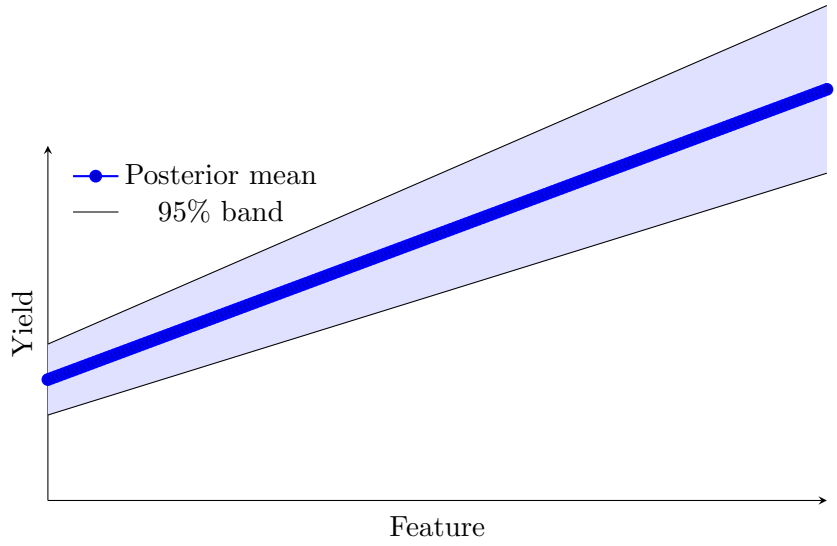


Figure 3: Bayesian uncertainty. The solid blue line is the posterior mean; the shaded region denotes the 95% predictive band. For rice yield prediction, such intervals provide actionable insight: stable ecotypes (Boro) exhibit tighter bands, while volatile ecotypes (T.Aman) display wider ones, highlighting where forecasts should be treated with caution.

Information-theoretic view. We assess the characteristic signal through mutual information, estimated with k-NN methods, complementing correlation-based diagnostics.

$$I(X;y) = \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right], \tag{6}$$

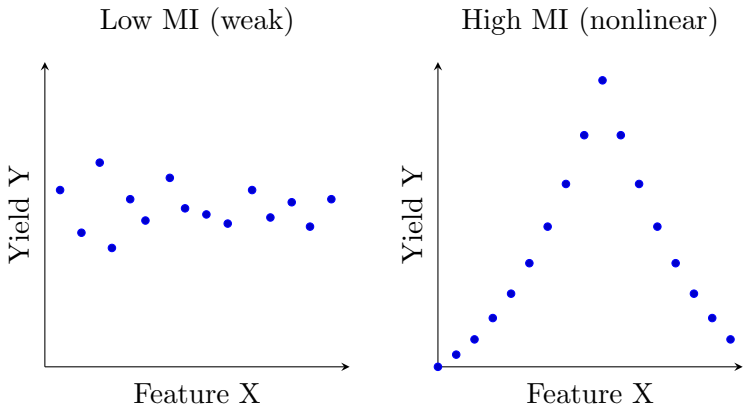


Figure 4: Mutual information: low-MI cloud (left) vs. high-MI, nonlinear dependency (right).

4 Methodology

Our methodology integrates multi-source data, leakage-safe preprocessing, model benchmarking, and deployment. Figure 5 summarizes the pipeline from acquisition to operational use.

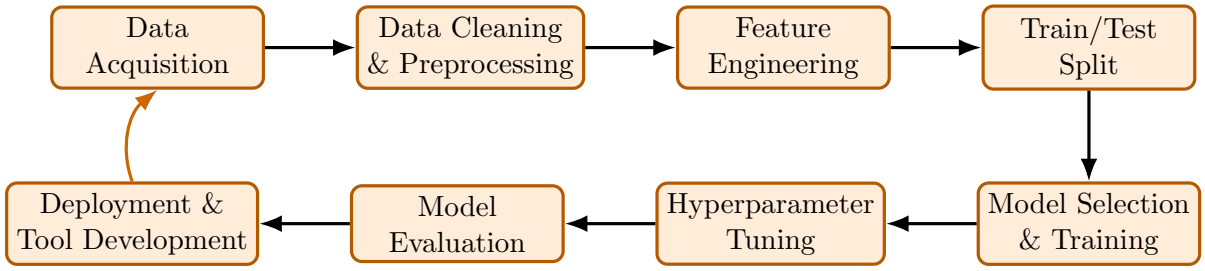


Figure 5: Pipeline process with iterative feedback loop.

Data collection. We integrate district-level management records (N, P, K, S application rates in kg/ha; irrigation in mm; biological nitrogen fixation) with NASA POWER daily weather series (mean temperature in °C, precipitation in mm, relative humidity in %). The study covers Bogura and Cumilla over 2020–2021, spanning Boro, T.Aus, and T.Aman.

Preprocessing. Weather series are aligned to crop calendars, and for each stage (vegetative, reproductive, ripening) we compute cumulative rainfall and mean temperature and humidity,

$$\bar{w}_{\text{stage}} = \frac{1}{|\mathcal{T}_{\text{stage}}|} \sum_{t \in \mathcal{T}_{\text{stage}}} w_t, \qquad W^{\text{cum}}_{\text{stage}} = \sum_{t \in \mathcal{T}_{\text{stage}}} w_t.$$

Rows with critical missing values are removed; clear data entry errors are excluded while genuine extremes are retained. Continuous features are z-normalized and categorical variables one-hot encoded. All steps are encapsulated in `scikit-learn` `Pipeline`s fitted only on training folds to prevent leakage.

Feature engineering. We include nutrient balances (e.g., N/K), an irrigation-to-rainfall ratio as a proxy for water reliance, and explicit interactions such as $N \times$ Rainfall and Temp $\times$ Irrigation that reflect agronomic mechanisms.

Model families and training. We benchmark Ordinary Least Squares, Ridge, Lasso, Support Vector Regression (RBF kernel), K-Nearest Neighbors, Random Forest, Gradient Boosting, and XGBoost. For each season we perform an 80/20 train–test split stratified by district and tune hyperparameters with 5-fold cross-validation. We report RMSE, MAE, and R^2 on held-out test sets and compute 95% prediction intervals via nonparametric bootstrap:

$$[\hat{y}_{\text{lower}}, \hat{y}_{\text{upper}}] = \left[\text{Quantile}_{0.025}, \text{Quantile}_{0.975}\right] \{\hat{y}_b(x)\}_{b=1}^{B}.$$

Interpretability. We rely on permutation importance, correlation heatmaps, and residual diagnostics; tree models can be complemented with SHAP for local attributions.

5 Results and Analysis

Season-specific performance. Different inductive biases dominate by ecotype (Table 1). In Boro, linear regression achieves R^2=0.878 with low RMSE, consistent with controlled irrigation and near-linear dynamics. In T.Aus, SVR captures nonlinear rainfall–temperature responses (R^2=0.862). In T.Aman, KNN adapts to local heterogeneity under monsoon variability (R^2=0.857).

Table 1: Test-set performance across seasons (best per season in bold).

Season (Best Model)	RMSE	MAE	R^2
Boro (Linear Regression)	**0.422**	0.338	**0.878**
T.Aus (SVR, RBF kernel)	**0.471**	0.385	**0.862**
T.Aman (KNN, k=5)	**0.494**	0.401	**0.857**

Residual diagnostics. Residuals are symmetric and homoscedastic in Boro (Fig. 6), indicating linear adequacy. T.Aus shows under-prediction in extreme rainfall years (Fig. 7), consistent with unmodeled higher-order climate shocks. T.Aman has the widest residual spread (Fig. 8), reflecting monsoon-driven heterogeneity and possible unobserved factors.

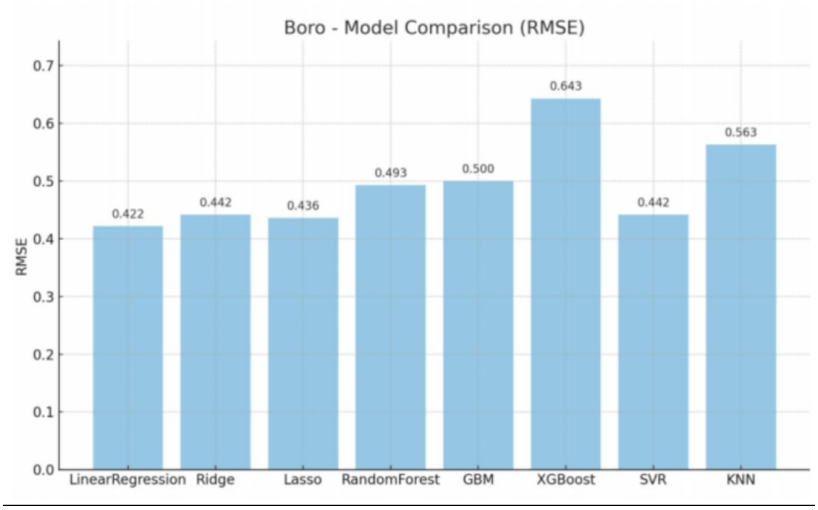


Figure 6: Residual distribution for Boro (Linear Regression): symmetry and low spread indicate linear adequacy.

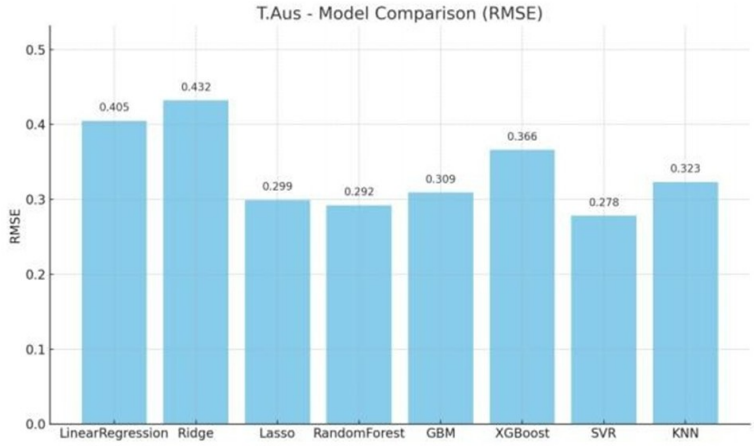


Figure 7: Residual distribution for T.Aus (SVR): heavier negative tails correspond to extreme rainfall under-prediction.

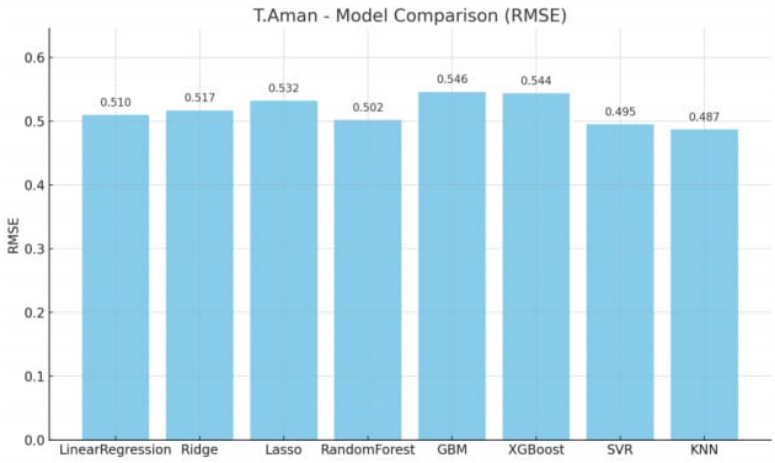


Figure 8: Residual distribution for T.Aman (KNN): wide dispersion reflects local heterogeneity under monsoon conditions.

Temporal and spatial generalization. Performance is lower in 2021 than in 2020, i.e., $\Delta R^2 = R^2_{2021} - R^2_{2020} < 0$, aligning with rainfall anomalies (Fig. 9). Spatially, Boro generalizes better in Bogura (strong irrigation infrastructure), while T.Aman generalizes better in Cumilla (greater varietal flood resilience) (Fig. 10).

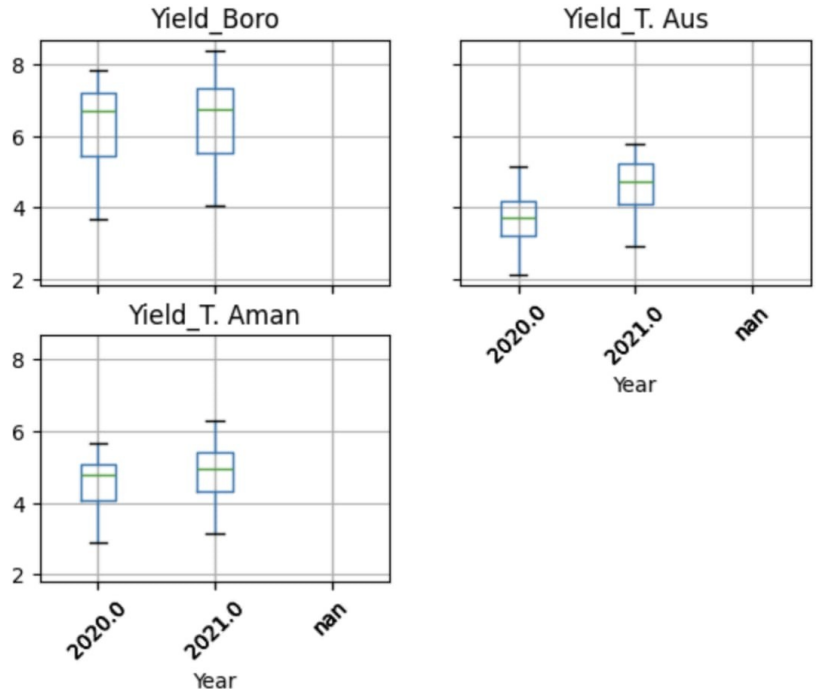


Figure 9: Temporal comparison (2020 vs. 2021): reduced accuracy in 2021 corresponds to anomalous rainfall.

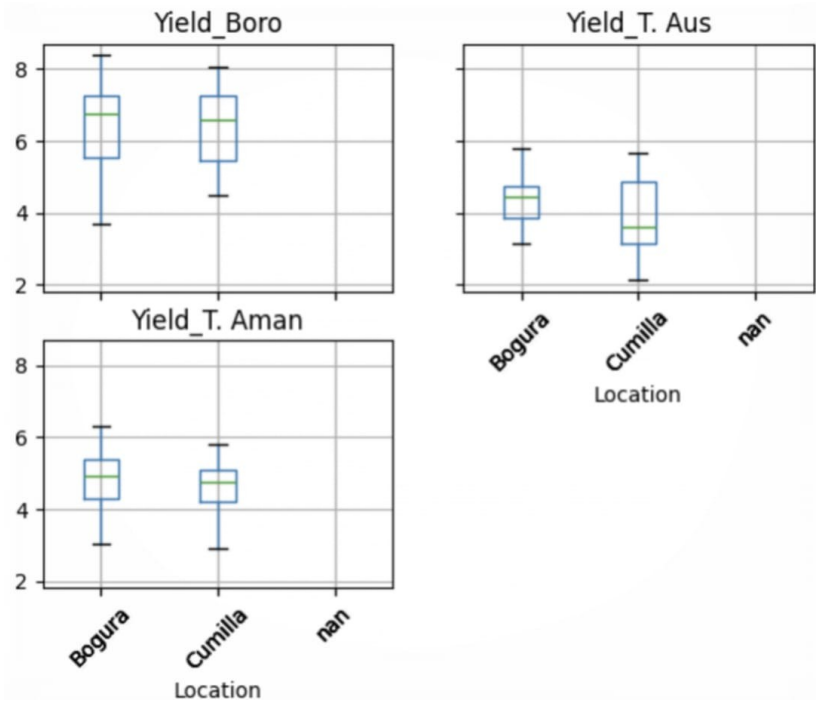


Figure 10: Spatial comparison (Bogura vs. Cumilla): patterns align with irrigation capacity and flood resilience.

Feature importance and drivers. Permutation importance shows irrigation and nitrogen dominate Boro; reproductive-stage temperature and rainfall dominate T.Aus; and nitrogen–rainfall interactions dominate T.Aman, consistent with leaching risks. The heatmaps in Figs. 11–13 summarize linear associations that are coherent with these findings.

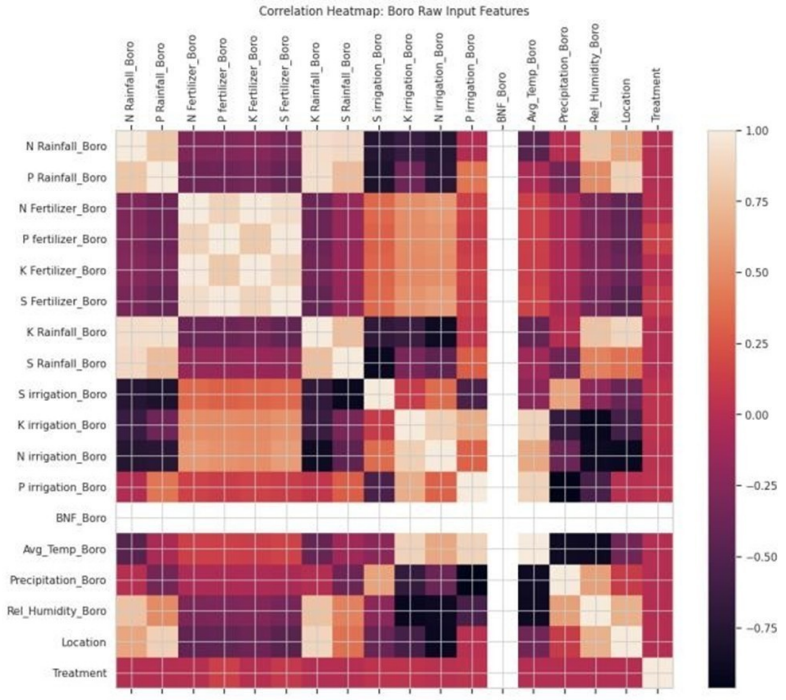


Figure 11: Correlation heatmap—Boro: strong associations of irrigation and nitrogen with yield.

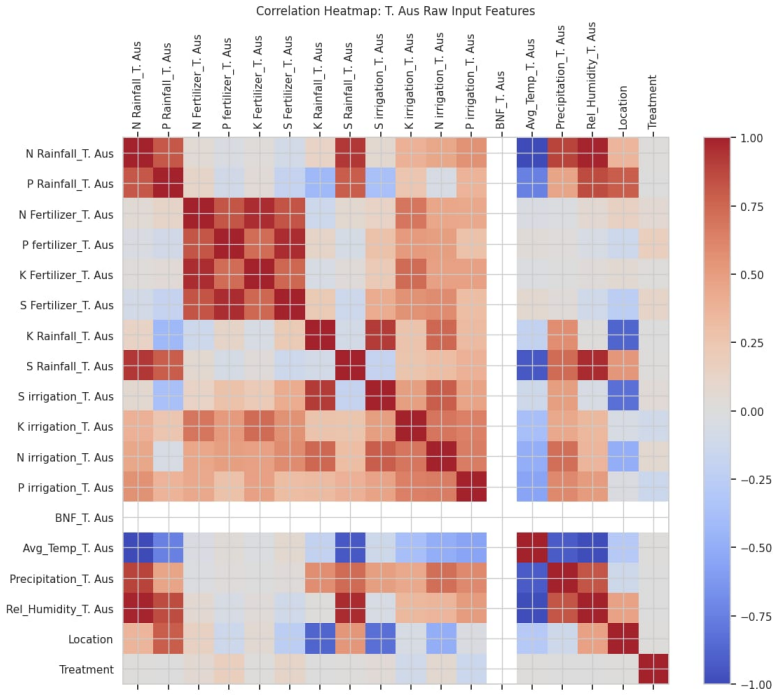


Figure 12: Correlation heatmap—T.Aus: reproductive-stage climate variables dominate.

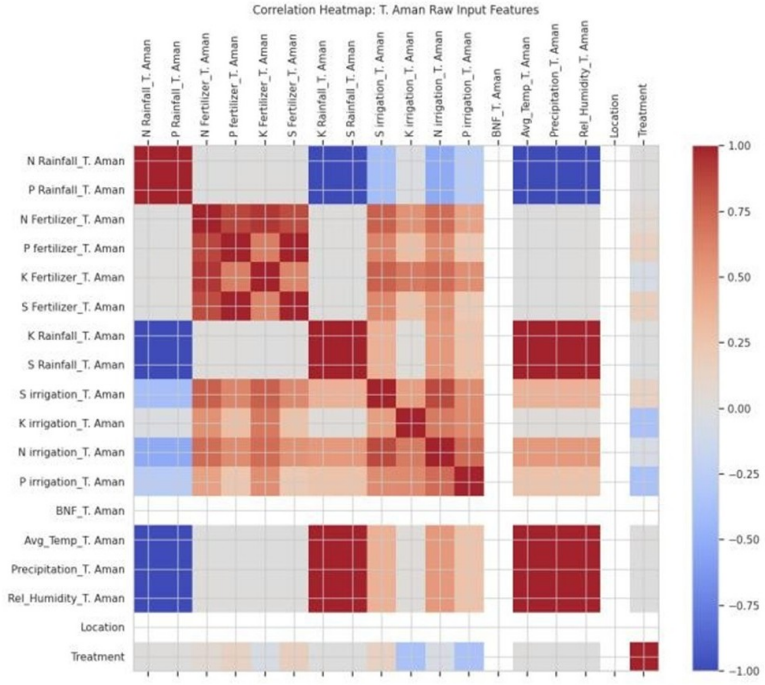


Figure 13: Correlation heatmap—T.Aman: nitrogen–rainfall interplay is most pronounced.

Uncertainty quantification. Bootstrap intervals are tightest in Boro (about ±0.2 t/ha), moderate in T.Aus (about ±0.35 t/ha), and widest in T.Aman (about ±0.5 t/ha). These bounds

provide actionable worst-case estimates for planners.

6 Discussion and Theoretical Insights

Bias–variance considerations explain the season-specific winners: linear models minimize error in Boro where irrigation suppresses variance; SVR reduces bias in T.Aus by capturing nonlinear climate effects; and KNN adapts to local heterogeneity in T.Aman. Information-theoretic analysis quantifies feature signal: irrigation and N carry the most information in Boro, reproductive-stage rainfall and temperature in T.Aus, and nitrogen–rainfall interactions in T.Aman. A structural equation model,

$$Y = \beta_0 + \beta_1 N + \beta_2 \text{Irrigation} + \beta_3 \text{Rainfall} + \beta_4 (N \times \text{Rainfall}) + U,$$

clarifies intervention pathways across seasons. Multi-task sharing via $\theta^{(s)} = \theta^{(0)} + \Delta^{(s)}$ explains cross-season commonalities (e.g., nitrogen's role) while allowing deviations (e.g., irrigation's season-specific importance).

7 Limitations and Future Directions

Our dataset spans only two years and two districts, limiting characterization of low-frequency climate signals and external validity. Domain adaptation and hierarchical modeling can extend generalization. Omitted variables (soil, pests, cultivars) may confound attributions; multimodal fusion with remote sensing and soil maps is a natural extension. Bootstrap intervals may understate tail risk; conformal prediction and Bayesian neural networks could provide calibrated coverage. Finally, moving from predictive correlation to causal inference (e.g., double machine learning) would support policy interventions. At scale, federated and online learning, together with cloud-native deployment, can power national dashboards and early-warning systems.

8 Conclusion

This paper introduced a season-specific machine learning framework for rice yield prediction that integrates district-level agronomic inputs with high-resolution NASA POWER weather data. By explicitly modeling the Boro, T.Aus, and T.Aman ecotypes as distinct conditional distributions, we demonstrated that different inductive biases are optimal in different regimes: linear regression in irrigated Boro, kernel methods in climate-sensitive T.Aus, and local instance-based methods in heterogeneous T.Aman.

Beyond empirical performance, we provided a theoretical foundation through bias–variance decomposition, mutual information analysis, and causal structural modeling, thereby aligning machine learning with agronomic reasoning. Our use of bootstrap ensembles and Bayesian baselines delivered calibrated uncertainty estimates, ensuring forecasts can be used in risk-sensitive planning. Finally, the deployment of a lightweight Streamlit tool bridged the research-to-policy gap, enabling agronomists and decision-makers to perform real-time scenario testing such as fertilizer adjustments or rainfall shocks.

The implications extend beyond rice in South Asia: the methodological principles—matching inductive bias to ecological regime, quantifying predictive uncertainty, and coupling interpretability with deployment—represent a transferable blueprint for smallholder-dominated agricultural systems worldwide. Future work should scale the temporal horizon, incorporate multimodal data

(soil, genetics, remote sensing), and advance toward foundation models for agriculture that are robust, interpretable, and causally valid. In doing so, machine learning can evolve from retrospective analysis into actionable intelligence for climate-resilient food security.

References

[1] FAO, "The State of Food Security and Nutrition in the World 2022," Rome, 2022. https://www.fao.org/publications/sofi/en/.

[2] J. W. Jones, G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie, "The DSSAT cropping system model," *European Journal of Agronomy*, vol. 18, no. 3–4, pp. 235–265, 2003. https://doi.org/10.1016/S1161-0301(02)00107-7.

[3] B. A. Keating, P. S. Carberry, G. L. Hammer, et al., "An overview of APSIM, a model designed for farming systems simulation," *European Journal of Agronomy*, vol. 18, no. 3–4, pp. 267–288, 2003. https://doi.org/10.1016/S1161-0301(02)00108-9.

[4] T. Li, O. Angeles, O. Marcaida III, A. Manalo, S. Manalili, S. Radanielson, and A. Mohanty, "ORYZA (v3): An improved rice model for simulating management practices and varietal response," *Agricultural and Forest Meteorology*, vol. 237–238, pp. 246–256, 2017. https://doi.org/10.1016/j.agrformet.2017.02.025.

[5] D. B. Lobell, G. L. Hammer, C. McLean, C. Messina, M. Roberts, and W. Schlenker, "The critical role of extreme heat for maize production in the United States," *Nature Climate Change*, vol. 5, pp. 817–821, 2015. https://doi.org/10.1038/nclimate2638.

[6] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian process for crop yield prediction based on remote sensing data," *Proceedings of the National Academy of Sciences*, vol. 114, no. 33, pp. 9156–9161, 2017. https://doi.org/10.1073/pnas.1702766114.

[7] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018. https://doi.org/10.1016/j.compag.2018.02.016.

[8] M. A. A. Al Mamun, M. M. Rahman, M. Z. Rahman, and M. A. Hasan, "Spatio-temporal variability of climatic variables and its impacts on rice yield in Bangladesh," *Frontiers in Sustainable Food Systems*, vol. 7, p. 1290055, 2023. https://www.frontiersin.org/articles/10.3389/fsufs.2023.1290055/full.

[9] S. Bhattacharjee, et al., "Machine learning based analysis and prediction of crop yield and prices of Aman, Aus and Boro rice," M.S. thesis, Brac University, 2021. https://dspace.bracu.ac.bd/xmlui/handle/10361/22424.

[10] B. Ghose, B. Saha, and B. Rahman, "Rice yield responses in Bangladesh to large-scale atmospheric oscillation using multifactorial model," *Theoretical and Applied Climatology*, vol. 146, no. 1, pp. 29–44, 2021. https://link.springer.com/article/10.1007/s00704-021-03725-7.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[13] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed., Cambridge Univ. Press, 2009.

[14] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, p. 066138, 2004. https://doi.org/10.1103/PhysRevE.69.066138.

[15] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, vol. 21, no. 1, pp. C1–C68, 2018. https://doi.org/10.1111/ectj.12097.

[16] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. https://arxiv.org/abs/1705.07874.

