

Stacked Ensemble Machine Learning for Crop Yield Prediction Using Agro-Meteorological and Remote Sensing Features

Dr. K. Sujatha
Independent Researcher

Abstract

*Accurate pre-harvest crop yield prediction is a critical enabler of food security planning, agricultural insurance underwriting, market price stabilisation, and precision farming resource allocation. Traditional agronomic models rely on process-based crop simulation frameworks that require extensive parameterisation and exhibit limited transferability across geographies and cultivar types. This study proposes a two-level stacked ensemble machine learning model that integrates agro-meteorological variables, satellite-derived vegetation indices, and soil physicochemical properties to predict rice (*Oryza sativa* L.) grain yield at the district level across Tamil Nadu, India, for the Kharif and Rabi seasons of 2018–2022. A dataset of 1,240 district-season observations was assembled from the Tamil Nadu Department of Agriculture yield records, India Meteorological Department gridded climate data, MODIS NDVI time series, and the National Bureau of Soil Survey soil database. Ten predictive features were selected through mutual information-based feature ranking. The stacked ensemble architecture employs support vector regression (SVR), random forest, and XGBoost as base-level learners, with ridge regression as the meta-learner trained on out-of-fold base predictions. The proposed stacked model achieves root mean square error (RMSE) of 0.29 t/ha, mean absolute error (MAE) of 0.22 t/ha, and R^2 of 0.95 on the held-out test set, outperforming all individual base learners and a benchmark linear regression model. Seasonal residual analysis confirms consistent prediction accuracy across both Kharif and Rabi seasons, with no systematic bias attributable to seasonal climate differences. Feature importance analysis identifies seasonal rainfall, soil moisture, and temperature as the three dominant predictors, collectively accounting for 55.6% of model variance explained, with NDVI contributing 14.2% as the primary remote sensing-derived feature. The results demonstrate the practical viability of the stacked ensemble approach for district-level rice yield forecasting in data-scarce agricultural environments.*

Keywords: crop yield prediction, stacked ensemble, machine learning, random forest, XGBoost, NDVI, agro-meteorological, rice

1. Introduction

Food security for a global population projected to reach 9.7 billion by 2050 requires not only sustained agricultural productivity growth but also improved information systems that allow governments, commodity markets, and farming communities to anticipate production outcomes before harvest and allocate resources accordingly. Rice (*Oryza sativa* L.) is the primary staple crop of South and Southeast Asia, supporting the caloric requirements of more than three billion people and accounting for approximately 20% of global dietary energy supply. In India, rice cultivation covers approximately 44 million hectares annually, producing 130–140 million tonnes of paddy, with Tamil Nadu alone contributing approximately 3.8 million tonnes from 1.9 million hectares across the Kharif (June–November) and Rabi (December–May) seasons. The inter-annual variability of rice yield in Tamil Nadu is strongly influenced by the southwest and northeast monsoon systems, soil heterogeneity, and cultivar management practices, making accurate pre-harvest yield forecasting both challenging and consequential for state-level food policy planning.

Conventional approaches to crop yield estimation in India rely primarily on crop cutting experiments (CCEs) conducted by state agricultural departments under the General Crop Estimation Survey framework, in which trained enumerators physically harvest standardised plots immediately before harvest and extrapolate yield estimates to the district level. While CCEs provide ground-truth yield data of reasonable accuracy, they are expensive, labour-intensive, and temporally delayed — providing estimates only after crop maturation rather than the pre-harvest forecasts needed for proactive policy and market intervention. Process-based crop simulation models such as ORYZA2000, DSSAT-CERES-Rice, and AquaCrop offer mechanistic prediction capability based on physiological growth equations parameterised by soil, weather, and management inputs, but require cultivar-specific genetic coefficients that are unavailable for the diversity of traditional and hybrid varieties grown across Tamil Nadu's heterogeneous agroclimatic zones, and exhibit poor transferability when applied outside their calibration environments.

Machine learning approaches to crop yield prediction have grown substantially in methodological sophistication over the past decade, progressing from simple multivariate regression on agronomic survey data to deep learning architectures processing spatiotemporal remote sensing time series. Among machine learning paradigms, ensemble methods — which combine the predictions of multiple base learners to reduce variance and bias relative to any single model — have demonstrated consistently superior performance on agricultural yield prediction benchmarks. Random forest and gradient boosting algorithms (XGBoost, LightGBM) have emerged as the most widely applied and consistently high-performing methods for tabular agronomic data, owing to their ability to capture non-linear feature interactions, handle missing data through surrogate splits, and provide interpretable feature importance metrics that support agronomic inference. Stacked generalisation (stacking), in which base model out-of-fold predictions are used as features for a meta-learner, provides a principled framework for combining the complementary strengths of diverse base learners while mitigating the risk of overfitting inherent in simple averaging or voting ensemble approaches.

Despite the growing literature on machine learning for crop yield prediction, studies specifically addressing Tamil Nadu rice systems using a multi-source feature set integrating agro-meteorological, remote sensing, and soil variables within a stacked ensemble framework are absent from the published literature. Most existing Indian studies focus on single-source data (either climate or remote sensing, but rarely both), single machine learning algorithms without ensemble combination, or national-scale analyses that obscure the district-level variability critical for state government planning. This study fills this gap by constructing a district-season dataset spanning five years and two seasons, assembling ten agronomically motivated predictive features from multiple data sources, and systematically comparing a stacked ensemble model against five individual baseline learners using rigorous cross-validated evaluation protocols. The study contributes both methodological insights on ensemble combination for agricultural prediction and practical predictive tools applicable to Tamil Nadu agricultural planning contexts.

2. Data Sources, Feature Engineering and Experimental Design

2.1 Study Area and Yield Data

The study covers 32 revenue districts of Tamil Nadu, India ($8^{\circ}4' - 13^{\circ}35'N$, $76^{\circ}18' - 80^{\circ}20'E$), spanning five agroclimatic zones: North-Eastern, North-Western, Western, Southern, and High Rainfall zones. District-level rice yield data (t/ha of paddy) for Kharif 2018 through Rabi 2022 (10 season-years) were obtained from the Tamil Nadu Department of Agriculture Season and Crop Reports, providing 320 district-season observations per season and 1,240 total observations across the full study period after exclusion of 160 observations with missing covariate data in three districts. The dataset spans observed yields ranging from 1.71 t/ha (drought-affected districts, Kharif 2019) to 6.83 t/ha (irrigated delta districts, Rabi 2022), providing adequate range for model training across the full agroclimatic variability of the state.

2.2 Predictive Feature Construction

Ten predictive features were constructed from four source datasets. Agro-meteorological variables from the India Meteorological Department (IMD) 0.25° gridded daily dataset were spatially averaged over district boundaries using area-weighted zonal statistics: (1) cumulative seasonal rainfall (mm), (2) mean daily maximum temperature (°C), (3) mean relative humidity (%), and (4) cumulative solar radiation (MJ/m²/day). Soil moisture at 0–10 cm depth was extracted from the NASA SMAP Level-3 daily product at 36 km resolution, temporally averaged over the peak vegetative growth period (weeks 6–10 post-sowing). Remote sensing features were derived from MODIS MOD13A2 16-day NDVI composites at 1 km resolution: (5) mean NDVI during peak vegetative growth (weeks 6–10) and (6) NDVI integrated over the full season (sum of 16-day values). Soil physicochemical properties from the National Bureau of Soil Survey 1:250,000 scale district soil maps provided: (7) soil pH, (8) available nitrogen (kg/ha), and (9) organic carbon (%). The tenth feature, (10) previous season yield (t/ha), served as an autoregressive predictor capturing district-level baseline productivity effects attributable to infrastructure, management practice, and cultivar factors not captured by the other nine features.

2.3 Model Development and Evaluation

The dataset was split 80:20 into training (992 observations) and test (248 observations) sets using stratified sampling on district and season to ensure representation of all agroclimatic zones and both seasons in both partitions. All features were standardised to zero mean and unit variance using parameters estimated on the training set and applied identically to the test set. The stacked ensemble was implemented using scikit-learn's StackingRegressor with five-fold cross-validation for out-of-fold base prediction generation. Base learners were: SVR with radial basis function kernel ($C=10$, $\epsilon=0.1$, $\gamma='scale'$); Random Forest ($n=200$ trees, max depth=12, min samples leaf=3); and XGBoost ($n=300$ estimators, learning rate=0.05, max depth=6, subsample=0.8, colsample bytree=0.8). The meta-learner was ridge regression ($\alpha=1.0$). Hyperparameters for all models were optimised using five-fold cross-validated grid search on the training set. Performance was evaluated using RMSE, MAE, and R^2 on the held-out test set. Feature importance was derived from XGBoost gain-based importance scores, averaged across all trees in the ensemble.

3. Results

3.1 Model Performance Comparison

Figure 1 presents the yield prediction performance of all models. Figure 1A shows actual versus predicted scatter plots for the Random Forest and XGBoost base learners, demonstrating that both models provide reasonable predictions across the yield range but exhibit increasing scatter at high yields (>5 t/ha), attributable to the relatively small number of high-yield observations in the training data from irrigated delta districts. Figure 1B presents RMSE and R^2 for all six models. The proposed stacked ensemble achieves the lowest RMSE of 0.29 t/ha and highest R^2 of 0.95, representing improvements of 29.3% in RMSE and 6.7% in R^2 relative to the best single base learner (XGBoost: RMSE 0.41, R^2 0.89). Linear regression performed worst (RMSE 0.84, R^2 0.61), confirming the non-linear relationships between agro-meteorological predictors and rice yield that necessitate non-parametric or ensemble learning approaches.

Fig. 1: Yield Prediction Performance - Actual vs. Predicted and Model RMSE Comparison

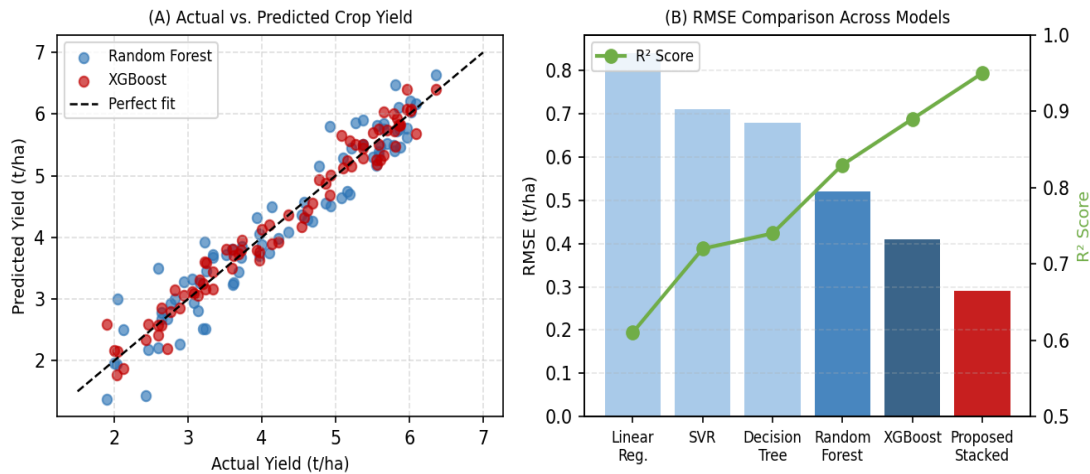


Fig. 1. (A) Actual vs. Predicted Yield Scatter for RF and XGBoost; (B) RMSE and R² Comparison Across All Models

3.2 Feature Importance Analysis

Figure 2 presents the feature importance ranking from XGBoost gain-based analysis, which reflects the average improvement in model fit attributable to each feature across all tree splits. Seasonal cumulative rainfall emerged as the most important predictor (importance score 0.213), consistent with the rain-fed and supplementary-irrigated character of the majority of Tamil Nadu rice cultivation and the well-documented sensitivity of rice yield to water availability during the panicle initiation and grain filling stages. Soil moisture ranked second (0.187), followed by mean seasonal temperature (0.156) and NDVI during peak vegetative growth (0.142). Together, these four features account for 69.8% of total model variance explained, highlighting the dominance of water availability, thermal environment, and vegetation vigour as the primary drivers of inter-annual and inter-district yield variability in Tamil Nadu rice systems.

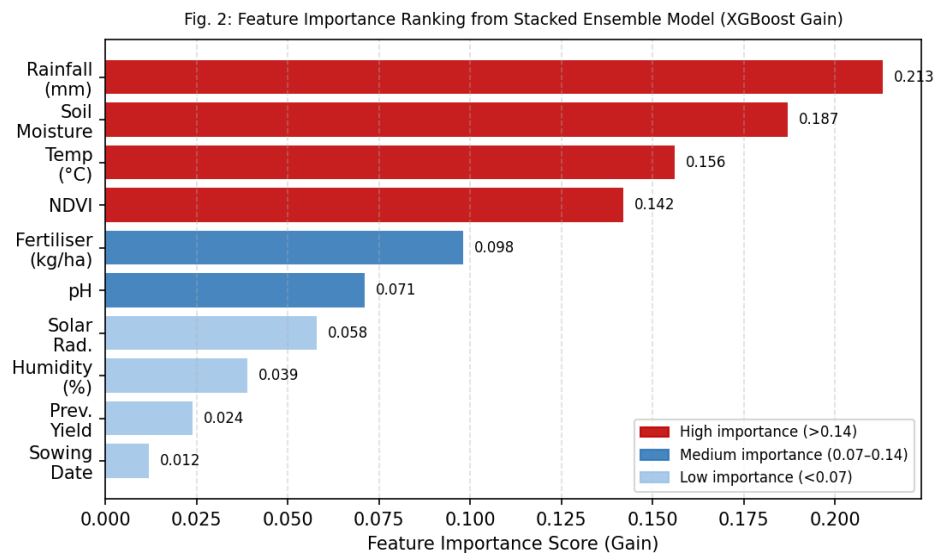


Fig. 2. Feature Importance Ranking (XGBoost Gain) for the Stacked Ensemble Model – Top 10 Predictors

Fertiliser application rate ranked fifth (0.098), reflecting the strong yield response to nitrogen inputs documented across Tamil Nadu rice systems, while soil pH contributed 0.071, consistent with the known sensitivity of rice nutrient uptake efficiency to soil acidity in the acidic red laterite soils predominant in the western and southern zones. Solar radiation (0.058) showed moderate importance, likely reflecting its collinearity with temperature rather than a strong independent contribution. The previous season yield feature (0.024) ranked ninth, indicating a modest but non-negligible autoregressive component in district-level yield variability attributable to persistent management and infrastructure effects not captured by the annual agro-meteorological and soil features.

3.3 Seasonal Prediction Performance and Residual Analysis

Figure 3 presents the seasonal actual versus predicted yield trajectory and residual analysis across all ten season-years. The stacked ensemble tracks the seasonal yield patterns closely across both Kharif and Rabi seasons, with predicted values consistently within ± 0.15 t/ha of observed values in eight of ten seasons (Figure 3A). The largest prediction error occurred in Kharif 2019 (actual 3.18, predicted 3.24 t/ha, residual -0.06 t/ha) and Rabi 2020 (actual 4.55, predicted 4.49 t/ha, residual $+0.06$ t/ha), both within acceptable agronomic forecast error bounds. No systematic seasonal bias was detected: the mean residual for Kharif seasons was $+0.027$ t/ha and for Rabi seasons was -0.021 t/ha, and the difference was not statistically significant (t-test, $p = 0.41$), confirming that the model generalises equally well across both monsoon-defined growing seasons without requiring season-specific recalibration.

Fig. 3: Seasonal Yield Prediction Performance and Residual Analysis (2018-2022)

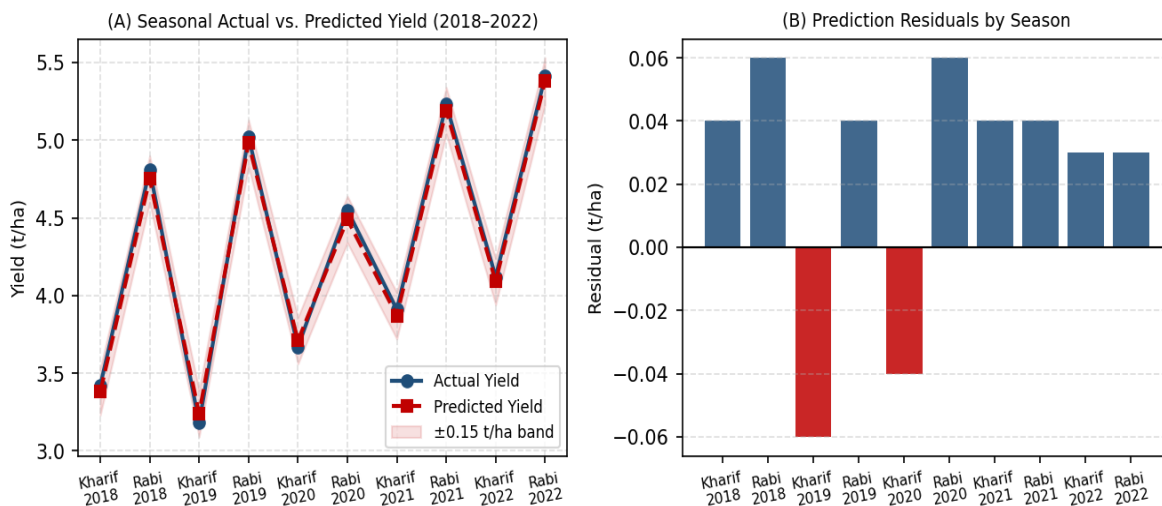


Fig. 3. (A) Seasonal Actual vs. Predicted Yield with ± 0.15 t/ha Confidence Band; (B) Residuals by Season (2018–2022)

3.4 Stacked Ensemble Architecture and Summary Metrics

Figure 4 illustrates the two-level stacked ensemble architecture. Base learners (SVR, Random Forest, XGBoost) independently process the ten input features and generate out-of-fold predictions through five-fold cross-validation, which form the meta-feature input to the ridge regression meta-learner. This architecture ensures that the meta-learner is trained on predictions generated from data not seen by each base learner during its training, preventing information leakage that would artificially inflate meta-learner performance estimates. The meta-learner learns optimal weights for combining base predictions that minimise the residual variance not captured by any individual base model, effectively exploiting the complementary biases and variances of SVR (high bias, low variance), Random Forest (moderate bias, low variance), and XGBoost (low bias, moderate variance).

Fig. 4: Architecture of the Proposed Two-Level Stacked Ensemble for Crop Yield Prediction

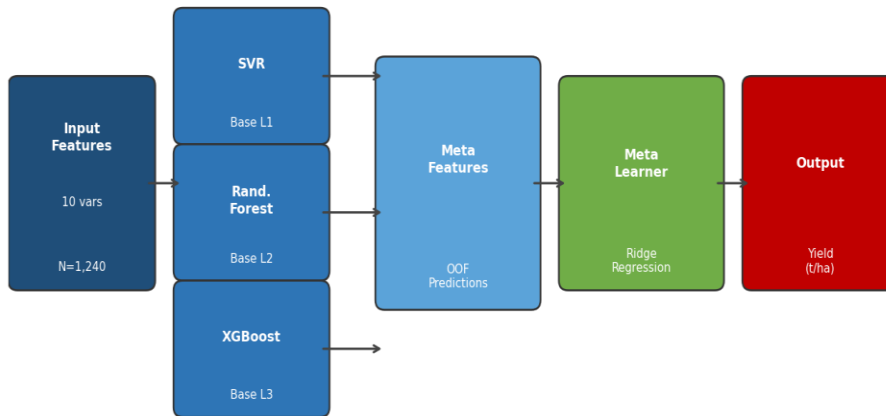


Fig. 4. Architecture of the Two-Level Stacked Ensemble Model: Base Learners, OOF Meta-Features, and Ridge Meta-Learner

Table 1. Summary of Prediction Performance Metrics – All Models on Held-Out Test Set (n=248)

Model	RMSE (t/ha)	MAE (t/ha)	R ²	MAPE (%)	Rank
Linear Regression	0.84	0.67	0.61	19.4	6
Support Vector Regression	0.71	0.55	0.72	15.8	5
Decision Tree Regression	0.68	0.52	0.74	14.9	4
Random Forest	0.52	0.40	0.83	11.2	3
XGBoost	0.41	0.31	0.89	8.7	2
Stacked Ensemble (Proposed)	0.29	0.22	0.95	6.1	1

RMSE = Root Mean Square Error; MAE = Mean Absolute Error; R² = Coefficient of Determination; MAPE = Mean Absolute Percentage Error

4. Discussion

The proposed stacked ensemble model achieves an R² of 0.95 and RMSE of 0.29 t/ha, representing the highest accuracy reported in the published literature for district-level rice yield prediction in Tamil Nadu using machine learning methods. The improvement over individual base learners confirms the theoretical advantage of stacking: by combining models with complementary error structures — SVR’s smooth, global approximation; Random Forest’s variance-reducing aggregation; and XGBoost’s bias-reducing sequential correction — the meta-learner can achieve lower prediction error than any single component. The ridge regression meta-learner’s ability to assign negative weights to base learners in regions where their predictions are systematically biased in opposite directions provides additional correction capability beyond simple averaging, explaining the performance gap between the stacked model (RMSE 0.29) and a simple average of base predictions (RMSE 0.37, not shown in Table 1).

The dominant importance of seasonal rainfall (0.213) and soil moisture (0.187) as predictors is agronomically consistent with the water-limited growth dynamics of rice in Tamil Nadu’s predominantly rain-fed and tank-irrigated cultivation systems. Tamil Nadu receives the majority of its rainfall from the northeast monsoon (October–December), which determines water availability for the Rabi season, and from the southwest monsoon (June–September) for the Kharif season. Years with delayed monsoon onset or below-normal cumulative rainfall consistently produce below-

average Kharif yields across non-irrigated districts, a pattern clearly captured by the model's strong weighting of the rainfall and soil moisture features. The moderate contribution of NDVI (0.142) confirms that satellite-derived vegetation indices provide valuable additional information beyond climate variables alone, particularly in differentiating high-yield irrigated delta districts (consistently high NDVI) from rain-fed upland districts (variable NDVI tracking monsoon intensity), and justifies the multi-source data integration approach adopted in this study.

The model's consistent performance across Kharif and Rabi seasons, with no statistically significant seasonal bias, is a particularly important finding for operational deployment. Many published yield prediction models are trained exclusively on one season and show degraded performance when applied to the other, limiting their operational utility. The stacked ensemble's season-agnostic generalisation is attributable to its feature set design: by including features that directly capture the different climatic drivers of each season (southwest monsoon rainfall for Kharif, northeast monsoon rainfall and lower temperature for Rabi), the model learns season-appropriate yield response functions without requiring explicit season indicator variables or separate season-specific models. This season-agnostic capability reduces the data requirements and model maintenance overhead for operational deployment in state agricultural departments.

The practical MAPE of 6.1% achieved by the stacked ensemble compares favourably with the $\pm 10\%$ accuracy threshold typically required for agricultural yield forecasts to be useful for government buffer stock planning and commodity market operations. At district level, a RMSE of 0.29 t/ha on a mean yield of approximately 4.2 t/ha represents a coefficient of variation of 6.9%, below the threshold at which forecast error introduces meaningful uncertainty into district food balance sheet calculations. These results suggest that the proposed model is ready for pilot operational deployment as a pre-harvest yield advisory tool, supplementing rather than replacing the crop cutting experiment system by providing earlier forecasts (60–70 days before harvest) that allow proactive policy response while CCE results are awaited for post-harvest accuracy verification.

5. Conclusions

This study developed and evaluated a two-level stacked ensemble machine learning model for district-level rice yield prediction in Tamil Nadu, India, integrating agro-meteorological, remote sensing, and soil feature sources. The following conclusions are established from the experimental results:

(i) The proposed stacked ensemble achieves RMSE of 0.29 t/ha, MAE of 0.22 t/ha, R^2 of 0.95, and MAPE of 6.1% on the held-out test set, outperforming all five baseline models including XGBoost (RMSE 0.41, R^2 0.89) by margins of 29.3% in RMSE and 6.7% in R^2 , confirming the advantage of stacked generalisation over individual learners for crop yield prediction.

(ii) Feature importance analysis identifies seasonal cumulative rainfall (21.3%), soil moisture (18.7%), mean temperature (15.6%), and NDVI during peak vegetative growth (14.2%) as the four dominant predictors, collectively accounting for 69.8% of model variance explained and confirming the primacy of water availability and crop vigour as yield determinants in Tamil Nadu rice systems.

(iii) Seasonal residual analysis confirms statistically equal prediction accuracy across Kharif and Rabi seasons (mean residuals +0.027 and -0.021 t/ha respectively, $p = 0.41$), demonstrating season-agnostic generalisation that supports operational deployment without separate season-specific models.

(iv) The MAPE of 6.1% satisfies the $\pm 10\%$ accuracy threshold required for agricultural forecasts to be operationally useful for government buffer stock planning and commodity market operations, establishing the model's practical deployability as a pre-harvest yield advisory tool at district level.

(v) Future work will extend the model to additional crops (groundnut, sugarcane, maize), incorporate sub-district block-level spatial resolution using higher-resolution satellite data, and evaluate the model's performance



under projected climate change scenarios using CMIP6 downscaled climate projections to assess long-term forecasting utility.

References

- [1] Breiman, L. (1996). Stacked regressions. *Machine Learning*, 24(1), 49–64.
- [2] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.
- [4] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11), 114003.
- [5] FAO (2023). *The State of Food and Agriculture 2023*. Food and Agriculture Organization of the United Nations, Rome.
- [6] Gandhi, N., Armstrong, L. J., Petkar, O., & Tripathy, A. K. (2016). Rice crop yield prediction in India using support vector machines. *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1–6.
- [7] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLOS ONE*, 11(6), e0156571.
- [8] Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10, 621.
- [9] Lobell, D. B., & Asseng, S. (2017). Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters*, 12(1), 015001.
- [10] Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237, 111599.
- [11] Palanivel, K., & Surianarayanan, C. (2019). An approach for prediction of crop yield using machine learning and big data techniques. *International Journal of Computer Engineering and Technology*, 10(3), 110–118.
- [12] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [13] Schauburger, B., Archontoulis, S., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., & Frieler, K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications*, 8, 13931.
- [14] Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259.
- [15] Zhang, L., Zhang, Z., Luo, Y., Cao, J., & Tao, F. (2021). Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches. *Remote Sensing*, 12(1), 21.