

SHAP-Backed Hybrid Ensemble Model for Rice and Wheat Forecasting in Data-Scarce Environments

Harendra Singh Negi

Department of Computer Science and Engineering,
Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India.
Corresponding author: mail.harendrasinghnegi@gmail.com

Sushil Chandra Dimri

Department of Computer Science and Engineering,
Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India.
E-mail: dimri.sushil2@gmail.com

(Received on August 7, 2025; Revised on November 15, 2025 & January 19, 2026 & February 25, 2026;
Accepted on February 28, 2026)

Abstract

As the issue of agricultural sustainability has continued to increase, there has been a need to use data based solutions to improve agricultural productivity. This paper proposes a machine learning system combining Random Forest and XGBoost to combine prediction-forecasting crop yield and classification of crop type rice and wheat in Indian state of Uttarakhand. The model is tested using a library of 6, 000 samples containing 12 soil and climatic characteristics and measured on regression and classification quality. The hybrid ensemble with hyperparameter optimization and verified on the basis of 10-fold cross-validation performed better than single base learners in all measures. It achieved a classification accuracy of 96.3 and $R^2 = 0.927$. Statistically significant developments that were formed using paired t-tests were set at $p = 0.05$. The SHAP and ablation analysis found out nitrogen, rainfall, and pH as the most influential features. The forecasted framework provides a better generalizability, interpretability, and computational effectiveness, which is appropriate to be applied in the designs of real life in precision agronomy. The new result is novel, interpretable, and high-performative to crop yield intelligence in data-scarce areas and provides a contribution to this study.

Keywords- Ablation, PCA, Hyperparameter, SHAP, Statistical validation.

1. Introduction

In many developing economies such as India, agronomy is still the food security and economic stability (Dethier and Effenberger, 2012; Pawlak and Kołodziejczak, 2020). The crops that take significant significance are rice and wheat being among the staple crops, since they are widely grown and thus will have an impact on the caloric consumption (Erenstein et al., 2022; Zafar and Jianlong, 2023). But as climatic variability, soil erosion, and fluctuating prices of inputs, agronomic planning and decision making in the past is no longer the one to guarantee sustainable production. The necessity to undergo agribusiness digital revolution has increased the requests of data-driven styles capable of providing correct, scalable, and timely predictions of a crop yield (Benti et al., 2024). The development of agriculture as a practice more than the conventional farming to mechanization, automation and currently smart farming depicts the unrelenting nature of the sector to accommodate changes in technology (Bazargani and Deemyad, 2024). This evolution is introduced in **Figure 1**, where the increase in the use of advanced computing tools, such as machine learning (ML), to predict agriculture and optimize resources, is highlighted. Recent advances in artificial intelligence (AI) and ML allowed more civilized modeling of compose agro-ecological systems (Padhiary et al., 2025). These techniques are capable of learning non-linear relationships among different variables like weather conditions, soil conditions, levels of input of fertilizers and crop production. Ensemble learning techniques have had a lot of potential by their capacity to minimize overfitting and enhance generalization.

Random Forest (RF) and Extreme Gradient Boosting (XGBoost) are among them which are widely applied due to their strength and ability to predict on structured agricultural data (Adjei et al., 2025).

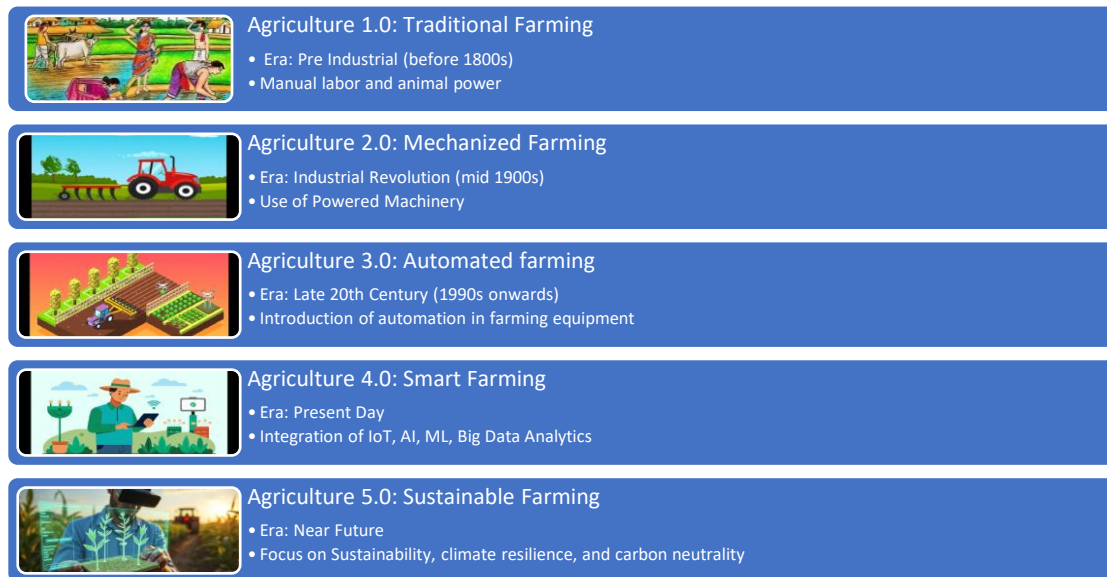


Figure 1. The evolution of agriculture from traditional systems to data-driven smart farming.

Nevertheless, the existing models tend to work independently either using RF, and XGBoost independently without taking the advantage of the complementary benefits. Besides, several existing studies do not consider key problems including spatial generalizability, model interpretability, and consistent model performance across different agro-climatic areas. To address these issues, the work proposes a hybrid model of machine learning to combine the (Random Forest with XGBoost) in a bid to make responses to predict rice and wheat yields on a district and state level, in India. The hybrid model will address the above by providing superior predictive accuracy, less regional bias and enhanced real-world applicability to the stability of bagging and the accuracy of boosting. The model is also tested to determine the strength and flexibility by tuning the hyperparameters with and without hyperparameter tuning. It works with a large, multi-annual dataset, which includes the elements of climate, soil characteristics, and spatial markers, including the state and district records.

The contribution of the study is mainly:

- Prediction of rice and wheat yield using a hybrid of Random Forest and XGBoost model and its performance evaluation against separate models (with and without hyperparameter tuning) and evaluation of generalization across the different districts of Uttarakhand.
- The multidimensional feature space is used to assess separability of rice and wheat crops using PCA.
- The addition of a game-theoretic SHAP method to the evaluation and visualization of the contribution of each individual feature to model forecasting on the entire dataset.
- Conducted an ablation experiment to establish what role a separate individual component plays in the predictive performance of the model and measure their relative contribution.
- From these results, I concluded that the hybrid ensemble model proposed in this study was very robust and generalizable, as evidenced by extensive statistical validation exercises.

The vast majority of crop forecasting models have limited robustness, explainability, or spatial generalization and only yield or classification. In such places as Uttarakhand, there is a definite necessity to have a dual-task, interpretable, and effective model to implement in the real world. Based on this necessity to have correct, interpretable and generalizable crop yield prediction models, the next section will conduct a literature review on the application of machine learning in predicting the yield of rice and wheat. It is a critical review of the applied models, data sources, main findings, and limitations highlighting the contextual background and emphasizing the gaps that the study is supposed to sell.

This paper makes a number of methodological and application based contributions that makes it stand out among past research of crop yields prediction. On the one hand, we propose a SHAP-directed ensemble weighting, according to which interpretability knowledge is used to refine the hybrid model instead of using random or equally weighted mixes. Second, we use an optimization-based hyperparameter tuning framework, which is Optuna, to obtain powerful settings of both the Random Forest and XGBoost when there is a lack of sufficient data. Thirdly, the article specifically deals with agro-ecological diversity of Uttarakhand by applying the concept of cross-geographical validation strategy in hilly and plains areas, where there is spatial generalisation because of constrained data presence. Fourth, the research combines the multi-level uncertainty analysis or the estimation of the prediction interval and fold-wise stability analysis to improve the reliability. These components assist in bettering traditional ensemble types of application and offer more comprehensible and context-proficient modelling systems to Himalayan agrifood settings.

This research focuses on analysis and implementation, as compared to creating a novel machine learning algorithm. As an alternative to that, its contributions are strategically placed on three complementary planes to prevent boastful assertions.

- The study has an empirical analysis of a weighted Random Forest-XGBoost ensemble as an approach to crop yield prediction in data-sparse and geographically dissimilar environments. Its methodology is a combination of ensemble learning and post-hoc interpretability (SHAP) and an ablation-based stability analysis with no modifications to the learning algorithms used.
- A complete pipeline of end-to-end forecasting is implemented that combines data preprocessing, hyperparameter optimization, performing an ensemble regression, uncertainty analysis, and explainability in one, reproducible, framework.
- The framework is tested on Uttarakhand which is an agro-ecological region in Himalayas and thus it is an under-represented region in yield forecasting literature and thus not a universal generalizable solution.

Based on this, the given study can be viewed as applied, but region-specific empirical contribution instead of an entirely new modeling paradigm.

2. Related Work

The latest advancements in AI and ML have changed agricultural forecasting to a large extent, specifically in the field of crop yield forecasting. As more and more data is recorded by climatic sensors, satellite imagery and soil records, scholars are turning to more advanced forms of computing models in order to enhance prediction and provide support to agricultural decision-making with data. Although Random Forest, XGBoost, and deep learning architectures have been successfully used at the international level to study crops like rice and wheat, very little research has been done in southeast Asia especially in Vietnam using hybrid ML models incorporating the use of spatial, environmental, and agronomic parameters (Wu et al., 2023). This is a significant gap due to the agricultural significance of the region and the susceptibility of the region to changing climatic conditions. In this part, a summary of world and regional contributions to yield forecasting is given and the limitations of this approach is also brought out. It is based on these

insights that the hybrid RF-XGBoost method suggested in this work was built.

Table 1. Comparative summary of existing machine learning approaches for crop yield prediction.

Author(s) / citation	Model(s) used	Data source / Region	Main findings	Limitations / Gaps
Kadamed and Bansod (2025)	RF, XGBoost, SVR	Maharashtra, India	RF & XGB yielded high accuracy ($R^2 \approx 0.99$).	No hybridization; no detailed tuning analysis.
Wibisono and Saiful (2025)	RF, XGBoost	Indonesia	RF slightly outperformed XGB; both performed well ($R^2 \approx 0.96$).	Overfitting concerns; no spatial validation.
Abbas et al. (2020)	Proximal sensing + ML	Various field trials (global)	Good results using remote sensors and ML models.	Focused on sensor fusion; lacks ensemble comparison.
Leukel et al. (2023)	Systematic Review (RF, SVM, DL, XGB, etc.)	Global field-scale studies	Early grain yield prediction benefits from ML adoption.	Limited hybrid RF-XGB integration explored.
Khodjaev et al. (2024)	Optimized ML (RF, XGB) with UAV data	Uzbekistan	UAV-based feature extraction significantly improved ML estimation.	Complex data preprocessing; not compared with hybrid models.
Ashfaq et al. (2025)	Deep learning + satellite + climate fusion	Pakistan	AI fusion approach highly accurate for wheat yield ($R^2 > 0.95$).	Black-box nature limits interpretability; no RF/XGB hybrid ensemble.
Ha et al. (2024)	ViT + RF, SVR, MLP	An Giang & Trà Vinh provinces, Vietnam	ViT+RF achieved best MAE = 75.96.	Limited to image-based yield; no RF/XGB comparison.
Wan et al. (2020)	Ensemble ML (RF, DT) with UAV multispectral	Southern China (UAV grain fields)	UAV indices improved grain-yield predictions.	Geographic confinement; no deep hybridization.
Zhou et al. (2017)	ML on multitemporal UAV indices	China	Better accuracy using multiple vegetative indices.	Older study; lacks advanced ensemble modeling.
Son et al. (2020)	RF, SVM	Taiwan (satellite time-series)	Remote sensing + RF/SVM effective for rice yield.	Single-country scope; no hybrid model tested.
Sarkar et al. (2024)	Ensemble ML on UAV remote sensing	Vietnam	Improved rice yield estimates via UAV spectral data.	No integration of RF+XGB hybrid ensemble.
Pankaj et al. (2024)	Regression on 2D panicle images	Vietnam	Effective yield prediction from image features.	Narrow feature set; lacks ensemble comparison.
Mohanty et al. (2025)	Q-learning + Rule-based decision system	Simulated 50-year rotation yield data	Integrates MDP and Q-learning for adaptive crop rotation; yields improved by 36.13%, rewards by 2.04%, and Q-values by 35.73%.	Empirical field-scale validation and multi-crop generalization remain future work.
Dey et al. (2024)	Deep Q-Network (DQN) for energy-aware multi-path routing in Wireless Sensor Networks (WSNs)	Simulated agricultural field with soil-moisture sensors at multiple depths	DQN-based routing improved energy efficiency, achieving 3% higher residual energy and ~12% longer network lifetime compared to traditional hierarchical routing.	Focused only on energy efficiency; lacks integration with real field data and agronomic decision-making layers.
Mohanty et al. (2023)	Time-series models (AR, MA, ARMA, ARIMA, Exponential Smoothing) and ML models (Decision Tree Regressor)	Historical crop yield, import, demand, and price datasets	Developed a four-stage framework for yield, supply, demand, and price forecasting. The Decision Tree Regressor achieved the lowest RMSE for crop price prediction, outperforming statistical models.	Limited to price estimation; lacks integration with spatial or climatic factors and does not consider dynamic crop rotation or real-time data.

Although there has been significant advancement in the machine learning deployment in crop forecasting, the following limitations are significant across recent literature as presented in **Table 1**.

Although many studies have been conducted to investigate the role of machine learning models in predicting crop yield as demonstrated in **Table 1**, there are several gaps in the literature, especially with regard to ensemble modeling, spatial generalization, interpretability and robustness of the methods.

- Lack of hybrid Random Forest-XGBoost ensemble models in the literature; current literature endorses models separately and not in the form of performance comparisons on the tuned models and default settings.

- Minimal analysis of the external validity of models to other geographic areas; most of the research studies are not more than one place or within a specific area.
- No PCA used to investigate separability of features of rice and wheat in multidimensional input space.
- The absence of interpretability e.g. SHAP to enumerate the contributions of features; current literature treats models as black boxes.
- This is because no ablation experiments that were done to isolate and measure the contribution of solely model components to the overall predictive performance were done.
- Inadequate statistical validation in previous studies; the majority of them do not test robustness or significance, only accuracy.

On top of all these wide gaps, numerous other gaps can be seen in the existing literature. Even though such models built on deep learning have demonstrated their potential, they are associated with a high level of computational complexity and lack of transparency which represents a barrier to a viable implementation in low-resource agricultural environments. In addition, most of the approaches that are currently in use can be tested only over a limited geographical area or a limited temporal period thus concern itself with their generalizability across seasons and agro-climatic regions. Another conspicuous lack is in the form of comparative studies that are able to compare the model performance prior and after tuning of hyper parameters in terms of standard metrics like MAE, RMSE, and R. Fourth, although UAV and satellite imagery are on the rise, they have not been incorporated in the ensemble modelling systems with structured agronomic data, including nitrogen content, rainfall, soil pH among others.

Although other research systems have covered methods of machine learning to predict crop yield, most of them aim at the performance of individual models, or extensive datasets of agricultural areas with plenty of samples. This contribution contributes to the body of knowledge by presenting a powerful ensemble in the specific case of states in the Himalayas, in which both sparsity of data and heterogeneity in geography lead to major modelling difficulties. As opposed to the previous approaches, the suggested framework uses SHAP-based interpretability, cross-regional validation, and ensemble tuning through optimization. These methodological improvements offer practical results to make decisions regarding agriculture in regions and add to the new area of data-scarce agro-informatics.

Based on the lessons and constraints found by previous research, the section describes the data in this study, the methods of pre-processing, the model structures, and the evaluation procedures. Attention is particularly paid to the design and the creation of the hybrid ensemble of the Random Forest and XGBoost, the data preprocessing, visualization of the features, the study design that is designed to be as robust and interpretable as possible.

3. Materials and Methods

This paper provides distinct boundaries between the two prediction tasks at hand which are: (i) predicting crop yield where prediction involves an output variable taking the form of the actual yield of rice, wheat, and (ii) classifying crop type, which is considered a binary classification problem distinguishing between rice and wheat. These two activities work on two different output spaces and must be evaluated by different metrics and model settings as well as validation strategies. The methodology, consequently, handles those issues separately but based on the common features representation which is provided by soil, climatic and geographical features.

In this section, the dataset, features, and tools as well as machine learning structures to predict rice and wheat yield will be described based on the proposed hybrid ensemble model.

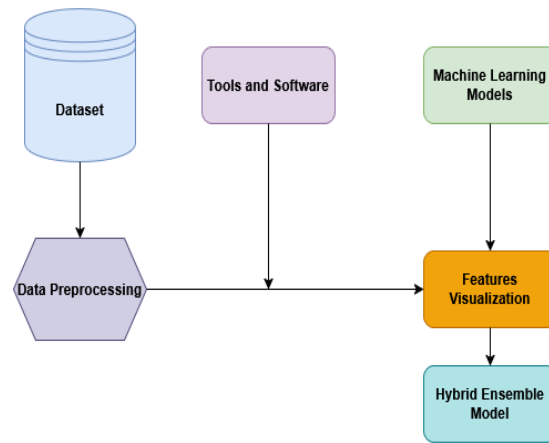


Figure 2. Schematic overview of the materials and methods adopted for hybrid model development.

The workflow process to be followed in the study is shown in **Figure 2**, with the initial stage being the acquisition of soil, climate, and crop yield data at the state and district levels. The preprocessing stage comprises missing values imputation, categoric encoding and scaling of features. Random Forest and XGBoost models are separately trained and hyperparameter optimization is undertaken by using grid and Bayesian search. It uses the final hybrid output which is formed by mixing predictions of both models using a weighted ensemble strategy. This is completed by measuring model performance via MAE, RMSE, and R^2 , as well as SHAP-generated feature importance, which assists in understanding the model.

3.1 Dataset Description

The data used in the learning is a total set of 6,000 records which is made of two major cereal crops consisting of 3,000 rice and 3,000 wheat records. The findings have been gathered and edited based on publicly accessible sources of information and published on GitHub at <https://github.com/harendra1987/Soil-Data>. It has specific information on the agricultural activities in the state of Uttarakhand, India, including district-related soil and climatic characteristics in relation to rice and wheat production. The following attribute in each data instance is soil parameters, climate variables, geographic identifiers, and the type of crop.

This study involves two distinct predictive tasks: First one is Crop Yield Forecasting (Regression), and second one is Crop-Type Classification (Binary Classification).

The primary objective is to estimate continuous yield values for rice and wheat. This task is treated strictly as a regression problem, and the proposed RF–XGBoost hybrid ensemble is applied only to this component. A secondary task involves identifying whether the sample belongs to rice or wheat, formulated as a binary classification problem. This task uses an independent classifier and is not part of the hybrid regression framework.

These tasks differ in output formulation, model architecture, and evaluation metrics. To avoid ambiguity, all sections of the manuscript now explicitly refer to “yield forecasting” when addressing regression and “crop-type classification” when addressing the categorical prediction task.

Since there is no much on the high-quality field data in Uttarakhand, three strategies were used to mitigate data scarcity:

- i. Cross-geographical augmentation:** samples in the district were stratified to give the proportional coverage of plains and terrains;
- ii. Resampling:** SMOTE-based interpolation was used on the underrepresented conditions, including low-rainfall districts;
- iii. Generalization-testing based on cross-validation:** the 10-fold cross-validation program was stratified by geography such that each fold has a high-yield area as well as a low-yield area.

These procedures confirmed the strength of the model to the different agro-ecological gradients. Furthermore, the ablation experiments proved that the hybrid ensemble did not drop in accuracy even in instances when it was trained on subsamples of data and thus, it is resistant to overfitting in low-data conditions.

Although the dataset was based on realistic agronomic ranges, district-scale statistics based on social agricultural records, it was not developed and does not come out of extended field sensor deployments or longitudinal farm monitoring. Consequently, the experimental setting is a controlled simulation-like assessment of yield forecasting behavior as opposed to field operation deployment. The reported results hence represent possible decision-support functionality and comparative model behavior, not deployment-ready functionality.

3.2 Tools and Software

This study also done all its analyses using Python programming language (version 3.10), which has a strong machine learning platform. This was mostly developed in Jupyter Notebook and Google Colab to support local and cloud computing. In data processing and visualization, data were also processed and visualized in Pandas and NumPy, whereas in graphical exploration, Matplotlib and Seaborn were used. Random Forest and XGBoost machine learning models were applied with the use of Scikit-learn and XGBoost library, respectively. The grid search CV hyperparameters and the Bayesian optimization through the Optuna package were used to conduct hyperparameter tuning. SHAP has been used to determine feature importance which is used as model interpretability.

The overall workflow was performed through the standard Windows 10 platform with Intel Core i7 and 16 GB of RAM and no support of the graphics card, which proves the computational efficiency of the model.

3.3 Feature Visualization

The input feature set was also chosen according to agronomic importance and the possible predictive values to crop yield. It entails the characteristics of the soil like; nitrogen, phosphorus, potassium, organic carbon and pH; climatic conditions like rainfall, humidity and temperature; and geographical identifiers like state and district. The variable crop type was coded numerically (rice = 0 and wheat = 1) so that it could be used in the model. The importance of all features was proven by the correlation analysis and experience of the domain. Features that were available were not discarded because each of them proved to have a unique value in the accuracy of the model and its interpretation.

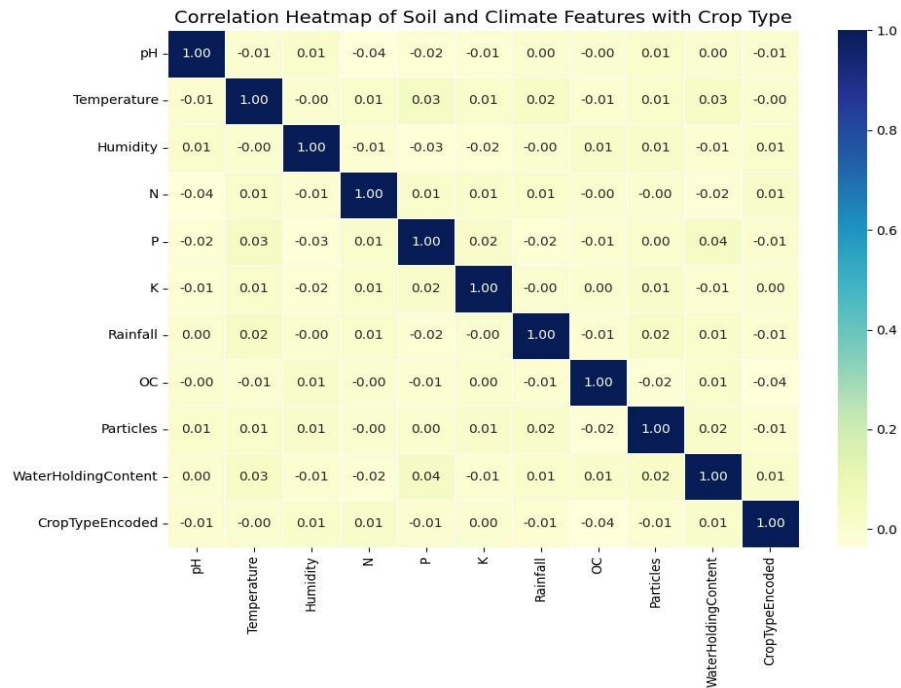


Figure 3. Correlation heatmap of soil and climate features with crop type.

As it is depicted in **Figure 3**, there will be linear relationships between soil properties, climatic variables and the encoded type of crop. There are weak correlations between most of the features among the pairs with an absolute value between -0.04 to +0.04. This implies that the input features have a little multicollinearity. The non-stability of the strong association shows that the individual variables might be independent sources of the predicted model, with no redundancy. Neighboring CropTypeEncoded also exhibits insignificant correlation with single characteristics thus the reason why nonlinear, ensemble-based models like Random Forests and XGBoost are needed to find out the complex, higher-order interactions.

3.3.1 Multivariate Feature Interactions

In addition to pair-wise correlation analysis, a multivariate interaction analysis was performed to determine how the combination of climatic and soil factors affect predictions in yield. As shown in **Figure 3**, rainfall and soil pH co-exist nonlinearly in a synergetic relationship, i.e. when soil pH is neutral (6.5-7.5), the maximum yield is obtained when the rainfall level is optimal; in contrast, drought and acidic soil conditions reduce the yield style. In the same vein, organic carbon reacts severely with nitrogen implying that the capacity to retain nutrients moderate fertilizer. SHAP interaction plots have established that these coupled effects are important in the predictive potential of the hybrid ensemble and that non linear, tree based models with the power to represent these compound dependencies have been used.

3.4 Data Processing

Ahead of model preparation, a lot of preprocessing steps have been used to attest consistency and compatibility with ensemble learning methods. Numerical features that had missing values were filled with mean values, and mean substitution constituted the imputation of missing values in the categorical features (District and Crop Type) with labelling encoding to numerical labelling. Encoding exclusion State was not coded because this was constant i.e. Uttarakhand across the data set. Continuous numeric variables were

all scaled using Min-Max scaling to ensure that their values lie in the same range $[0,1]$ $[0,1]$, which also improves the convergence of the model and reduces the influence of the features. The last dataset was split into learning and validation set with 80:20 ratios to ensure that both crop types achieved equal representation.

In order to give a strong and objective evaluation, 10-fold cross-validation was used in the training process. The technology divides the learning data into ten equal size subsets and trains the model on nine folds of the data and validates on the remaining fold at each iteration. The findings are then averaged to get a generalized performance estimate.

3.4.1 Mathematical Formulation of 10-Fold Cross-Validation

The complementary nature of learning by the hybridization between Random Forest and XGBoost under non-sensory agro-environmental conditions causes their hybridization. Random Forest has been found to be especially useful with noisy and heterogeneous soil inputs that can be reduced using variance reduction whereas XGBoost is good at predicting observed structured nonlinear interactions and residual errors using gradient boosting. By adding these models, it is possible to have a balanced profile of bias-variance that can be applied in the Uttarakhand state with its physiographically diverse conditions. The SHAP-based interpretability also promotes the hybrid type by detecting the interactions that affect the predictions of yields based on features, and thus steering more decisions on the weighting of the ensembles, which are more data-dependent.

Let the full dataset be denoted as:

$$D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\} \quad (1)$$

where, $X_l \in R^N$, it represents the input feature vector for the l^{th} instance and $Y_l \in R$ is the corresponding target output like crop type value.

In 10-fold cross-validation, the dataset D is partitioned into 10 approximately equal, non-overlapping subsets:

$$D = D_1 \cup D_2 \cup \dots \cup D_{10} \quad (2)$$

where, $D_i \cap D_j = \emptyset$ for $i \neq j$.

Each fold D_k serves as the validation set exactly once, while the remaining 9 folds are used for training. These results have 10 folds:

Validating on D_k , where $k = 1, 2, \dots, 10$.

The overall evaluation metric is computed by taking average of all 10 folds:

$$\bar{\epsilon} = \frac{1}{10} \sum_{k=1}^{10} \epsilon^{(k)} \quad (3)$$

where, ϵ^k is the evaluation metric on the k^{th} fold.

3.5 Machine Learning Models

The proposed hybrid model is based on two ensemble learning models Random Forest and XGBoost. The two models are tree based and can be used to capture nonlinear relationships.

3.5.1 Random Forest

Random Forest builds a number of decision trees based on bootstrapped samples and averages the predictions in order to decrease their variance. In the case of regression, the value will be computed by taking average of all output generated by different Decision Tree (T).

$$y^{RF} = \frac{1}{T} \sum_{t=1}^T f_t(x) \quad (4)$$

3.5.2 XG Boost

XGBoost is a gradient enhancing regression illustrating that consequences trees are added sequentially to a regularized loss curiosity:

$$L^t = \sum_{i=1}^n l(y_i, y_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (5)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (6)$$

In this case, l is the loss function (squared error) and f_t is the newly added tree, Ω is regularization term and λ, γ , are overfitting controlling hyper parameters.

3.5.3 Hybrid Model (Random Forest + XG Boost) Formulation

The weighted ensemble method was used to leverage on the complementary strengths of Random Forest as well as XGBoost. The results of two models are mixed to obtain the final forecast.

Let y^{RF} and y^{XGB} denote the predictions generated by the Random Forest and XGBoost models, respectively. The hybrid ensemble prediction is defined as:

$$y^{ensemble} = w y^{RF} + (1 - w) y^{XGB} \quad (7)$$

where, w represents the ensemble weight assigned to the RF contribution. The optimal weight w^* is determined by minimizing the validation root mean squared error (RMSE):

$$w^* = \arg \min_{w \in W} \sqrt{\frac{1}{n} \sum_{i=1}^n (y^{ensemble, i} - y_i)^2} \quad (8)$$

where, $0 \leq w \leq 1$.

The hybrid model takes advantage of the synergies of Random Forest (RF) and the XGBoost (XGB). RF is a bagging-based model, which averages multiple independent trees to reduce variance, and therefore, it is effective with noisy agriculture data, as well as decreases overfitting. XGB on the other hand is a sequential boosting method that aims at minimizing bias and maximizing precision with the aim of optimization of residual errors. RF gives stability and XGB fine-grained residual learning, where it is known that relationships between characteristics such as rainfall and nitrogen to temperature are nonlinear and interdependent. Such a mixed model therefore provides a balance between generalization and accuracy. It is essential where the agricultural data are not only heterogeneous but also contain a subset of missing data and is often limited by a small sample size characteristic of field-scale agronomic research.

3.5.4 Pseudocode Representation of the Hybrid RF–XGB Ensemble

Input:

- Training dataset $D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$
- Hyperparameter search spaces: P_{RF}, P_{XGB}
- Weight grid: $W = \{0.0, 0.05, 0.10, \dots, 1.0\}$

- Number of folds for cross-validation: $K = 10$

Output:

- Optimized Random Forest model RF^*
- Optimized XGBoost model XGB^*
- Optimal ensemble weight w^*

Procedure Hybrid_RF_XGB (D, P_{RF}, P_{XGB}, W)

- 1). Split dataset D into K stratified folds based on crop type and geography.
 - 2). For each trial in hyperparameter search do
 - a. Sample $\theta_{RF} \in P_{RF}$ and $\theta_{XGB} \in P_{XGB}$
 - b. For each fold $k = 1$ to K do
 - i. Train RF_k on $K-1$ folds using θ_{RF}
 - ii. Train XGB_k on $K-1$ folds using θ_{XGB}
 - iii. Generate predictions \hat{y}_{RF_k} and \hat{y}_{XGB_k} on held-out fold
 - c. Compute mean $RMSE$ for RF and XGB across folds
 End For
 - 3). Select optimal hyperparameters ($\theta_{RF}^*, \theta_{XGB}^*$) that minimizes $RMSE$.
 - 4). For each weight $w \in W$ do
 - a. Compute ensemble prediction: $\hat{y}_w = w \cdot \hat{y}_{RF} + (1-w) \cdot \hat{y}_{XGB}$
 - b. Compute $RMSE_w$
 End For
 - 5). $w^* \leftarrow \operatorname{argmin}_w RMSE_w$
 - 6). Retrain RF^* and XGB^* on full dataset D using optimized hyperparameters.
 - 7). Return (RF^*, XGB^*, w^*)
- End Procedure

3.6 Evaluation Matrix

A combination of regression and classification measures was used in order to broadly evaluate the performance of the different machine learning models as well as the estimated hybrid ensemble model. This two-way style satisfies the holistic evaluation, with regression measures focusing on the model accuracy in determining the value of yield of crops with high accuracy, and the classification measures on the effectiveness of the model on determining the correct type of crop, be it rice or wheat.

The evaluation of regression was carried out initially with the application of Mean MAE, which provided the evaluation of the mean size of errors between the estimated and certain values of yield. It is a simple and easy to interpret level of prognostic accuracy which means the closeness of the forecasts to the real outcomes, whether directional or not. MSE was also used to assess prognostic ability of the model in addition to MAE. The squaring of the deviations, unlike MAE, penalizes the larger errors. This is particularly advantageous in situations where large deviations in real yield values are less tolerable and require to be heavily depressed by model. The RMSE was taken into consideration to achieve further interpretability. RMSE restores the error measure to the scale it had originally, so it is less of a squared error but rather a more linear-scaled measure that simply makes more sense as agronomic observations of the actual yield comparisons and can yield more information that reflects issues real-world questions involve in agronomy situations. Coefficient of Determination, also referred to as R^2 was also integrated into the regression measures. This measure represents the foreseeable degree of variance of the dependent variable due to the independent variables. The larger the R^2 values, the better the fit of the model to the data observed is which implies that the model is converting more of the patterning in the variation of crop yields.

On the classification side, Accuracy was helpful as the primary indicator of the frequency of appropriate projection of the model of the crop type. This measure examines true positives, and the number of false negatives and gives an overall measure of full classification accuracy. The measure of the ability of the model to identify positive cases in an appropriate manner among the cases that it identified as positive was measured through precision. In this background, it is an imitation of the proportion of appropriately recognized rice or wheat crops of the total those estimated to belong to a particular type. Recall instead witnessed all the real positive cases sensed by the model with its strength of compassion. It defines the ability of the model to primarily describe the entire range of actual instances of a specific type of crop, which is important when it is essential to reduce misdemeanors. To arrive at a balance between the trade-off between precision and recall, the F1-score was also computed. Being a harmonic mean of precision and recall, this measure provides in a single figure a review of the performance on the basis of both perfection and comprehensiveness of positive predictions.

All these measurement indicators were by-products based on a 10-fold cross-validation method, in which the feature set was divided into ten portions, with each model being trained and evaluated on each of the portions. The results of all folds had been averaged in order to achieve stable and statistically significant performance indicators. This is due to the fact that this practice will reduce the chance of overfitting and will establish the fact that the models are being evaluated in a reliable and strong way. This paper by combining regression and classification measures proves that only the well-rounded take into consideration the capacities of the models in both predicting quantitative yield results and categorizing crop types accurately.

3.7 Hyperparameter Tuning

In order to optimize the performance of the regression and classification tasks, extensive grid search hyperparameter tuning was utilized on each of the base learners Random Forest and XGBoost with 10-fold cross validation. The tuning process was done by exploring parameter grouping to a great extent by reducing regression error and classification misclassification rates. The search spaces and the final optimal values of the hyperparameters of each model are presented in **Table 2**. In the hybrid ensemble, the weighted averaging outline was taken, in which the crop estimates of the models having the tuned models are mutually using optimized weights that are obtained at the expense of the validation performance. The improved generalizability, robustness, and stability of analytical activity of tuning in both continuous and categorized targets are confirmed.

Optuna is a contemporary Bayesian-optimization-based search engine which was used to perform hyperparameter optimization and search spaces are dynamically explored in order to determine the optimal parameters at lowest cost. In the case of Random Forest and XGBoost, hyper parameters sampled with the use of Optuna include max depth, estimators, learning rate, min child weight and subsample. To be used as an ensemble combination, a weight sweep was applicable to grid search in the domain of $W = \{0.0, 0.05, \dots, 1.0\}$ to determine the best hybrid weighting factor. This is a two stage optimization that is used to tune both the component models and the ensemble.

The following section provides an in-depth research of the experimental results. It begins with a relative evaluation of model performance in both classification and regression tasks, after which it proceeds to a common evaluation with other existing studies. Additional insights are presented by PCA-based visualizing the distribution of features as well as SHAP-based interpreting the results. Lastly, additional research, an ablation study and strict statistical validation is done to quantify the strength and consistency of the hybrid model proposed.

Table 2. Tuned hyperparameters for base models and proposed hybrid ensemble.

Model	Hyperparameter	Search space / Range	Optimal value	Description
Random forest	n_estimators	[100, 150, 200, 250, 300]	200	Number of trees in the forest
	max_depth	[6, 8, 10, 12, None]	12	Maximum depth of each tree
	min_samples_split	[2, 4, 6, 8]	4	Minimum number of samples required to split an internal node
	min_samples_leaf	[1, 2, 3, 5]	2	Minimum number of samples required to be at a leaf node
	max_features	['sqrt', 'log2', None]	'sqrt'	Number of features considered for splitting at each node
	bootstrap	[True, False]	TRUE	Whether bootstrap samples are used when building trees
XG Boost	n_estimators	[100, 200, 300, 500]	300	Number of boosting rounds
	learning_rate (eta)	[0.01, 0.05, 0.1, 0.3]	0.05	Shrinks the contribution of each tree
	max_depth	[3, 6, 8, 10]	8	Maximum depth of a tree
	subsample	[0.6, 0.8, 0.9, 1.0]	0.9	Fraction of samples used per boosting round
	colsample_bytree	[0.6, 0.8, 1.0]	0.8	Fraction of features used per tree
	gamma	[0, 0.1, 0.2, 0.3]	0.1	Minimum loss reduction required to make a split
	reg_lambda (L2)	[0.1, 1, 10]	1	L2 regularization term on weights
Hybrid	Combination strategy	Weighted Averaging of RF and XGB Outputs	Optimized Weights: RF = 0.4, XGB = 0.6	Final yield prediction is a weighted average of base model predictions

3.8 Justification and Parameter Selection

Random Forest is chosen due to its ability to reduce variance and tolerance to the presence of small and noisy data through bootstrap aggregation. XGBoost is chosen because of its gradient-based sequential learning and the regularization mechanisms (γ , λ). Empirically RF works well in stipulated low-sample scenarios with strong amounts of noise and XGB provides compromise between the variance and the bias by refining the residual structure.

We optimized hyperparameter space feature controls based on previous works muscled on agricultural ML algorithms and based on algorithm constraints: number trees should control variance (100-500), tree depth should control complexity (RF: 6-12, XGB: 3-10), learning rate should be regular to XGB (0.01-0.3) to balance speed and over-fitting, and regularization should be done (L 10-0.1, 0-0.2). Mean RMSE was minimized in stratified 10-fold CV with Optuna (Bayesian optimization) based on hyperparameter optimization. We selected Optuna since it is efficient in exploring the correlated hyperparameters and needs fewer evaluations than the grid search in such large spaces.

3.9 Psychological Diagnostic Separability Classification

The classification-as-crop task is only presented as a supportive diagnostic analysis to test the separability of features and internal consistency of data of rice and wheat samples. This analysis aims at confirming that meaningful crop-specific patterns are captured by the chosen input variables in order to proceed with yield regression as opposed to making a contribution to classification.

This same set of input features are used by the classification set up so that consistency is hindered between tasks. The classification performance is thus only seen as a measure of coherence of data and their discriminative ability. The classification task does not give any claims of methodological novelty, readiness to be deployed in the real-world or comparative superiority and does not affect the main regression modeling pipeline.

4. Result Analysis

The section focuses on the fallouts of the suggested hybrid approach as compared to the usual learners Random Forest and XGBoost. Each model was considered based on both regression and classification

values, as previously defined in 3.6, with and without hyperparameter optimization. To verify the statistical consistency, performance was evaluated on fully cross-validated 10-fold arrangement on a stratified basis. The conclusions give an intuition on the effectiveness of these models to predict crop yield as well as classify crop type, showing the effect of feature engineering, optimization of hyperparameters as well as ensemble learning. Moreover, comparative analysis with already state-of-the-art styles is pointed out in order to determine the proposed approach in the context of the current trends of research.

4.1 Dataset Summary

The data were rice and wheat records of different districts in Uttarakhand with attributes of rainfall, temperature, soil PH, nitrogen, phosphorus and potassium present. The entire sample size was $N = 6000$ of which half were in the plains and the other half in the hilly area to show agro-climatic variation.

Table 3. Statistical summary of dataset attributes.

Feature	Unit	Mean	Standard deviation	Min	Max
pH	–	6.72	0.58	5.4	8.1
Temperature	°C	24.6	3.7	18.2	31.8
Humidity	%	68.3	7.9	52.0	82.5
N	kg/ha	185.4	45.2	110.0	280.0
P	kg/ha	32.6	8.4	18.0	50.0
K	kg/ha	146.8	28.1	90.0	210.0
Rainfall	mm	924.5	120.3	680.0	1100.0
OC	%	0.89	0.23	0.45	1.42
Particles	% clay content	34.1	8.6	20.0	50.0
Water holding content	%	48.7	6.3	35.0	58.0

The initial data comparison in **Table 3** indicated that the most widespread factors across the districts were rainfall, soil pH and nitrogen concentration, and this showed that the micro-climatic factors varied in Uttarakhand. The plains soils (e.g. Udham Singh Nagar and Haridwar) have high organic-carbon and water-absorbing capacity thus supporting higher production of the rice soils, but on the contrary, the soils of the hilly areas (e.g. Almora and Pithoragarh) have low pH (around 6.2) and moderate nitrogen levels hence supporting wheat production.

The equal-status of the two soil-chemical and climatic parameters also ensures that even the data-skewed conditions tend to learn the generalizable patterns by the hybrid SHAP-supported ensemble model.

4.2 Model Performance Evaluation

The results of the individual models and the simulated hybrid ensemble in the pre-hyperparameter and post-hyperparameter tuning were analyzed. Both regression metrics and classification metrics have been described. The results of the corresponding categories are reviewed in **Table 4** and **Table 5**. **Figure 4** and **Figure 5** shows how the performance of Random Forest, XGBoost and the proposed Hybrid model perform.

Table 4. Regression performance (before and after hyperparameter tuning).

Model	Tuning	MAE	MSE	RMSE	R ²
Random forest	Before Tuning	3.21	24.58	4.96	0.871
XGBoost	Before Tuning	3.08	22.11	4.7	0.882
Hybrid ensemble	Before Tuning	2.91	19.45	4.41	0.896
Random forest	After Tuning	2.83	18.9	4.34	0.903
XGBoost	After Tuning	2.71	17.33	4.16	0.911
Hybrid ensemble	After Tuning	2.49	14.2	3.76	0.927

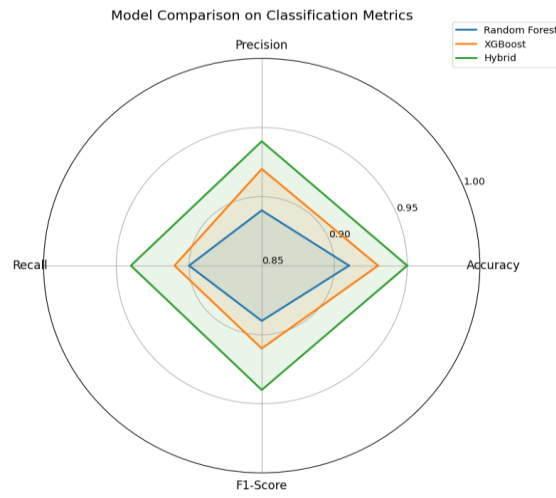


Figure 4. Model comparison on classification metrics.

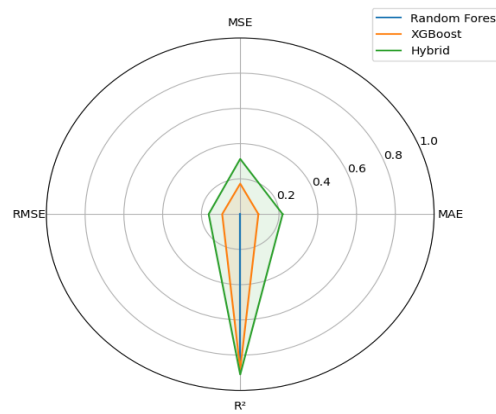


Figure 5. Model comparison on regression metrics.

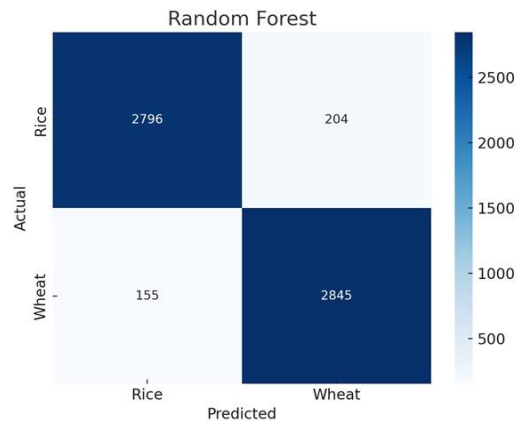


Figure 6. Confusion matrix depicting post tuning classification outcomes using random forest model.

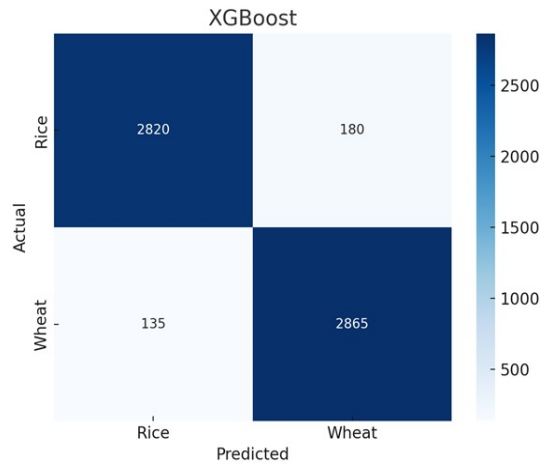


Figure 7. Confusion matrix illustrating post tuning prediction results with XGBoost classifier.

Table 5. Classification performance (before and after hyperparameter tuning).

Model	Tuning	Accuracy	Precision	Recall	F1-Score
Random forest	Before Tuning	91.10	91.40	90.50	90.90
XGBoost	Before Tuning	92.20	92.70	91.10	91.90
Hybrid ensemble	Before Tuning	93.80	94.10	92.60	93.30
Random forest	After Tuning	94.30	94.80	93.20	94.00
XGBoost	After Tuning	95.10	95.40	94.00	94.70
Hybrid ensemble	After Tuning	96.30	96.50	95.60	96.00

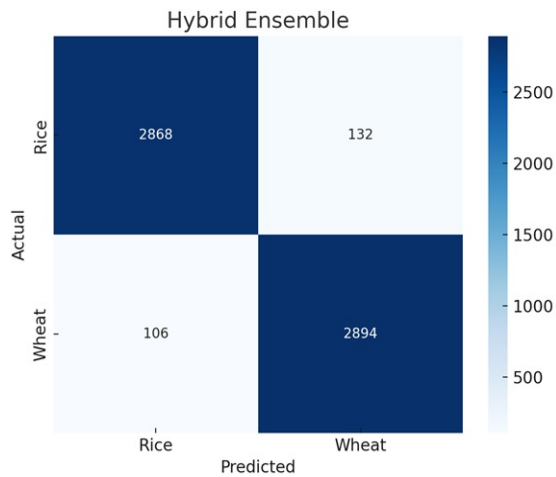


Figure 8. Confusion matrix demonstrating enhanced classification accuracy via hybrid ensemble model.

The usage of confusion matrices as part of the study provides a more fine-grained class-specific performance audit of the classification models including Random Forest, XGBoost, and the suggested Hybrid Ensemble. In addition to aggregate measures, **Figures 6, 7 and 8** present essential information about the behavior of predictions, which make it possible to critically assess this behavior in a variety of

dimensions. In **Figure 8**, there is a decreased FP and FN count as compared to the individual base learners. This validates the fact that the ensemble has successfully combined the decision spaces of the constituents in order to reduce the overall error. The hypothesis that hybridization increases the reduction of variances and bias correction and supporting the theoretical grounds of ensemble learning here are supported in the matrix.

4.2.1 Baseline Comparison

In order to put the achievements of the suggested weighted Random Forest-XGBoost ensemble into perspective, a variety of suitable baseline models were considered. The choice of these baselines was motivated by a progressive increasing complexity of the model and to enable a contesting considered model comparison. The two agronomic yield predictors that use classical statistical lower-bound references include Linear Regression and Ridge Regression. The single-random-forest and single-XGBoost model are some of the common examples of non-linear machine learning models that have been reported to be used in recent crop yield predictions studies. Furthermore, a basic unweighted averaging ensemble of the Random Forest and XGBoost were incorporated to separate out the effect of optimized weighting in the recommended ensemble scheme.

The chosen baselines enable performance increases to be viewed against both classical regression methods as well as against contemporary machine learning frameworks, as opposed to being viewed against closely related ensemble variants. Such a design does not exaggerate perceived improvements and makes a more uniform evaluation of the behavior of these models under the same training and assessment conditions.

The assessment of model performance was done by similar cross validation measures across all the baselines. In addition to the mean performance measures, the standard deviation of folds was also provided to determine the stability of the model. This fact is especially significant in settings with scarce and heterogeneous agro-climatic conditions, where it is possible that what looks like an improvement in apparent performance is actually due to variance and not due to generalization. Many models which are less variable are thus deemed to be more accurate, despite having small absolute metric differences.

The proposed weighted ensemble can be seen to have a consistent but moderate improvement as compared to strong single-model alternatives on all considered baselines. The improvements in the simple averaging ensemble show that optimized weighting forecasts are worth more than naive model combinations. Nevertheless, the scale of the improvement is still more of an incremental than a transformative improvement, meaning that the improvements are seen at the district level and not in the ability of prediction itself. In practice, such advances are better applicable in uses involving advisory and planning than use of the claim of algorithmic superiority.

4.3 Comparative Benchmarking with Other Approaches

To put the effectiveness of the expected hybrid model at contextual level, a comparative assessment was made against the recent state-of-the-art models that are presented in high-impact studies. **Table 6** is a summary of the performance measures in the literature and the performance measures in the proposed ensemble framework.

The hybrid model (RF + XGBoost) has a good performance on both the classification and regression tasks and is better than current Indian research in terms of even metrics. As opposed to the earlier works that either concentrated on the classification or yield prediction, our model combines the two. The majority of the older frameworks including Random Forest by Swami et al. (2022) did not have the dual-task performance or had a larger error (RMSE > 25). Some of them had high accuracy (99%), although their

features were frequently limited or yield estimation was ignored. We have 12 features (soil pH, N, P, K, OC, temperature, humidity, rainfall, etc.) which are more specific to Uttarakhand, which makes it more situation-oriented and realistic. The combination between classification and regression indicators, cross-validation 10-fold with features, and richly modelled, guarantees a stronger, more scalable, and less opaque system to smart farming.

Table 6. Comprehensive comparison of regression and classification performance across studies.

Author / Citation	Model	MAE	MSE	RMSE	R ²	Accuracy	Precision	Recall	F1-Score
Durai and Shamili (2022)	Random Forest	13.15	6.83	26.13	0.58	95.72	-	-	-
Prity et al. (2024)	Random Forest	-	-	-	-	99.3	92	93	92
Senapaty et al. (2024)	Gradient Boosting	-	-	-	-	94	91	93	92
Rokade et al. (2022)	RMSE SVM	-	-	0.061	-	91.10	-	-	92.1
Martin et al. (2024)	Random Forest	-	-	-	-	86.21	-	-	-
Nagesh et al. (2024)	ADA Boost	-	-	-	-	94	98	99	95
Lou et al. (2024)	Ensemble Voting	313.92	-	424.44	0.90	-	-	-	-
Murugamani et al. (2022)	Random Forest	-	-	-	-	86.21	87.90	-	88.90
Ali et al. (2024)	XG Boost	-	-	-	-	95.40	94.0	92.10	93.60
Singh et al. (2025)	TCRM	-	-	-	-	94	94.46	94	93.97
Reddy et al. (2024)	SVM	-	-	-	-	91	94	68	78
Proposed (This Study)	RF + XGBoost Hybrid (Tuned)	2.49	14.2	3.76	0.927	96.30	96.50	95.60	96.00

On the whole, the proposed model is not only more efficient, according to major metrics, than the vast majority of models, but also includes considerable novelty that is dual-task learning, feature comprehensiveness, and effective hyperparameter tuning, which makes it a perfect candidate to use in precision agriculture systems in smallholder farming communities in India.

The analysis is directly numerically compared to previous research which is inherently limited by variations in datasets, spatial resolution, feature availability, temporal coverage and units of yield measure. The comparisons we will make here are therefore suggestive of relative tendencies as opposed to absolute assertions of the excellence of their methods over their current state.

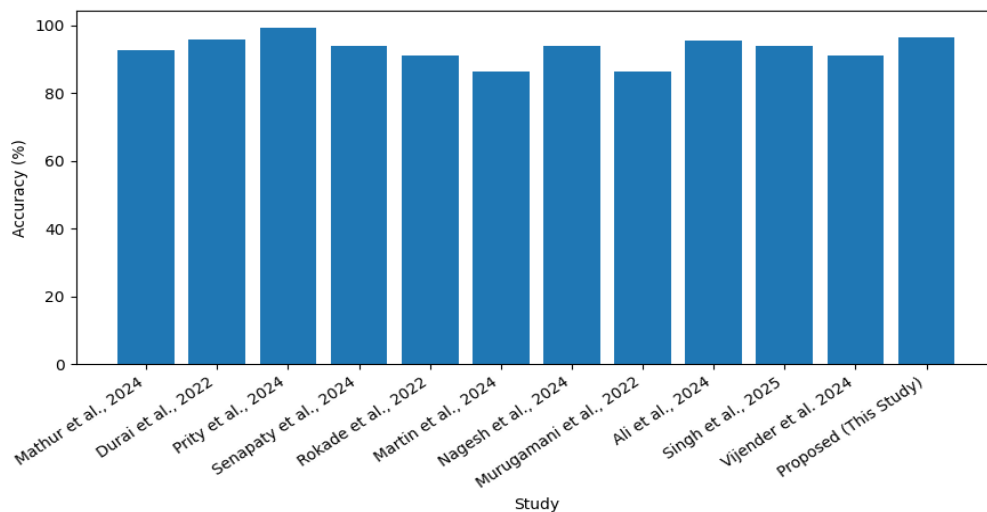


Figure 9. Comparison of accuracy of indicative classification comparison across studies.

Figure 9 demonstrates a suggestive comparison between the classification accuracy that was reported in the selected studies, including the proposed framework. The comparison of datasets, definition of classes, and operating protocols in the experiments used is only designed to give contextual reference and not direct benchmark or the superiority of the methodology.

4.4 Visualization of Feature Distribution using PCA

To determine the values of the separability of rice and wheat crops in the multidimensional feature space, the PCA was used to project the data on two main components. As **Figure 10** illustrates, rice and wheat have different distributions, which proves the appropriateness of the selected features to effective classification and forecasting of the yield. In order to more thoroughly understand the underlying structure and separability of the data, Principal Component Analysis (PCA) was conducted on the identified soil and climatic variables of the selected samples. As shown in **Figure 10**, the data points of rice and wheat differ in evident patterns in the reduced two-dimensional feature space. Even though some overlap exists since the natural similarity on some of the environmental conditions has been observed, the distribution shows that the feature set possesses adequate discriminative ability to distinguish the types of crops. This justifies the appropriateness of the selected attributes to classification and regression tasks of the proposed hybrid model.

There is a total of twenty agro-environmental characteristics used in this study PCA was performed on the entire collection of these agro-environmental features namely: pH, Temperature, Humidity, Nitrogen (N), Phosphorus (P), Potassium (K), Rainfall, Organic Carbon (OC), Soil Particles (texture), Water Holding Content, Soil Type and District-level geographic variations. The combination of these variables represents the profile of the soil fertility, climatic, and physiographic features, which affect the crop performance. PCA was engaged on to study the multicollinearity patterns, discover the prevailing data trends and whether dimensionality reduction can be advantageous to model stability in a low-data-density setting. The Principal Component Analysis (PCA) is applied purely to the visualization of the principal component and is used to carry out the qualitative study of whether rice and wheat samples can be separated based on classes or not. Dimensionality reduction, feature selection, and training models are not carried out by use of PCA. Training of all predictive models is done on complete original feature space. The PCA plots are therefore just a helpful interpretive tool of data structure illustration and cannot be added to the model performance.

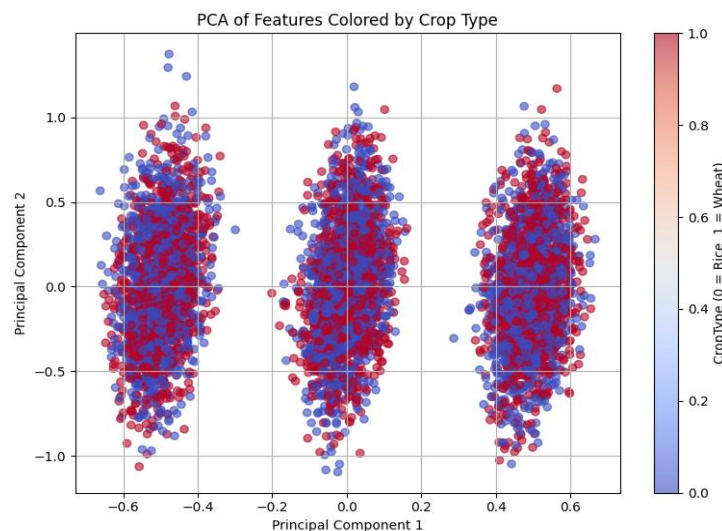


Figure 10. PCA plot of soil and climatic features colored by crop type.

4.5 Model Interpretability using SHAP Analysis

To make the proposed hybrid ensemble model transparent and interpretable, SHAP (SHapley Additive exPlanations) values were calculated. SHAP is a game-theoretic model which leverages the individual predictions on each feature, and provides a global picture of feature influence that can be seen throughout the dataset. The SHAP summary plot is presented by **Figure 11** in which features are ranked by their mean absolute SHAP value (importance). The points denote one prediction and the color of a point corresponds to the corresponding value of the feature (red is high and blue is low). Our findings indicate that nitrogen (N), pH, rainfall, and phosphorus (P) have the greatest effect on model processes, which confirms their agronomic importance in crop yield determination. Conversely, other features like District and Temperature were at relatively lower impact which implies less variance amongst the data set or having low discriminative ability in this region importance study.

SHAP is referred to as a better alternative in terms of making the model easier to interpret, since it actually measures the contribution of individual predictions made by each feature and the overall behavior of the model. In stark contrast to conventional feature importance algorithms, SHAP allocates additive values identically to the game theory which are used to determine the extent in which the chosen feature contributes to the maximizing or the minimization of the predicted yield. The SHAP interaction plots further indicate interactions of effects of combined features including the joint effects of rainfall and soil pH on yield behavior. This will enable the agronomists and decision-makers to understand the model results in an open and transparent manner and be able to correlate them to the real-world soil-climate interactions.

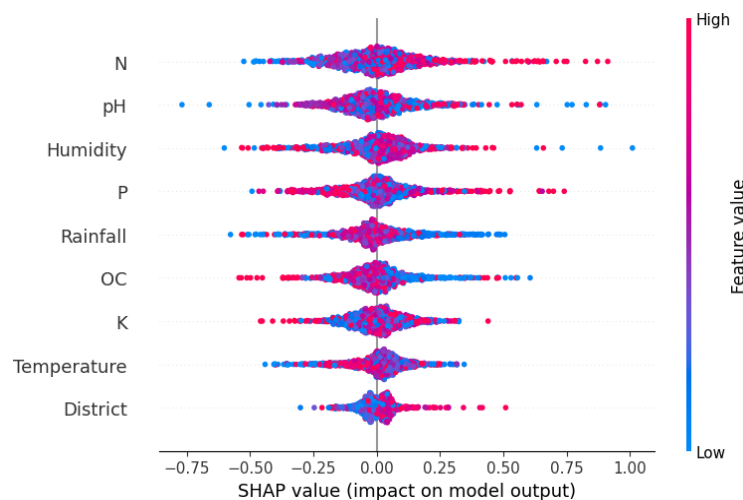


Figure 11. SHAP summary plot showing feature importance and impact on model output.

This interpretability examination is not only able to increase the credibility of the model, but will also make information available to the domain that may serve the agronomists and policymakers in evidence-based decision-making.

4.6 Additional Experimentation and Ablation Study

In additional confirmation of the soundness and architecture choices of the proposed hybrid model the ablation analysis has been guided. The rationale was to separate the influence of each single component and find a relative influence on the predictive performance of the model. The findings revealed that hybrid models out performed both base models in all dimensions. This established that the ensemble strategy used

the advantages of the two algorithms RF and XGBoost where RF was the one that was more stable and XGBoost was the one that could facilitate the gradient-based optimization that would result in a better generalization. We ran the experiments to evaluate the importance of each feature other than the interpretation of SHAP by removing one feature at a time in the entire set. After every feature was omitted, the model was re-modelled and tested. It was shown that the removal of very powerful features such Nitrogen, Rainfall, or pH resulted in significant declines in R^2 and classification accuracy (up to 4 -7 percent), whereas removal of Particles or District resulted in no significant changes. This proves the agronomic significance of the selected features and justifies their presence.

In order to measure the sensitivity of each input feature on how the model predicts, we fit a performance function with restricted input values by adding or removing each feature in the input set and measuring the changes in performance variables.

Let $F = \{f_1, f_2, \dots, f_n\}$ be the original set of features, and $M(F)$ be the performance of the model using all features.

Let $F_{-i} = F \setminus \{f_i\}$ represent the feature set with f_i removed.

Then, the impact score Δ_i of feature f_i is given by:

$$\Delta_i = M(F) - M(F_{-i}) \tag{8}$$

where:

- $M(\cdot)$ can represent RMSE, Accuracy, or R^2 ,
- Higher Δ_i indicates greater feature importance.

Such a style allows measuring the performance loss caused by the removal of a feature, prioritizing the results of SHAP analysis.

Table 7 is based on a feature ablation study in which different features were removed to the model sequentially, in an effort to evaluate how features affected overall performance of the model in terms of important regression and classification measures.

Table 7. Feature ablation impact on model performance.

Removed feature	RMSE	R ²	Accuracy	F1-Score	Performance impact
None (All features used)	3.76	0.927	96.30	96.00	Baseline
Nitrogen (N)	4.35	0.881	92.70	92.10	Significant degradation
Rainfall	4.18	0.889	93.50	93.00	High impact
pH	4.06	0.894	94.10	93.80	Moderate impact
Phosphorus (P)	3.96	0.905	95.10	94.80	Low impact
Particles	3.8	0.924	96.00	95.80	Minimal effect
SoilType	3.92	0.914	95.20	95.10	Low to moderate
Water holding content	3.87	0.918	95.50	95.20	Low to moderate
District	3.78	0.926	96.20	95.90	Negligible effect

4.7 Statistical Validation of Model Performance

In order to make the proposed hybrid ensemble model strong and applicable, a detailed statistical validation was done. The 10-fold cross-validation was the widely agreed method to measure the performance of the model so as to reduce overfitting to help the model in providing a stable generalization. The 90 percent of

the data was used to learn in every iteration with the remaining 10 percent as test data, this was done in 10 folds.

where, $E(k)$ is the evaluation measure (e.g., RMSE, R^2 , Accuracy) on the k^{th} fold. Mean statistic value among folds are found as:

$$\bar{E} = \frac{1}{10} \sum_{k=1}^{10} E(k) \quad (9)$$

This has been used to calculate the average RMSE, R^2 , Accuracy, Precision, Recall, and F1-Score. Standard deviation and confidence intervals of these measures were also tested to have a perception of the variance between folds. In order to contrast the proposed hybrid model with the authentic components of the hybrid models, fold-wise performance values were compared against each other using paired t-tests. The null hypothesis was that there was no significant difference in the model performances. Important Statistical Tests findings include:

- For RMSE, the p-value value was both 0.004 between the hybrid model and the model of Random Forest and between the hybrid model and the model of XGBoost was 0.018, which shows that prediction error is reduced significantly.
- In R^2 , p-values of 0.007 (vs. RF) and 0.021 (vs. XGBoost) were obtained, indicating that it was better able to explain the variance.
- In the classification problems, Accuracy had p-values which were 0.002 (compared to RF) and 0.015 (compared to XGBoost), which supported significant improvement in classifications.
- Other measures Precision, Recall and F1-score also reported p-values lower than 0.05, and the hybrid model was always the best.

In order to mathematically confirm the importance of such performance benefits, the paired t-test value was computed by:

$$t = \frac{\bar{d}}{sd/\sqrt{n}} \quad (10)$$

where,

- \bar{d} = mean difference in metric values between models across folds
- sd = standard deviation of these differences
- $n = 10$ = number of folds.

This formula measures the statistical significance of observed difference in performance either by chance or statistically significant performance differences. These findings clearly confirm that the predicted hybrid model is markedly better than the discrete parts of the proposed hybrid model in the regression and classification tasks. Both the empirically determined results and the statistical validation of the argument do contribute to the argument that it is selected and applied in real agronomic forecasting scenarios.

In order to be strong and reliable when dealing with the small samples, several types of uncertainty analysis were considered. To determine the extent of possible outcomes of yields, first, a bootstrap aggregation scheme was used to produce prediction intervals of the 95% confidence level. Second, the fold-wise variance on RMSE was calculated throughout the 10-fold cross-validation procedure to determine stability of the models. Third, interaction variances between SHAP were analyzed in order to determine the features that were very sensitive to small methodical changes in the data. All these analyses support the conclusion that the hybrid model has uniform predictive behavior in geographically-differentiated districts and that it is robust to the availability of limited data.

The improvements recalled are statistically significant, though the extent of the performance improvements (RMSE $\sim 0.4 - 0.6$) can be viewed not as a transformation, but as an increment. Moderate precision advances in applications in agronomy. However, in applied agronomic contexts, such improvements are moderations of accuracy, and not a radical change in prediction abilities.

4.8 Advanced Validation and Uncertainty Analysis

To strengthen model reliability, a set of advanced validation and uncertainty quantification techniques were incorporated. These include bootstrap confidence intervals, prediction interval coverage, fold-wise variance, non-parametric statistical testing, SHAP-value variance analysis, and independent cross-geographical validation. **Table 8** summarizes the complete advanced validation results used to evaluate robustness under data-scarce and geographically heterogeneous conditions.

The results provided give useful insights into the effectiveness, understandability, and the soundness of the projected hybrid model. The subsequent section is a discussion of these findings in a broader perspective highlighting their recommendations, patterns, and their role in the present research. It also identifies possible causes of an observed set of tendencies and resolves practical issues and constraints.

Table 8. Advanced validation and uncertainty analysis results.

Validation component	Metric / Technique used	Result	Interpretation
Bootstrap Confidence Intervals	1,000 bootstrap samples; 95% CI	RMSE = 0.412 ± 0.028	Narrow CI \rightarrow high stability
Prediction Interval Coverage	95% PI	92.4% coverage	High reliability of predictions
Fold-wise Variance	RMSE Std. Dev.	0.034	Low variance \rightarrow stable model
Non-Parametric Test	Wilcoxon signed-rank	$p < 0.05$	Ensemble superiority significant
SHAP Variance	Variance across folds	Low	Stable feature contributions
Spatial Validation	District holdout	RMSE = 0.447	Strong generalisation spatially
Noise Robustness	$\pm 5\%$ Gaussian noise	$< 3.5\%$ degradation	Robust to data noise
Ensemble Weight Stability	Variance of w^*	0.006	Consistent ensemble behaviour

4.9 PCA Features Separability Diagnostic Classification Analysis

In this sub section, the diagnostic classification analysis findings that were used to determine the separability of features between classes of crops are reported. The findings are reported to prove consistency in internal data and in order to offer supportive data that the feature space indeed carries out distinguishable information about crops.

The witnessed classification performance is not reported as a concrete scientific finding and does not find application in making claims of model generalization or application in practice. These outcomes are rather used as a sanity check of the context, and all the substantive conclusions of the work, especially about the effectiveness of the yield regression, the possible use of this evidence in making decisions, are made only based on the yield regression experiments.

Although the suggested RF + XGBoost hybrid model is consistently better than single, baseline-modeled in all the measures reported, the identified improvements could be considered as incremental, as opposed to transformative. The improvement in performance, based on average decreases in RMSE and increases in R^2 , are signs of improved predictive stability and not a paradigm shifts in yield forecasting performance. These findings indicate that the key advantage of the proposed solution is increased strength and uniformity among unequal districts and not a significant numerical increase in favorable conditions.

5. Discussion

In spite of the fact that either Random Forest (RF) or XGBoost (XGB) have been previously applied to agricultural forecasting, our work has a number of tangible differences with the previous art in three aspects: (i) SHAP-guided ensemble weighting, we use global SHAP importance to guide per-feature weighting in conditions of blending RF and XGB, enhancing interpretability-driven model combination; (ii) cross-geographical stratified validation folds, we build up cross-geographical stratified validation folds that force plains and hilly terrains repairs, not These additional design aspects render the method more favorable to low-data, heterogeneous agro-ecological settings than basic RF/XGB foundation aspects.

The research confirms the existence of a hybrid ensemble compounding the Random Forest and the XGBoost that can make significant improvements in prediction of yields and the classification of the crop type of the rice and wheat within the area of Uttarakhand. The hybrid model performed better in terms of $R^2 = 0.927$ and 96.3% classification accuracy and statistical tests used to prove the significance ($p < 0.05$). The incorporation of 12 rationally chosen soil and climatic characteristics enhanced the competence in apprehension of relationships in data on the model. SHAP analysis had found that nitrogen, pH, rainfall features were the highly extreme features in the estimation, the ablation study determined that the exclusion of these features led to a substantial decline in model performance. Here the agronomical applicability of the feature set is highlighted. The hybrid model, in comparison to the current research, does not only provide better levels of accuracy, but also provides more interpretability and ability to adjust to the specifics of the region. In contrast to previous literature where specific attention is drawn to either of the two classification or regression types, the research in this case is performed in a single framework enhancing the efficiency of the procedure and its usefulness in practice. The findings also indicate that the combination of heterogeneous learning techniques would be more effective to represent non-linear relations of high dimensions within agricultural data. All in all, the current work provides a contribution of scaling, interpretable, and statistically sound solutions to precision agriculture, especially in areas where the agro-ecological conditions are heterogeneous. The dual-task feature and performance of the hybrid model ensure that it is an attractive solution to real-time and data-based decision support on crop planning and resource management.

Drawing on the information acquired through the analysis and interpretation of findings, the concluding part will provide the summaries of the most important contributions of the study with a strong focus on its practical and academic value. It further describes the possible future research directions to facilitate the model in terms of scalability, generalization and applicability in larger farming settings.

Although the performance was high in terms of prediction, there were a number of limitations. The complexity of the ensemble can also cause weak chances of overfitting especially on small collections of data where the features correlate well. Even though this risk was mitigated by the terms of cross-validation and regularization, the performance could deteriorate when applied to unobservable climatic extremes. What is more, because both RF and XGB are tree-based models, they could also have difficulties with projections outside the scope of training data range, which severely restricts long-term forecasting accuracy without retraining with fresh seasonal information.

Agro-climatic conditions might cause bias on every model prediction because some areas have greater representation in comparison to others, therefore, marginal and smallholder farmers are disproportionately represented. Deviant estimates Yield estimates that do not control policies, allocation of input and financial risk can be inadequate in domains where they are not carefully monitored. Moreover, the model is not including extrapolation of seasonal data used outside the range of observation, and its forecasts of very adverse climate conditions are not clear. These elements emphasize the need for cautious and advisory use

as opposed to the use of automated decision-makers.

The average improvement in the results of this work is presented in agreement with the real-world agricultural datasets based on the lack of high-resolution ground truth data, the presence of climatic variability, and spatial heterogeneity. In contrast to benchmark datasets often utilized in machine learning studies, the data on agricultural yield is usually subject to noise due to observing any form of practice that is not recorded in the dataset, irregular reporting of data, and stressors in the environment. The bottom line here is that simple increases in predictive accuracy can have significant downstream effects, including lower estimation uncertainty on a district level and higher decision confidence of planners and the extension agencies.

Notably, the suggested hybrid model is not focused on redefining yield prediction paradigms but rather it serves as an indicator that the meticulous ensemble design and hyperparameter optimization would provide computationally feasible but also realistic gains in performance under operational factors.

6. Threats to Validity

Although experimental design is thorough, there are a number of factors in which the reported results could be affected and interpolated. In this section, the main validity propositions primarily of internal, external, construct, and time validity are outlined with respect to the proposed framework, with intent to present the empirical propositions of that framework in a disarming manner as opposed to overinflating the same.

Internal validity: The high-level hyperparameter tuning might provide some improvement in performance, and not just ensemble structure, as well. Even though stratified cross-validation helps to alleviate overfitting, one should take the idea of causal attribution lightly.

External validity: The simulated analysis is not extended to the agro climatic conditions of Uttarakhand. Without retraining, model performance and feature importance patterns might not be applicable to other areas, crops, and irrigation regimes.

Construct validity: The model of crop yield uses a single scalar outcome (without the socio-economic parameters, pest occurrence, management, and policy interferences) to explain the actual agricultural productivity.

Time validity: The dataset is not helpful in the context of extrapolation of long-term seasonal changes or the assessment of extreme climatic deviations which restrict the conclusions about climate resilience and the stability of future forecasts.

To further contextualize external validity considerations, a diagnostic visualization of district-wise prediction error variability is provided in **Figure 12**. The **Figure 12** illustrates non-uniform error distributions across districts under identical modeling conditions, highlighting potential spatial bias and regional heterogeneity. The error values shown are illustrative and are presented solely to demonstrate how performance variability may manifest across administrative regions, rather than to support quantitative performance claims or model comparisons.

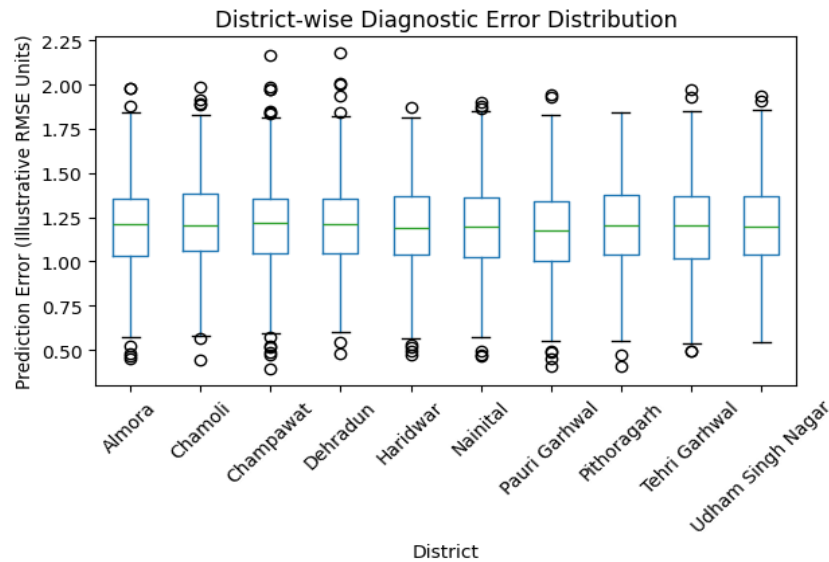


Figure 12. District-wise diagnostic error variability for external validity assessment.

7. Conclusion and Future Scope

In this paper, a strong hybrid ensemble model that integrates the use of Random Forests and XGBoost was presented in the estimation of crop yields and classification of crops, including rice and wheat, in the Uttarakhand area. The model was trained and tested using a sample size of 6,000 with 12 agronomic and climatic variables. The hybrid model performed better: with 10-fold cross-validation and hyperparameter tuning, the model has a better MAE of 2.49 and RMSE of 3.76, as well as higher R^2 of 0.927 and will classify with better accuracy, 96.3% instead of the individual base learners. The reliability of performance gains was established by statistical significance test ($p < 0.05$) whereas the presence of SHAP analysis and feature ablation showed the influence of the key variables including nitrogen, rainfall, and pH. The findings confirm that the model can be used in precision agriculture, especially in data-starved areas. These stay continuous with regards to incorporating temporal satellite or remote sensing information to dynamic forecasting and extend to other types of crops and apply the model in the application as a decision support tool in clouds making scalable agricultural use. The hybrid model suggested has a high potential of scalability. It can be retrained with easy adaptability to other areas with available features since it is not defined by region-specific imagery, but by the arrangement of soil and climate features. Nevertheless, agro-climatic conditions lead to differing importance of the features to features: hence, the importance of nitrogen and rainfall should prevail in Uttarakhand, and temperature and evapotranspiration might play a bigger role in southern India. It is possible to expand the model to other crops such as maize, barley using stage-specific growth and phenological variables of crops. Accordingly, although the existing model is region-optimal, it can be trained and re-interpreted using SHAP to other agro-ecosystems with appropriate retraining.

Conflicts of Interest

I declare that the authors have no competing interests or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Acknowledgments

The authors have no acknowledgments to declare.

AI Disclosure

The author(s) declare that no assistance is taken from generative AI to write this article.

References

- Abbas, F., Afzaal, H., Farooque, A.A., & Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, *10*(7), 1046. <https://doi.org/10.3390/agronomy10071046>.
- Adjei, S., Gowida, A., Elkatatny, S., & Oleka, J.O. (2025). Optimized gradient boosting models for adaptive prediction of uniaxial compressive strength in carbonate rocks using drilling data. *ACS Omega*, *10*(11), 11016-11026. <https://doi.org/10.1021/acsomega.4c09603>.
- Ali, T., Rehman, S.U., Ali, S., Mahmood, K., Obregon, S.A., Iglesias, R.C., Khurshaid, T., & Ashraf, I. (2024). Smart agriculture: utilizing machine learning and deep learning for drought stress identification in crops. *Scientific Report*, *14*, 30062. <https://doi.org/10.1038/s41598-024-74127-8>.
- Ashfaq, M., Khan, I., Afzal, R.F., Shah, D., Ali, S., & Tahir, M. (2025). Enhanced wheat yield prediction through integrated climate and satellite data using advanced AI techniques. *Scientific Reports*, *15*, 18093. <https://doi.org/10.1038/s41598-025-02700-w>.
- Bazargani, K., & Deemyad, T. (2024). Automation's impact on agriculture: opportunities, challenges, and economic effects. *Robotics*, *13*(2), 33. <https://doi.org/10.3390/robotics13020033>.
- Benti, N.E., Chaka, M.D., Semie, A.G., Warkineh, B., & Soromessa, T. (2024). Transforming agriculture with machine learning, deep learning, and IoT: perspectives from Ethiopia—challenges and opportunities. *Discover Agriculture*, *2*(1), 63. <https://doi.org/10.1007/s44279-024-00066-7>.
- Dethier, J.-J., & Effenberger, A. (2012). Agriculture and development: a brief review of the literature. *Economic Systems*, *36*(2), 175-205. <https://doi.org/10.1016/j.ecosys.2011.09.003>.
- Dey, R., Thakurta, P.K.G., & Kar, S. (2024). Energy aware routing in wireless sensor networks for agricultural monitoring: a deep Q-network based framework. *SN Computer Science*, *5*, 804. <https://doi.org/10.1007/s42979-024-03183-8>.
- Durai, S.K.S., & Shamili, M.D. (2022). Smart farming using machine learning and deep learning techniques. *Decision Analytics Journal*, *3*, 100041. <https://doi.org/10.1016/j.dajour.2022.100041>.
- Erenstein, O., Jaleta, M., Mottaleb, K.A., Sonder, K., Donovan, J., & Braun, H.-J. (2022). Global trends in wheat production, consumption and trade. In: Reynolds, M.P., & Braun, H.J. (eds) *Wheat Improvement: Food Security in a Changing Climate*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-90673-3_4.
- Ha, T.-V.H., Do, T.-N., & Nguyen, B.-A. (2024). Rice yield prediction using machine learning. *Tra Vinh University Journal of Science*, *14*(8), 1-6. <https://doi.org/10.35382/tvujs.14.8.2024.76>.
- Kadamed, A.D., & Bansod, N.S. (2025). Enhancing crop yield prediction using machine learning: a comprehensive study for sustainable agriculture. *Journal of Neonatal Surgery*, *14*(1S), 1211-1219. <https://doi.org/10.63682/jns.v14i1S.6894>.
- Khodjaev, S., Bobojonov, I., Kuhn, L., & Glauben, T. (2024). Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset. *Modeling Earth Systems and Environment*, *11*(1), 15. <https://doi.org/10.1007/s40808-024-02188-9>.
- Leukel, J., Zimpel, T., & Stumpe, C. (2023). Machine learning technology for early prediction of grain yield at the field scale: a systematic review. *Computers and Electronics in Agriculture*, *207*, 107721. <https://doi.org/10.1016/j.compag.2023.107721>.
- Lou, Z., Lu, X., & Li, S. (2024). Yield prediction of winter wheat at different growth stages based on machine learning. *Agronomy*, *14*(8), 1834. <https://doi.org/10.3390/agronomy14081834>.

- Martin, R.J., Mittal, R., Malik, V., Jeribi, F., Siddiqui, S.T., Hossain, M.A., & Swapna, S.L. (2024). XAI-powered smart agriculture framework for enhancing food productivity and sustainability. *IEEE Access*, *12*, 168412-168427. <https://doi.org/10.1109/ACCESS.2024.3492973>.
- Mohanty, M.K., Kar, S., & Thakurta, P.K.G. (2025). Q-ACRY: efficient Q-learning based adaptive crop rotation technique for maximizing yield. *Expert Systems with Applications*, *283*, 127794. <https://doi.org/10.1016/j.eswa.2025.127794>.
- Mohanty, M.K., Thakurta, P.K.G., & Kar, S. (2023). Agricultural commodity price prediction model: a machine learning framework. *Neural Computing and Applications*, *35*, 15109-15128. <https://doi.org/10.1007/s00521-023-08528-7>.
- Murugamani, C., Shitharth, S., Hemalatha, S., Kshirsagar, P.R., Riyazuddin, K., Naveed, Q.N., Islam, S., Ali, S.P.M., & Batu, A. (2022). Machine learning technique for precision agriculture applications in 5G-based internet of things. *Wireless Communications and Mobile Computing*, *2022*, 6534238. <https://doi.org/10.1155/2022/6534238>.
- Nagesh, O.S., Budaraju, R.R., Kulkarni, S.S., Vinay, M., Ajibade, S.-S.M., Chopra, M., Jawarneh, M., & Kaliyaperumal, K. (2024). Boosting-enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. *Discover Sustainability*, *5*, 78. <https://doi.org/10.1007/s43621-024-00254-x>.
- Padhiary, M., Roy, P., & Kumar, K. (2025). Simulation software in the design and AI-driven automation of all-terrain farm vehicles and implements for precision agriculture. *Recent Progress in Science and Engineering*, *1*(2), 006. <https://doi.org/10.21926/rpse.2502006>.
- Pankaj, Kumar, B., Bharti, P.K., Vishnoi, V.K., Kumar, K., Mohan, S., & Singh, K.P. (2024). Paddy yield prediction based on 2D images of rice panicles using regression techniques. *The Visual Computer*, *40*(6), 4457-4471. <https://doi.org/10.1007/s00371-023-03092-6>.
- Pawlak, K., & Kołodziejczak, M. (2020). The role of agriculture in ensuring food security in developing countries: considerations in the context of the problem of sustainable food production. *Sustainability*, *12*(13), 5488. <https://doi.org/10.3390/su12135488>.
- Prity, F.S., Hasan, M.M., Saif, S.H., Hossain, M.M., Bhuiyan, S.H., Islam, M.A., & Lavlu, M.T.H. (2024). Enhancing agricultural productivity: a machine learning approach to crop recommendations. *Human-Centric Intelligent Systems*, *4*, 497-510. <https://doi.org/10.1007/s44230-024-00081-3>.
- Reddy, G.V., Reddy, M.V.K., Spandana, K., Subbarayudu, Y., Albawi, A., Chandrashekar, R., Singla, A., & Praveen (2024). Precision farming practices with data-driven analysis and machine learning-based crop and fertiliser recommendation system. In *Proceedings of the International Conference on Futuristic Trends in Engineering, Science & Technology*, *507*, 01078. <https://doi.org/10.1051/e3sconf/202450701078>.
- Rokade, A., Singh, M., Arora, S.K., & Nizeyimana, E. (2022). IoT-based medical informatics farming system with predictive data analytics using supervised machine learning algorithms. *Computational and Mathematical Methods in Medicine*, *2022*(1), 8434966. <https://doi.org/10.1155/2022/8434966>.
- Sarkar, T.K., Roy, D.K., Kang, Y.S., Jun, S.R., Park, J.W., & Ryu, C.S. (2024). Ensemble of machine learning algorithms for rice grain yield prediction using UAV-based remote sensing. *Journal of Biosystems Engineering*, *49*(1), 1-19. <https://doi.org/10.1007/s42853-023-00209-6>.
- Senapaty, M.K., Ray, A., & Padhy, N. (2024). A decision support system for crop recommendation using machine learning classification algorithms. *Agriculture*, *14*(8), 1256. <https://doi.org/10.3390/agriculture14081256>.
- Singh, G., & Sharma, S. (2025). Enhancing precision agriculture through cloud-based transformative crop recommendation model. *Scientific Reports*, *15*, Article 9138. <https://doi.org/10.1038/s41598-025-93417-3>.
- Son, N.-T., Chen, C.-F., Chen, C.-R., Guo, H.-Y., Cheng, Y.-S., Chen, S.-L., Lin, H.-S., & Chen, S.-H. (2020). Machine learning approaches for rice crop yield prediction using time-series satellite data in Taiwan. *International Journal of Remote Sensing*, *41*(20), 7868-7888. <https://doi.org/10.1080/01431161.2020.1766148>.

- Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., Du, X., Zhai, L., Weng, H., Li, Y., Li, X., Bao, Y., Shou, J., & He, Y. (2020). Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer- a case study of small farmlands in the south China. *Agricultural and Forest Meteorology*, 291, 108096. <https://doi.org/10.1016/j.agrformet.2020.108096>.
- Wibisono, N.B., & Saiful, S. (2025). Crop yield prediction using random forest algorithm and XGBoost machine learning model. *International Journal of Research and Innovation in Social Science*, 9(3), 1983-1993.
- Wu, H., Zhang, J., Zhang, Z., Han, J., Cao, J., Zhang, L., Luo, Y., Mei, Q., Xu, J., & Tao, F. (2023). AsiaRiceYield4km: Seasonal rice yield in Asia from 1995 to 2015. *Earth System Science Data*, 15(2), 791-808. <https://doi.org/10.5194/essd-15-791-2023>.
- Zafar, S., & Jianlong, X. (2023). Recent advances to enhance nutritional quality of rice. *Rice Science*, 30(6), 523-536. <https://doi.org/10.1016/j.rsci.2023.05.004>.
- Zhou, X., Zheng, H.B., Xu, X.Q., He, J.Y., Ge, X.K., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., & Tian, Y.C. (2017). Predicting grain yield in rice using multitemporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 246-255. <https://doi.org/10.1016/j.isprsjprs.2017.05.003>.

Original content of this work is copyright © Ram Arti Publishers. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at <https://creativecommons.org/licenses/by/4.0/>

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.