International Journal of Elite Research and Innovation

Vol. 01 Issue 02, July, 2026

Crop Yield Forecasting Under Changing Climatic Conditions

Anurag Yadav

Civil Engineering Department, MNNIT Allahabad, Prayagraj, Uttar Pradesh, India

Article History

Received Date: 13/06/2026
Revised Date: 14/07/2026
Accepted Date: 21/07/2026
Published Date: 28/07/2026

Abstract
Agricultural planning, food-security management and decision making for climate-resilient agriculture require climate-sensitive crop-yield forecasting. In this research, machine learning models were built and tested for predicting the yield of various crops for different states in India as the climate changes. The analysis combined 19,652 observations from 55 crops, 30 states and union territories, six growing seasons and the time frame 1997-2019. The climatic predictors were temperature, rainfall and humidity, and the contextual variables were soil nitrogen, soil phosphorus, soil potassium, soil pH, cultivated area, crop type, season, state and year. Chronological training/validation/testing partitions for eight forecasting approaches were compared. Significant positive rainfall and national average humidity trends were observed, while there was no significant national average temperature trend. XGBoost had the lowest testing log-RMSE of 0.323 and the highest (R2) of 0.926. The most influential were crop type, area cultivated, season, soil nutrients and soil temperature. The accuracy of the forecasts was mixed between the crops and regions rice and wheat, pulses and oilseeds were comparatively good while maize forecasts were weak. The climate sensitivity analysis showed that with increased warming and a decreased level of rainfall stress or humidity, yields were predicted to decrease. The results show that the use of nonlinear machine learning models useful for agriculture planning that takes into account climate conditions however, higher resolution weather and soil-moisture and remote-sensing data required to generate more accurate future predictions.

Keywords:

- Crop-yield forecasting
- climate variability
- machine learning
- XGBoost
- climate-resilient agriculture

1. Introduction

Accurate forecasting of crop production is crucial for food security, for agricultural production planning, for efficient regulation of commodity markets, for land, water, fertiliser and other resource allocation purposes. Accurate predictions help governments and agricultural organisations to plan for shortages, procure food and improve disaster preparedness. They also assist farmers in their crop selection and input management/marketing decisions. In recent years, forecasting systems have evolved to involve the integration of agronomic, climatic and environmental information for growing more accurate forecasts than those based only on historical yield record (Elavarasan et al., 2018). Accurate yield data are also crucial to ensure food system resilience and resilience of agricultural systems to climate-induced shocks and stresses (Miles & Moore, 2026).

Climate changes have begun to affect agricultural productivity through increasing temperatures, changes in precipitation, high humidity in the air, drought and excess rainfall. The temperature influences the development of crops, evapotranspiration and the length of reproductive phases, while the rainfall influences the availability of water during germination, flowering and grain filling. Concurrently hot and dry weather during the growing season result in significant yield losses in large agricultural areas (Heino et al., 2023). Climate variability may also impact planting dates, soil-water availability, pest incidence, and consequently long-term agricultural stability and annual production (Baffour-Ata et al., 2021). Climate change is making agriculture more vulnerable and driving the growing demand for forecasting systems that can support adaptation (Pörtner et al., 2022).

The climatic requirements of major crops vary highly and so do tolerance limits. Monsoon timing and rainfall intensity and availability play a critical role in rice production, especially in rain-fed areas of India (Bowden et al., 2023). Weather based forecasting is especially beneficial for wheat growing areas as wheat is highly sensitive to terminal heat stress and moisture stress during grain filling (Dhakar et al., 2022). The relationship of yield of maize to temperature and rainfall and interaction between management is non-linear (Chang et al., 2023). Pulses and oilseeds are generally sensitive to extended periods of drought and erratic rainfall, and sugarcane requires consistent moisture over a relatively long period. These disparities necessitate crop-specific instead of uniform climate yield models.

Typical methodology used in forecasting involves previous trends in yields and statistics on cultivated areas and production. These techniques are likely to ignore non-linear climatic impacts, interactions with the environment, and regional variability. They can also give false-looking good accuracy if production, a mathematically related predictor, is added. Traditional statistical models often fail to capture the interactions between various climatic, crop types, soil types and crop management variables (Ansarifar et al., 2021). Forecasts using only historical averages may, therefore, be poor in abnormal climatic years.

Machine-learning models can predict the non-linear and interactive relationship that conventional models cannot. Climatic, soil, spatial and agricultural parameters can be combined in a single predictive framework using decision trees, random forests, support vector machines or neural networks (Iniyan et al., 2023). Temporal dependencies of weather also be captured using deep-learning approaches, and give rise to data yields (Gavahi et al., 2021). It has been shown that using neural networks, satellite observations, and climate information can be very helpful in crop yield estimation and forecasting (Jhajharia & Mathur, 2022). Process-based knowledge and data-driven learning integrated into a hybrid crop-modelling and machine-learning system further enhancing prediction (Chen & Tao,

2022). As a result, efficient machine-learning frameworks are increasingly relevant to agricultural forecasting (Gopal & Bhargavi, 2019). The effects of climatic variables beneficial within the optimum range but detrimental above the crop-specific threshold. Models which ignore these interactions are unlikely to be regionally and temporally generalizable (Crane-Droesch, 2018).
Past research tends to look at a single crop and/or a single area and/or a few climate parameters. Very few incorporate long-term trends in temperature, rainfall, and humidity, coupled with the state level characteristics of soil, crop seasonality, and temporal validation and alternative climate scenarios. Furthermore, several prediction studies focus only on model accuracy, without being able to interpret climate sensitivity and test model performance under climate extremes (Jhajharia & Mathur, 2022).
The aim of the study is to create and test climate-responsive machine-learning models for predicting crop yields in Indian states under altered climatic conditions. It explores temporal trends of temperature, rainfall and humidity, climate yield relationships, compares responses by crop, season and state, includes climatic, soil, spatial and agricultural predictors, compares statistical and machine-learning models, estimates yield under alternative climates, and pinpoints influential climatic determinants.

2. Materials and Methods

2.1 Research Design and Study Area

This study was quantitative, longitudinal, and predictive in design with secondary data pertaining to the agriculture, climate and soil conditions of the Indian states and union territories. India was selected as study area because of the agroclimatic diversity present in India with tropical, arid, semi-arid, temperate and high-altitude regions. While the original data period was 1997-2020, the analysis was based on the years 1997-2019. The year 2020 was not considered due to the limited availability of data on crops in the year (37 cases in Uttarakhand only).

2.2 Dataset Description

Three separate CSV datasets were merged into a single study (Mishra, 2025). The Crop Yield data contained 19,689 observations on 30 states/UTs, 55 crop categories and six agricultural seasons, viz. Kharif, Rabi, Summer, Winter, Autumn and Whole year. The sample size was 19,652 after excluding the data that were not complete in 2020. There were 720 state-year observations in the weather data set from 1997 to 2020. It comprised average annual temperature, total annual rainfall and average annual humidity. One soil record was supplied per state, with measurements of nitrogen, phosphorous, potassium and soil pH.

2.3 Variable Selection

For all analyses, yield of crops was chosen as the dependent variable. Climatic data like temperature, rainfall, humidity and soil properties like N, P, K and pH were considered as independent variables. Other context variables like crop type, season, state, area and year were also added to the models. Production was deliberately omitted from the analysis since it is mathematically linked to yield and result in a leakage of the target and bias the model output. In the same way, the amounts of raw fertiliser and pesticide usage were omitted from the principal models as they were highly correlated with cultivated area and not properly representative of the rates of usage by crop.

2.4 Data Integration and Preprocessing

Crop names, state names and labels for seasons were standardized for consistency before the datasets were merged. The datasets for the crop and weather were joined on the common fields of state and year, and the soil dataset was joined on state. During the integration process the full data set was produced, there were no missing values or duplicate records.
Descriptive statistics (mean, standard deviation, and range) and graphical methods (histograms and box plots) were used to examine the distribution of the crop yield to identify possible outliers. This was because yields of coconuts were found to be significantly higher than for most other crops, which might affect the performance of the models. To address this, two sets of model estimates were conducted using the original yields and a log-transformed version, log (1 + yield). Sensitivity testing also was applied to coconut. All categorical variables were transformed into dummy variables, and all continuous variables were standardized using parameters only from the training set.

2.5 Climate-Trend and Statistical Analysis

Trends in temperature, rainfall and humidity were analysed at both national and state levels. The results of these trends were tested for statistical significance using the Mann-Kendall test and the rate of change over time was quantified using Sen's slope estimator. The temperature, rainfall, and humidity climate anomalies were derived for each state against the baseline period (1997-2019). Other indicators, such as standardised rainfall deviations and a composite climate variability index were also created. Data were analysed for data distribution, crop and season summary, correlation analysis, analysis of variance and variance inflation factor analyses to understand the data distribution, correlation among various variables and presence of multicollinearity.

2.6 Forecasting Models

Various predictive models were used to estimate crop production. Such as Multiple Linear Regression, Ridge Regression, Lasso Regression, Decision Tree Regression, Random Forest Regression, Gradient Boosting Regression, XGBoost Regression and Support Vector Regression. All eligible crops were included using a pooled modelling to allow the crop type, state and season to be categorical variables. Furthermore separate models were built for the major crops like rice, wheat, maize, sugarcane, potato, pulses and oilseed crops to recognise the variability in yield scale and sensitivity to climatic conditions.

2.7 Model Development and Validation

The data was split chronologically, keeping the temporal context of the data. The training period was from 1997 till 2015, and the validation period was from 2016 till 2017, while the testing period was from 2018 till 2019. Data was not split randomly, to avoid having data from the same climatic periods in both the training as well as testing sets. An approach based on time-series cross validation was used to tune hyperparameters by performing a rolling cross validation across the time series. For this reason, all of the preprocessing steps such as data transformations, encoding and standardisation were done on the training data only to prevent information leakage. Random seeds were fixed along the modelling process to ensure repeatability.

2.8 Model Evaluation and Climate-Sensitivity Analysis

Several measurements were used to assess the model performance, such as coefficient of determination (R^2), root mean square error, mean absolute error, normalised root mean square error, mean absolute percentage error and prediction bias. Performance was evaluated overall and by crop, season, state and by climatic conditions. Feature importance analysis, permutation importance, SHAP values and partial dependence plots were used to interpret the behaviour of the models. These techniques were used to determine important predictors, the non-linearities and interactions between climate and soil variables.

Best model was used to create hypothetical scenarios to test climate sensitivity. These scenarios comprised 1°C or 2°C temperature rise, and ±10% or ±20% rainfall changes, as well as combinations of above scenarios such as temperature and rainfall changes, higher temperatures and lower humidity, excessive rainfall and lower humidity. The scenarios were designed to test the sensitivity, not to make a formal projection of future climate conditions.

2.9 Robustness and Sensitivity Analysis

The strength of the results was assessed by comparing the performance of the models under varying conditions. This encompassed evaluating models with and without soil variables, with and without cultivated areas and with raw and log transformed yield values. Further comparisons were conducted between the pooled model and the crop specific models and between the models for seasons. The effect of adding or dropping coconut data was also investigated and the performance of the models in both the normal and the climate anomaly years were compared. All these analyses helped to guarantee the consistency of the results for the different modelling options and data conditions.

3. Results

3.1 Descriptive Profile of the Agricultural Dataset

The analytical data were integrated to contain 19,652 observations over 30 Indian States and UTs, 55 crop categories and 6 growing seasons from 1997 to 2019. The 2020 data were excluded due to limited observations and not having been nationally representative. The distribution of cultivated area and crop yield was highly skewed right as shown in Table 1. Mean cultivated area was 180,218.60, whereas the median was only 9,333.00. The same said of the mean, which was 80.10, and the median, which was 1.03. The difference between the mean and the media indicated high yielding crops, especially coconut, sugarcane and banana crops. The maximum number of observations were recorded in Kharif followed by rabi and whole year crop. The smallest representation was for autumn and winter crops.

Table 1. Descriptive Profile of the Agricultural Dataset

Category	Indicator	Value
Dataset coverage	Observations	19,652
Dataset coverage	States and union territories	30
Dataset coverage	Crop categories	55
Dataset coverage	Growing seasons	6
Dataset coverage	Study period	1997–2019
Cultivated area	Mean	180,218.60
Cultivated area	Median	9,333.00
Cultivated area	Standard deviation	733,482.30

Cultivated area	Range	0.50–50,808,100.00
Crop yield	Mean	80.10
Crop yield	Median	1.03
Crop yield	Standard deviation	879.13
Crop yield	Range	0.00–21,105.00
Seasonal distribution	Kharif	8,215
Seasonal distribution	Rabi	5,732
Seasonal distribution	Whole year	3,712
Seasonal distribution	Summer	1,190
Seasonal distribution	Autumn	414
Seasonal distribution	Winter	389

3.2 Temporal Patterns in Climatic Conditions

The normal annual courses of temperature, precipitation, and humidity are shown in Figure 1. There was significant variability from year to year in all three variables. However, the temperature had shown no national direction with the fluctuation around its long-term mean, while the rainfall and humidity had been gradually showing positive standardised values in the later years.

The Mann-Kendall analysis showed that there was no statistically significant national temperature trend ($\tau=0.028$), ($p=.876$). For the annual temperature changes in the period between 1905 and 2000, Sen's slope was 0.0007°C per year. The rainfall showed a significant positive trend ($\tau=0.526$), ($p<0.001$) and a rainfall increase of 19.40 mm per year. There was also a significant increase in humidity ($\tau=0.534$), ($p<.001$), of about 0.25 percentage points per year. Results were varied at the state level, so that the national averages masked regional variations.

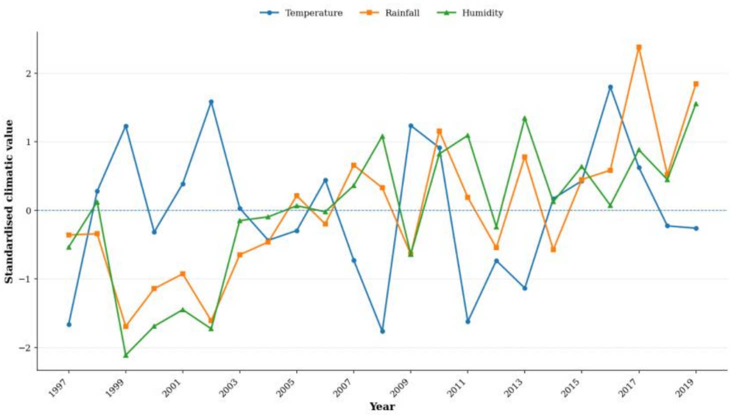

Figure 1. Standardised annual patterns in temperature, rainfall and humidity across Indian states, 1997–2019

3.3 Spatial and Temporal Patterns in Crop Yield

The overall trend of the crop yield was an upward one in the middle of the study period. 2019 had the highest annual median with 1.176 followed by 2017 and 2018. The minimum annual median was 0.880 in 2002. However, there was considerable variation in annual mean yields due to differences in crop composition from year to year.

There was also significant spatial variation. Delhi had the highest median yield among all states followed by Goa, Puducherery and Kerala with their higher means yields due to more representation of the coconut and other high yield crops. The median yield was the lowest in Chhattisgarh at 0.496. Coconut and sugarcane were the top two crops in terms of median yield per crop, followed by banana, tapioca and potato. The seasonal median for whole year crops was the highest while that for the rabi crops was the lowest. Table 2 presents the most noticeable spatial and temporal differences, crop-specific and seasonal differences in the crop yield obtained during the study period.

Table 2. Selected Spatial and Temporal Patterns in Crop Yield

Dimension	Category	Observations	Mean yield	Median yield
Highest-yield year	2019	1,079	74.32	1.176
Lowest-yield year	2002	815	80.66	0.880
Highest-median state	Delhi	203	13.12	2.590
Lowest-median state	Chhattisgarh	915	1.97	0.496
Highest-yield crop	Coconut	172	8,652.00	8,466.316
High-yield crop	Sugarcane	604	51.68	53.152
High-yield crop	Banana	245	26.85	17.949
High-yield crop	Tapioca	201	16.67	11.118
High-yield crop	Potato	626	13.34	10.175
Seasonal pattern	Whole year	3,712	413.52	4.294
Seasonal pattern	Summer	1,190	3.00	1.679
Seasonal pattern	Winter	389	5.29	1.545
Seasonal pattern	Autumn	414	3.92	1.092
Seasonal pattern	Kharif	8,215	2.48	0.924
Seasonal pattern	Rabi	5,732	1.99	0.903

3.4 Relationships Between Climatic Conditions and Crop Yield

The climate yield relationships were far from similar for the various crop groups as evidenced in Figure 2. The groundnut (r_s=0.536), rice (r_s=0.319), and gram (r_s=0.250) showed the highest positive correlation with sugarcane, with temperature coming next (r_s=0.584). Rapeseed and mustard yield was negatively correlated with temperature (r_s=-0.096) and the relationship with soybean was weak and statistically insignificant.

Rainfall was negatively associated with maize, (r_s=-0.221), sugarcane, (r_s=-0.208), rice, (r_s=-0.144), wheat, (r_s=-0.125), and rapeseed and mustard, (r_s=-0.108). Rainfall was positively associated with soybean (r_s=0.123) and groundnut (r_s=0.084). The wheat yield was particularly negatively associated with humidity (r_s=-0.429), and moderately negatively associated with the rapeseed and mustard yield (r_s=-0.233). On the other hand, the groundnut and soybean yields were positively related to humidity. These results showed differences in the responses of climatic variables among crops.

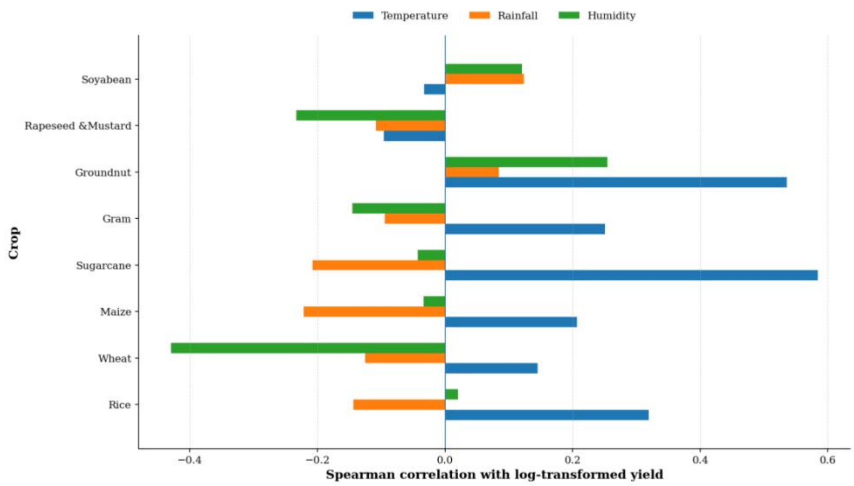

Figure 2. Crop-specific Spearman correlations between climatic variables and log-transformed yield

3.5 Effects of Climate Anomalies on Crop Yield

The comparison of crop productions under normal and anomalous climatic conditions is given in Table 3. The highest crop-standardised log-yield score was observed in high-humidity conditions (0.129), while the highest median yield was observed in high-humidity conditions (11.65%) compared to normal conditions. A positive standardised yield score was also linked to better than average conditions, as well as an 8.88 per cent greater median yield.
The crop-standardised log yield score was slightly negative, but median yield was slightly higher as the warmer weather came. The apparent increase in yield was primarily due to changes in crop composition, as had drier than average conditions with a high raw mean yield but a negative standardised score. The unadjusted median score rose slightly but the combined heat and rainfall stress score was the lowest (-0.120). Thus, the standardised results gave a more satisfactory basis for comparison than raw mean yield.

Table 3. Crop-Yield Differences Under Climate-Anomaly Conditions

Climatic condition	Observations	Mean yield	Median yield	Standardised log yield	Median change from normal
Normal climatic conditions	1,630	56.51	0.982	−0.109	0.00%
Warmer-than-average conditions	3,165	68.29	1.014	−0.028	3.21%
Drier-than-average conditions	2,685	99.06	1.000	−0.096	1.81%
Wetter-than-average conditions	3,659	72.96	1.068	0.080	8.68%
High-humidity conditions	3,343	71.74	1.097	0.129	11.65%
Low-humidity conditions	3,245	86.42	1.000	−0.139	1.81%
Combined heat and rainfall stress	2,265	65.89	1.000	−0.120	1.81%

3.6 Comparative Performance of Forecasting Models

The results of the eight forecasting models were considerably different. The most successful model was XGBoost, which had the lowest log-RMSE of 0.323, according to the prespecified log-RMSE selection criterion. It accounted for 92.59 % of the variation in yield and had an RMSE of 233.13, an MAE of 16.65 and a prediction bias of −3.86. Log-RMSE ranked the Random Forest at #2 with a value of 0.335 and Support Vector Regression at #3 with a value of 0.371.
Original-scale (R2) was the highest and original-scale RMSE was the lowest with Decision Tree Regression at 0.944 and 203.49 respectively. But it had a log-RMSE of 0.395 and a MAPE of 49.47% which were worse than XGBoost. This disparity suggested that the decision tree consistently fit with the large-yield observations but was not consistent over the entire highly skewed yield distribution. Therefore, XGBoost was kept as the main model for forecast.
Linear and regularised regression methods had less potency. Multiple Linear Regression had a log-RMSE of 0.454, Ridge Regression had a value of 0.459 and Lasso Regression had a value of 0.464. The superior performance of XGBoost and Random Forest showed the importance of capturing the nonlinearities and interactions between the crop, climate, soil and spatial variables. The comparative testing performance of the eight crop-yield forecasting models in terms of the log-RMSE is shown in Figure 3.

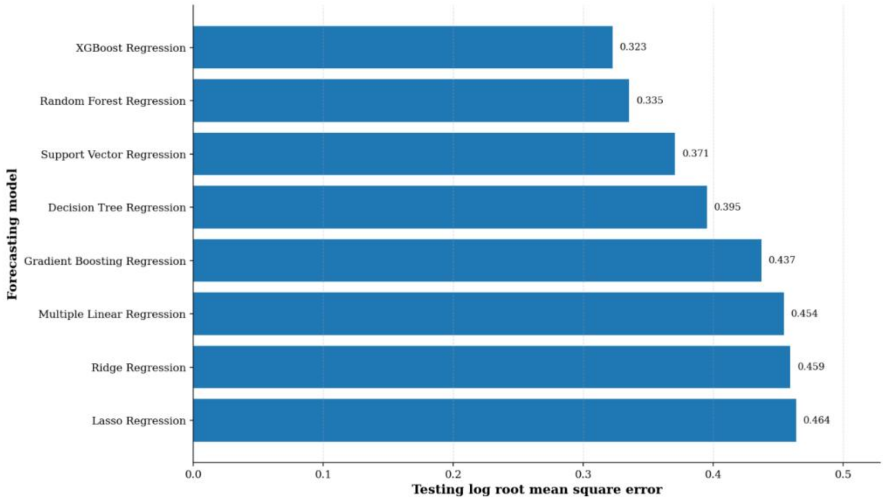

Figure 3. Comparative testing log-RMSE of the crop-yield forecasting models

3.7 Crop-Specific Forecasting Performance

Crop-specific XGBoost models produced different levels of forecasting accuracy, as reported in Table 4. Rice got the minimum log-RMSE of 0.162 and MAPE of 16.57%. Log-RMSEs were also consistent for oilseeds (0.167) and pulses (0.169). Wheat had the highest R-squared value (0.819) among these groups, but with a MAPE of 34.07%.
The sugarcane forecasting was moderately successful with (R2=0.713); the original-scale RMSE was higher though, as the sugarcane yield values were significantly larger than those of cereals and pulses. Maize performed poorly, producing a negative (R2) of -149.686, an nRMSE of 643.48% and a MAPE of 68.25%. A negative (R2) showed that the maize model specific to the crop was doing worse than simple mean-yield benchmark in the testing period.

Table 4. Crop-Specific Testing Performance of the XGBoost Model

Crop or crop group	Test observations	(R^2)	RMSE	MAE	MAPE	Log-RMSE
Rice	122	0.519	0.525	0.361	16.57%	0.162
Oilseeds	408	0.675	0.442	0.249	32.86%	0.167
Pulses	525	0.536	0.562	0.238	26.07%	0.169
Wheat	52	0.819	0.505	0.391	34.07%	0.191
Sugarcane	52	0.713	16.526	10.323	30.83%	0.413
Maize	101	−149.686	18.951	3.165	68.25%	0.477

3.8 Seasonal and Regional Forecasting Performance

The accuracy of the forecast was different according to the season. Autumn had the lowest log-RMSE (0.165, n=46). The highest R2 value was found for the winter season (0.966) and log-RMSE value was 0.198. Summer performed well (R2=0.884), while Rabi showed moderate accuracy (R2=0.665, log-RMSE 0.244). Kharif performed weaker (R2=0.537, log-RMSE 0.358, nRMSE 244.58%). The R2 value was high (0.924), however RMSE (625.51) and log-RMSE (0.420) were large for whole-year crops because of high yielding crops such as coconut. Regionally, performance varied. It is followed by Jharkhand with log-RMSE of 0.119, Arunachal Pradesh with 0.147 and Nagaland with 0.159 with Sikkim having the lowest log-RMSE of 0.032. Delhi had poor performance (log-RMSE 0.688, negative R2, n=15). Errors were also higher in Andhra Pradesh, Telangana, Madhya Pradesh and Puducherry, suggesting a difference in the performance of the models across states.
3.9 Predictor Importance, Climate Scenarios and Robustness
Type of crop was the most significant predictor (1.150), followed by area of crop (0.040) and season (0.039). Phosphorus, nitrogen and temperature were key environmental factors with low importance for rainfall. The mean and median of the predicted yields were 71.29 and 1.149 at the baseline. A 1°C temperature rise reduced yield by 1.76%, and 2°C by 1.42%. The combined 2°C warming and ±20% precipitation resulted in a change in yield of

approximately −1.43% to −1.44%. The decrease was the greatest (−1.83%) under high temperature and low humidity. Rainfall-only effects were minimal (−0.01% to +0.13%). Rainfall had less influence than did temperature and humidity. Overall, the nonlinear models had better performance than the linear models, depending on the crop and region, however, with poorer performance for maize and some states. Figure 4 shows the yield trend in percentage under different climate stress conditions in terms of temperature, rainfall, humidity and combined climate stress scenarios.

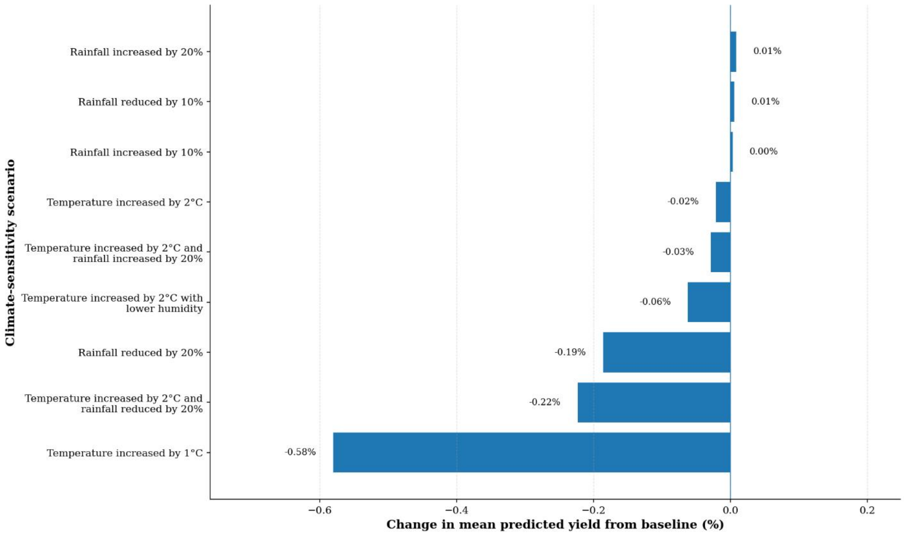

Figure 4. Percentage change in mean predicted crop yield under alternative climate-sensitivity scenarios relative to baseline conditions

4. Discussion

The results indicated that there existed statistically significant trends for increasing rainfall and humidity, but not for the linear changes in the national average temperature. XGBoost had the smallest testing log-RMSE and a well-rounded overall performance. Crop type, cropland area, season, and soil nutrients and temperature were the main factors affecting the predicted yield. This indicates the importance of using machine learning for large-scale yield estimate in diverse agricultural systems (Paudel et al., 2021).

The impact of temperature varied widely across the crops. Sugarcane, groundnut, rice and gram showed positive association; while rapeseed and mustard showed weak negative association. However, under the temperature increases of 1°C and 2°C the scenario analysis did predict yield reductions. This seeming dichotomy reflects nonlinear relationships where moderate warmth can promote growth for some crops, but can still be detrimental if crop-specific heat thresholds are surpassed. Such nonlinear environmental relationships are well-represented by deep neural networks (Khaki & Wang, 2019).

Rainfall impacts varied by crop and were not necessarily favorable. There was a positive correlation between rainfall and standardised yields for maize, rice, wheat and sugarcane and a negative correlation for all other crops. Too much rainfall can lead to waterlogging of soil, loss of nutrients and disease; too little rainfall means soil-water is limited. Such geographical unevenly distributed rainfall effects have already changed agriculture production due to climate change (Ray et al., 2019).

The highest standardised yield score was linked to high humidity, and combined heat–rainfall stress gave a negative score for yield. But a combination of higher temperatures and lower humidity resulted in the biggest scenario-based yield loss. Under moderate levels, humidity can enhance crop-water balance, but it can also create disease pressure if it gets too high. This interaction makes models that can capture temporal climatic dependencies justified (Khaki et al., 2020).

The performance forecasting was different for each crop such as rice, wheat, maize, sugarcane, pulses and oilseeds. Maize had a poor performance while rice had the lowest error. Seasonal differences were related to differences in growing season, irrigation requirements and exposure to monsoon conditions. Crop-group-specific indicators can thus enhance regional projections by identifying the unique phenological and climatic needs of specific crops (Ronchetti et al., 2023).

Crop identity and season were the most important factors in the model, with nitrogen, phosphorus, potassium and soil pH also providing some contribution in the prediction. Spatial variations between the states were due to the differences in agroclimatic conditions, cropping pattern and soil characteristics. The incorporation of satellite data, weather and location-specific information has also helped in enhancing forecasting performance at county level (Ju et al., 2021).

XGBoost was superior to traditional regression models as it was able to model nonlinearities, interactions and threshold effects without specifying fixed functional relationships. Decision Tree Regression was consistently good on the original scale, but was less consistent across the skewed yield distribution. Agricultural predictors have also been shown to be advantageous for deep convolutional approaches over simpler ones when the agricultural predictors are complex (Nevavuori et al., 2019). A combination of crop knowledge and machine learning make them even more stable and generalisable (Shahhosseini et al., 2021).

Predicted mean yield reductions were modest for all warming scenarios and the rainfall-only scenarios resulted in comparatively small changes. The greatest reduction was with the combined effects of warming and reduced humidity. The findings corroborate previous evidence showing that anthropogenic climate change has reduced the rate of growth of agricultural productivity and exacerbate production risks in the future (Ortiz-Bobea et al., 2021).

The findings used for climate adaptive crop selection, sowing date adjustments, irrigation scheduling, drought preparation and adaptation at state level. Since yield responses are due to interacting environmental conditions, soil and nutrient management should be included in climatic adaptation. The use of remote-sensing data to complement field level monitoring and timely management decision making is supported (Marshall et al., 2022).

The forecasting framework a valuable tool for farmers to help with production planning and for governments and extension agencies to help identify risks early. It also help crop-insurance companies and food-security planners and market authorities predict supply changes. Uncertainty-aware prediction is especially useful when prediction results are inputs for high-risk agricultural decision-making (Ma et al., 2021).

The key strengths were the longitudinal representation, representation of multiple crops and states, climate–soil integration, chronological validation, multiple model comparison, and multiple climate-sensitivity scenarios. The multimodal structure is in line with the fact that the combination of complementary agricultural information can enhance the yield estimation ability (Maimaitijiang et al., 2020).

Annual state level weather and static soil data were used for analysis rather than growing season or field level measurements. Some of the irrigation, extreme-weather and detailed management variables were unavailable and observations for 2020 were incomplete. Yield skewness was also caused by differences in the measurements of crops. The use of multi-temporal remote-sensing data help overcome these limitations (Qiao et al., 2021).

Future studies should incorporate daily weather, soil moisture, irrigation and district level observations/vegetation indices. In addition to the projections from the climate models (CMIP6) and scenarios (SSP), the LSTM, CNN-RNN and spatiotemporal models should be evaluated. Real-time satellite monitoring also help operate early-warning systems and more accurate yield forecasts.

5. Conclusion

Crop yield forecasting based on climate information is a useful tool for analysing and managing agricultural production under climate change. The study showed significant variation in yield in time, space, season and across various crops based on 19,652 observations from 55 crops in 30 states and union territories and years 1997-2019. The increasing trends for rainfall and humidity were found to be highly significant while the national average temperature did not reveal any statistically significant linear trend. However, temperature was still an important predictor and caused a negative yield in the warming scenarios developed. XGBoost was the most balanced forecasting method with the lowest testing Log-RMSE and accounted for a high percentage of the yield variation among the forecasting methods considered. Crop type, crop area, crop growing season, soil nutrients and temperature came out as the key forecasting factors. But prediction accuracy were significantly different for different crops and areas. The forecasts for maize continued to be weak while rice and wheat, pulses and oilseeds were more reliably forecasted, indicating that the pooled model performance should not be uniformly generalised. Climate-scenario analysis suggested that increases in temperature (in combination with decreased humidity or rainfall stress) lead to losses in agricultural productivity. Consequently, crop selection, sowing schedules, irrigation planning, nutrient management and regional adaptation should be based on climate-sensitive forecasting. The incorporation of high-resolution weather, remote-sensing and soil moisture and future climate-projection data can further bolster the architecture of agricultural early-warning, food-security planning and climate-resilient decision making.

References

1. Ansarifar, J., Wang, L., & Archontoulis, S. V. (2021). An interaction regression model for crop yield prediction. Scientific Reports, 11(1), 17754.
2. Baffour-Ata, F., Antwi-Agyei, P., Nkiaka, E., Dougill, A. J., Anning, A. K., & Kwakye, S. O. (2021). Effect of climate variability on yields of selected staple food crops in northern Ghana. Journal of Agriculture and Food Research, 6, 100205.
3. Bowden, C., Foster, T., & Parkes, B. (2023). Identifying links between monsoon variability and rice production in India through machine learning. Scientific Reports, 13(1), 2446.
4. Chang, Y., Latham, J., Licht, M., & Wang, L. (2023). A data-driven crop model for maize yield prediction. Communications Biology, 6(1), 439.

5. Chen, Y., & Tao, F. (2022). Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large area. Field Crops Research, 276, 108398.
6. Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environmental Research Letters, 13(11), 114003.
7. Dhakar, R., Sehgal, V. K., Chakraborty, D., Sahoo, R. N., Mukherjee, J., Ines, A. V., Kumar, S. N., Shirsath, P. B., & Roy, S. B. (2022). Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing. Agricultural Systems, 195, 103299.
8. Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. Computers and Electronics in Agriculture, 155, 257–282.
9. Gavahi, K., Abbaszadeh, P., & Moradkhani, H. (2021). DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. Expert Systems with Applications, 184, 115511.
10. Gopal, P. M., & Bhargavi, R. (2019). A novel approach for efficient crop yield prediction. Computers and Electronics in Agriculture, 165, 104968.
11. Heino, M., Kinnunen, P., Anderson, W., Ray, D. K., Puma, M. J., Varis, O., Siebert, S., & Kummu, M. (2023). Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. Scientific Reports, 13(1), 3583.
12. Iniyan, S., Varma, V. A., & Naidu, C. T. (2023). Crop yield prediction using machine learning techniques. Advances in Engineering Software, 175, 103326.
13. Jhajharia, K., & Mathur, P. (2022). Machine learning approaches to predict crop yield using integrated satellite and climate data. International Journal of Ambient Computing and Intelligence (IJACI), 13(1), 1–17.
14. Ju, S., Lim, H., Ma, J. W., Kim, S., Lee, K., Zhao, S., & Heo, J. (2021). Optimal county-level crop yield prediction using MODIS-based variables and weather data: A comparative study on machine learning models. Agricultural and Forest Meteorology, 307, 108530.
15. Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621.
16. Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN framework for crop yield prediction. Frontiers in Plant Science, 10, 1750.
17. Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sensing of Environment, 259, 112408.
18. 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.
19. Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., & Nelson, A. (2022). Field-level crop yield estimation with PRISMA and Sentinel-2. ISPRS Journal of Photogrammetry and Remote Sensing, 187, 191–210.
20. Miles, A. F., & Moore, E. R. H. (2026). Food system resilience, disaster preparedness & response. In Frontiers in Sustainable Food Systems (Vol. 10, p. 1797675). Frontiers Media SA. https://www.frontiersin.org/journals/sustainable-food-systems/articles/10.3389/fsufs.2026.1797675/full
21. Mishra, A. (2025). Crop yield Data with Soil And Weather Dataset. https://www.kaggle.com/datasets/anshumish/crop-yield-data-with-soil-and-weather-dataset
22. Nevavuori, P., Narra, N., & Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture, 163, 104859.
23. Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G., & Lobell, D. B. (2021). Anthropogenic climate change has slowed global agricultural productivity growth. Nature Climate Change, 11(4), 306–312.
24. Paudel, D., Boogaard, H., De Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. Agricultural Systems, 187, 103016.
25. Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R. A., Betts, R., Kerr, R. B., & Biesbroek, R. (2022). Climate change 2022: Impacts, adaptation and vulnerability. https://www.researchgate.net/profile/Vincent-Moeller/publication/362431678_Climate_Change_2022_Impacts_Adaptation_and_Vulnerability_Working_Group_II_Contribution_to_the_Sixth_Assessment_Report_of_the_Intergovernmental_Panel_on_Climate_Change/links/6529092e06bdd619c48be291/Climate-Change-2022-Impacts-Adaptation-and-Vulnerability-Working-Group-II-Contribution-to-the-Sixth-Assessment-Report-of-the-Intergovernmental-Panel-on-Climate-Change.pdf
26. Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Zhang, L., & Tian, Z. (2021). Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks. International Journal of Applied Earth Observation and Geoinformation, 102, 102436.
27. Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., & Chatterjee, S. (2019). Climate change has likely already affected global food production. PloS One, 14(5), e0217148.
28. Ronchetti, G., Manfron, G., Weissteiner, C. J., Seguini, L., Scacchiafichi, L. N., Panarello, L., & Baruth, B. (2023). Remote sensing crop group-specific indicators to support regional yield forecasting in Europe. Computers and Electronics in Agriculture, 205, 107633.
29. Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports, 11(1), 1606.