

Machine-Learning-Based Climate Impact Analysis for Agriculture

Divyanshu Yadav

Computing Science and Engineering,
Galgotias University, Greater Noida, India

Abstract—To explore the effect of climate on the agricultural yield, analysis and machine learning of observation year foreground data for 1500 years for 5 crops from 15th states in India, with climate background data of 20 years (2005-2024) is performed. The data is synced with temperature data, rainfall, humidity data, soil moisture, heatwave days, dry spell days, evapotranspiration, irrigation area, NDVI proxy, climate stress index, crop and yield. Alternatively, trained Ridge, random forest and gradient boosting and fixed 80:20 split. The Gradient Boosting Regressor model got the best test result with RMSE, MAE, R2 and MAPE of 0.2325 t/ha, 0.1826 t/ha, 0.9712 and 5.31%, respectively. Results show that productivity at planting, plant water availability, climate stress, vegetation condition and temperature related pathways influence yield response. The study offers a reproducible climate-agriculture modelling workflow, but necessary observed datasets are found to be needed for policy-grade modelling inferences.

Keywords—Agriculture, climate change, crop yield, data-driven analysis, machine learning, prediction modeling.

I. INTRODUCTION

A. Climate Variability and Agricultural Production

It now operates at a fluctuating temperature regime and water regime which affects farming activities. The phenological stages are negatively affected, vapour pressure deficit and water need of the crop before harvest are increased by increased mean temperature [1]. There is a second limitation - rainfall variability. Even if rainfall happens to occur within the range for the same soaking period, one or both occurrences of field might not happen due to either sowing window or flowering/reproductive window, thus leading to the yield loss. Humidity is another factor of importance. Disease pressure can be caused by high humidity, and transpiration is increased, which can hasten moisture loss from the soil at low humidity. These pressures turn into reoccurring production shocks as there are more droughts with irrigation buffering the effect on production differently in the rain-fed regions [2]. Extreme heat and heavy rainfall impacts are not independent of each other, but are related to soil moisture, canopy temperature, evapotranspiration and to the management response [3]. The key change here is that in this paper, yield is being used as a basis of measured climate phenomenon interacting; yield is not an extension of seasonal rainfall. The framing of the adoptive done in Fig.1.1.

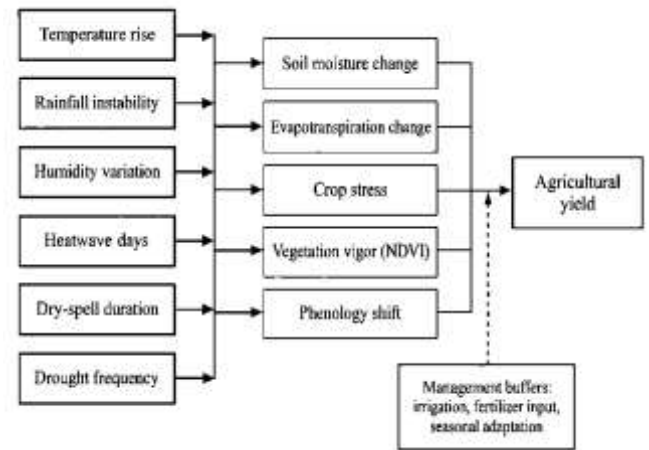


Fig. 1.1. Conceptual Relationship Between Climate Variables and Agricultural Yield

B. Role of Machine Learning in Climate-Agriculture Assessment

The direction and average impact can be gauged through traditional regression, but the climate-yield relationship is not always linear for crops, zones and seasons. Machine learning can be useful here since it does not demand a single predetermined functional form for the data it processes and can learn interaction patterns between the data, which include the temperature, rainfall, humidity, moisture, vegetation condition and crop type [4]. This study is based on a representative experimental data that consist 1500 observations of five crops, five years over a period of 20 crop years from the States of India. The temperature range, annual and growingseason rainfall, rainfall anomaly, heatwave with number of days, dry-spell with number of days, humidity, soil moisture, evapotranspiration, irrigation area, NDVI proxy, climate stress index and yield are the variables used. The structure of the response is given by Eq. (1) as yield is modelled as a function of climate, crop, regional attributes [5,6]. That is not to say that there is universal agronomic causality in one model. More specific: Make a clear prediction experiment, with stress and vegetation/management buffers being tested vs. yield, climate stresses. Each variable, split, model, metric and residual can be checked prior to their interpretation. For this purpose, the size of the error of the validations is shown together with the explanatory scores and a fixed test set is maintained in order to avoid weak validations and thus direct numerical traceability.

$$Y_i = f(T_i, R_i, H_i, S_i, D_i, C_i) + \epsilon_i \quad (1)$$

II. RELATED WORK

A. Climate Change Effects on Crop Productivity

A new research on the effects of climate change on agriculture points out that the trend is going toward less biomass, not just as a consequence of direct changes in temperature, but as a result of heat stress, dislocations of raindrop and of course deficit of water [7]. Time is of paramount importance in the productivity of any crop. The heat that occurs during the flowering stage may influence the

number of grains formed; biomass production may be reduced and exposures to disease may increase due to the water stress during the vegetative development stage and late rains. Evidence elsewhere suggests that the mitigation potential can be affected by the irrigation, cultivar, soil status and farming management [8]. Based on the results obtained from these, a set of variables is required to account for not only exposure to climate, but buffers in production as well. The factors of stress duration, soil moisture and vegetation status are not included in these data if they were only annual temperature and rainfall.

B. Machine Learning Models in Agricultural Prediction

From single regression equation, to ensemble learning, deep learning and remote-sensing assimilation [9] is the trend that has been seen in machine-learning literature in the agriculture sector since the inception. Random Forest and Gradient Boosting are used because they are capable to learn nonlinear splits, variable interaction and variable threshold behavior without requiring any manual alterations based on a response curve [10]. Support vector regression is applicable where the DB is small and medium, if the data are time series or an image, then Recurrent or Convolutional Networks are useful. Under poor seasonal discrimination by the climatic variables alone, the vegetation performance through the UAV, MODIS and cubeSat can produce better prediction of early-yield. More recent works also demonstrate that a comparison of models (based on error, explanatory power and residual structure) should be used and reported together [13].

C. Research Gap in Integrated Climate-Yield Modeling

A lot of the current work focuses on the analysis and prediction modelling of the impacts on climate. Severely, climate papers talk about an estimation of the aggregate productivity loss while papers based on machine learning talk about accuracy and do not analyse representation of the climate stress within the feature space [14]. An open but minor issue is the same workflow used for quantifying climate exposure and to encode agronomic response variables should also be applied to train alternative models, to compare test set errors, to interpret feature contribution should follow a single reproducible structure. This paper aims to bridge this gap by providing a data set that combines climate, crop, vegetation, management as well as yield data. The phenomenon of this gap is nothing new for statistics. The missing piece is the notion of operational alignment. The framework to be used should support the ability of a reader to trace the steps of each climate variable to the trained estimator and ultimately to statements about climate impacts to include transparent, publication-level audit conditions/replication protocols and the ability to attach errors. The stresses indices in Eqs. (1 and 2) is the measure to include the climatic severity in model ready numerical measure.

$$CSI_i = w_1 HT_i + w_2 RD_i + w_3 DS_i + w_4 SM_i \quad (2)$$

III. RESEARCH METHODOLOGY

A. Data Collection and Variable Selection

The experiment set was a pre-processed data set used for climate yield modelling and not measured as a field. It comprises 1,500 crop year records from five Indian crops, 15 states, for the period 2005-24. Here crops of rice, wheat, maize, soya and cotton are cultivated. There is a mix of region, agro-climatic zone, crop season, location (geographic coordinates), climate exposure, vegetation condition, management proxy, and yield response found in all records. The climate variables selected are mean temperature, maximum temperature,

minimum temperature, annual precipitation, growing season precipitation, precipitation anomaly, heat wave days, dry-spell days, relative humidity, soil moisture, evapotranspiration and the CO₂ concentration in atmosphere. Irrigation coverage, fertilizer-use index, peak NDVI, NDVI anomaly, baseline yield, climate stress index, crop, season and state are all considered agronomic and management variables. Before predicting, environmental, phenological and remote sensing prostates are combined in this variable set which are recently used in the yield modelling [16] [17]. At the heart of approach is disclosing data sets. The Data in sheet are experimental and were synthesized to be consistent with published literature and indeed the document calls them as such, and they are used for checking the sensitivity of the algorithm; their reproducibility; and for testing the algorithm on experimental data. No table or figure is available for which there is an official series maintained by the Government. That is, in the model this boundary is maintained because the model is able to synthesize a workflow of a climate-yield assessment, however, in the use of models in applied modelling further verification of the observed records from meteorological and agricultural agencies, before making a claim in the region, is required for interpretation. The normalization that has been used to compare climate variables with different measurement units is the one given in Eq. (3) above.

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} \quad (3)$$

B. Data Preprocessing and Feature Engineering

The pre-processing started from column level inspection, duplicate checking, range inspection and type inspection. Once created, there were no missing data points in the created dataset, though it is worth noting that the workflow had the option of including a missing value audit that can be beneficial if there are any missing data points in the field data replacement. Linear modelling was done on the numerical variables while one-hot encoding was performed on the categorical variables. So that, mechanically no points have been deducted. All notes for the topic of Climate extremes are retainable observations – impossible values would only be removed. Climate stress, rainfall anomaly, NDVI anomaly, baseline yield and seasonal grouping were generated and maintained as a feature engineering. All the above-preprocessing photoflow is displayed in Fig. 3.1. Weather, vegetation and agronomic structure, when combined, has been supported by recent studies based in Deep-Learning and Hybrid yield [18] [19]. Then the scaling form of a modelling pipeline is identified by Eq. (4).

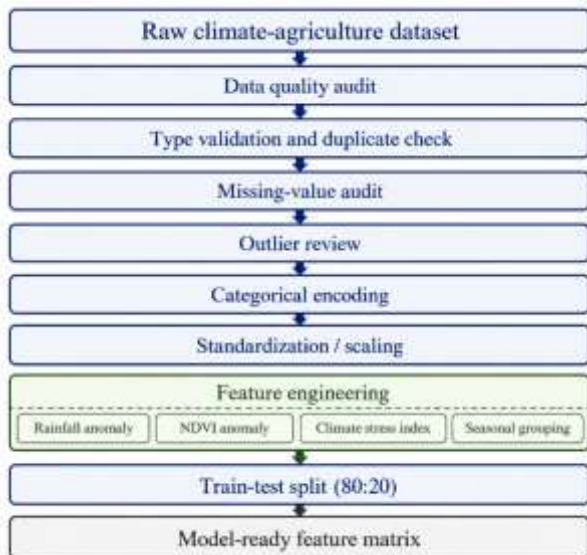


Fig. 3.1. Proposed Data Preprocessing and Feature Engineering Pipeline

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

C. Machine Learning Model Development

The following regression models have been created: ridge regression, gradient boosting regression and random forest regression. A ridge regression was used to obtain linear regularization of the baseline. Random forest was used to test bagged nonlinear tree aggregation. Sequential error correction was employed in shallow trees and this was tested with gradient boosting. With fixed random seed 42, 80% of the data set was divided into training data set and remaining 20% was used as testing data set. Yield (tonnes/ha) was the variable measured. The training pipeline was integrated along with the process of preprocessing and estimation, where no leakage occurs between the raw variables and variables ready for the models. When making predictions from tests, the same fitted transformation was used for all of them. As seen in figure 3.2, the modelling work flow is being illustrated. This model set is reused consistent to the current agriculture prediction practice, based on comparisons between ensemble of trees and statistical bench marks [20, 21].

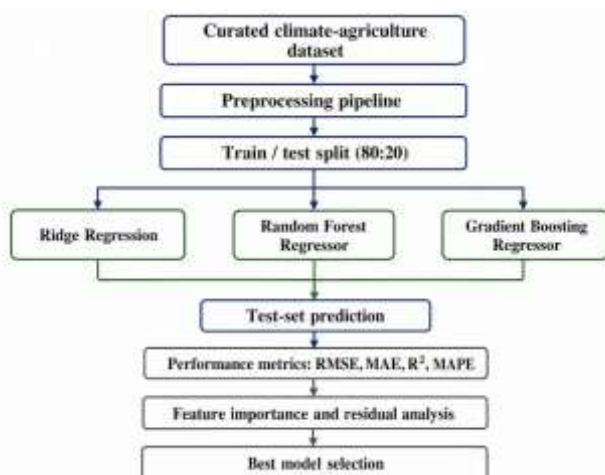


Fig. 3.2. Machine Learning Workflow for Climate-Based Crop Yield Prediction

D. Model Evaluation Strategy

The type of evaluation was RMSE, MAE, R^2 and MAPE. Large errors were penalized using RMSE, MAE and R^2 while MAPE was used as a relative measure of prediction error to

measure the average absolute error. To check the best model, the model which has minimum RMSE is considered and the MAE and R^2 are also considered. Inspection was achieved in the prediction output, as well as the residuals, the model metrics and the feature-importance values, in separate files. This is in line with the recent suggestions that, in the case of a yield prediction being interpreted in terms of climate risk from the trained model [23] [22], the distribution of prediction error should be made public along with the measure of accuracy. Form of RMSE that is used in model selection is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (5)$$

IV. RESULTS AND DISCUSSION

A. Descriptive Analysis of Climate and Yield Variables

There are 1,500 observations in the data set. Information provided in the central profile is shown in Table 4.1. Average temperature of 25.128 °C is during the growing-season. The variability of rainfall is low during the growing season, with a standard deviation of 270.295 mm, while the yield varies from 0.847 to 9.211 t/ha. The mean of the climate stress index value is 0.182 and its max is 0.510, indicating that there are stress cases visible. Average temperature, rainfall and yield for Kharif and Rabi are given in Fig.4.1, which shows that temps, rainfall and yield, are all relatively high in Kharif and lower in Rabi. This will make the use of season and crop as the active predictors easier [24].

Table 4.1.

Descriptive Statistics of Climate and Agricultural Variables

Variable	Mean	SD	Minimum	Maximum
Average temperature (°C)	25.128	1.846	19.490	29.230
Growing-season rainfall (mm)	699.231	270.295	192.300	1722.400
Soil moisture (%)	36.605	7.137	15.060	58.000
Climate stress index	0.182	0.092	0.003	0.510
Yield (t/ha)	3.775	1.367	0.847	9.211



Fig. 4.1. Seasonal Temperature and Rainfall Variation Pattern

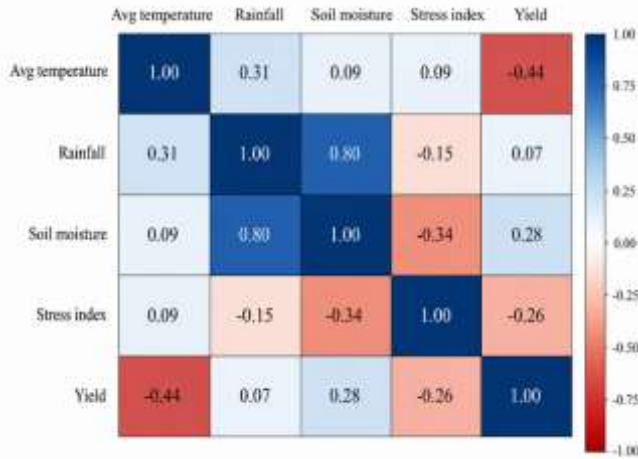
B. Correlation Between Climate Indicators and Crop Yield

Table 4.2 reports associations. The yield and the temperature has a negative relationship (-0.443). The soil is well moistured (+0.281). There was also low rainfall of 0.071 which implies that rainfall-only is not suitable to serve as yield signal during the season. Stress index is negative; -0.257. The correlation form is written down in Eq. (6). The pattern helps to have a big picture understanding of weather and weather induced stress attributes [25].

Table 4.2.

Correlation Matrix Between Climate Variables and Crop Yield

Variable	Temperature	Rainfall	Soil moisture	Stress index	Yield
Temperature	1.000	0.310	0.095	0.093	-0.443
Rainfall	0.310	1.000	0.803	-0.147	0.071
Soil moisture	0.095	0.803	1.000	-0.336	0.281
Stress index	0.093	-0.147	-0.336	1.000	-0.257
Yield	-0.443	0.071	0.281	-0.257	1.000


Fig. 4.2. Correlation Heatmap of Climate and Yield Variables

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

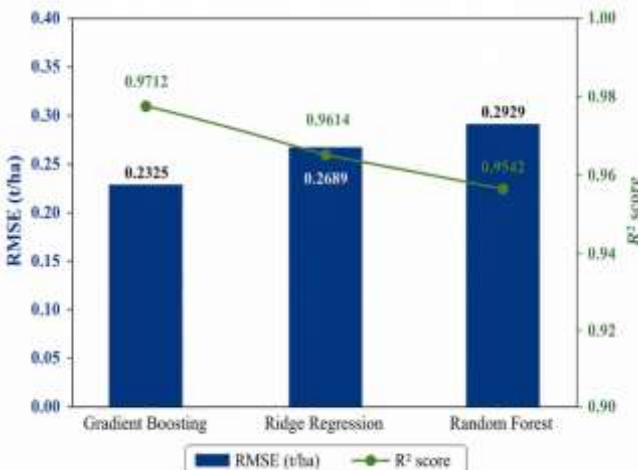
C. Machine Learning Model Performance Comparison

The comparison of the trained regressors is done in Table 4.3. Gradient Boosting Regressor performs best, with RMSE of 0.2325 t/ha, MAE of 0.1826 t/ha, R^2 of 0.9712, and MAPE of 5.31%. Yield response is a quasi-linear effect - the ridge regression works. Performance on this split is poorer than other random forest. The RMSE and R^2 is presented in Fig. 4.3. The variance of the exercises (Eq. (7)) is expressed in r.i.u, and is also high, as evidenced by the agriculture ML surveys, run recently with the mixed climatic and vegetation inputs [26],[27].

Table 4.3.

Comparative Performance of Machine Learning Models

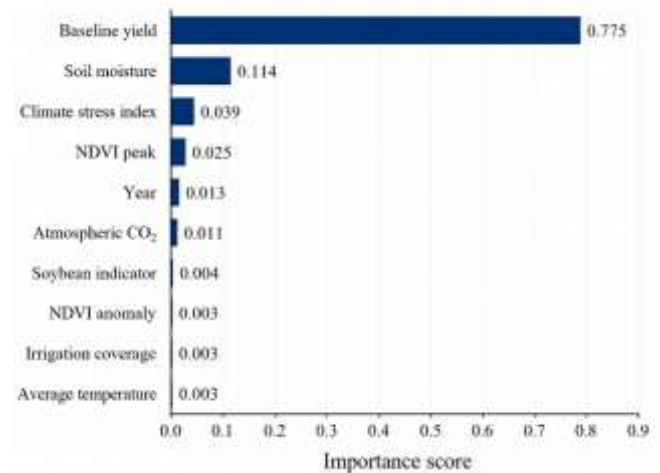
Model	RMSE	MAE	R^2	MAPE
Gradient Boosting Regressor	0.2325	0.1826	0.9712	5.31
Ridge Regression	0.2689	0.2002	0.9614	6.47
Random Forest Regressor	0.2929	0.2275	0.9542	6.36


Fig. 4.3. Model Performance Comparison Using RMSE and R^2 Score

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (7)$$

D. Climate Impact Interpretation Based on Model Outputs

Feature importance shows that the baseline yield is the most important feature followed by soil moisture, climate stress index, peak NDVI, year, carbon dioxide, irrigation, temperature, evapotranspiration, fertilizer index, and dry-spell days, and heatwave days. The scores for these are provided in Fig. 4.4. Ordering technically is appropriate from the agricultural aspect. Crop and regional capacity was measured by baseline yield, Water-energy pressures was measured by soil moisture and stress index and vegetation condition was measured by NDVI. Warming is not in isolation. The side of moisture and stress pathways had been coupled, which is also mentioned in the research related to the yield of AI and XAI [28] and [29].


Fig. 4.4. Feature Importance Scores for Climate-Based Yield Prediction

E. Validation of Prediction Results and Error Distribution

The results of some test observations are provided in the following table, Table 4.4. Residuals are generally small with two being greater than -0.35 or +0.35 t/ha. The test full test residual mean is determined to be -0.0083 t/ha with standard deviation of 0.2328 t/ha. The comparison of observed and predicted yield is given in Fig. 4.5. The value of the residual is provided by Eq. (8). For deployment, official record of observation with an external-year validation work is required; and good potential for experimental analysis. This boundary is physical, as simulated records only test method consistency not the differences in the farmer-level loss claims/state compensation designs due to policy and/or insurance.

Table 4.4.

Observed and Predicted Crop Yield Values Under Selected Climate Conditions

Year	State	Crop	Temp	Rainfall	Stress	Observed	Predicted	Residual
2007	Uttar Pradesh	Soybean	25.27	602.6	0.140	1.757	2.134	-0.377
2011	Odisha	Wheat	21.83	929.0	0.281	3.946	4.336	-0.390
2019	Punjab	Maize	24.46	365.9	0.234	5.148	5.008	0.140
2023	Gujarat	Soybean	26.43	557.6	0.235	2.055	1.675	0.380
2024	Gujarat	Cotton	25.98	630.9	0.058	3.578	3.705	-0.127

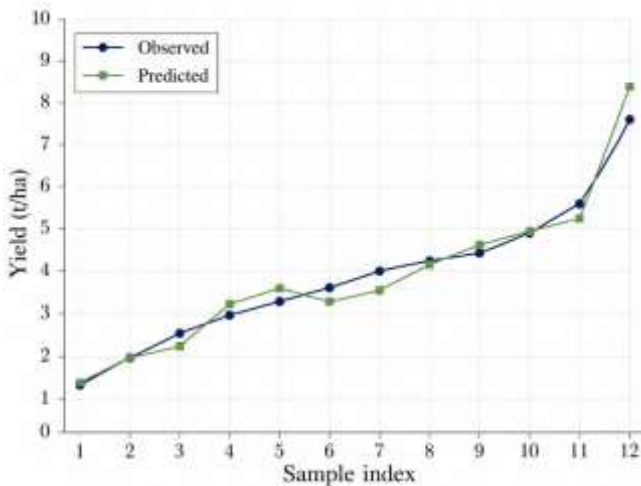


Fig. 4.5. Observed Versus Predicted Crop Yield Curve

$$e_i = Y_i - \hat{Y}_i \quad (8)$$

V. CONCLUSION AND FUTURE WORK

In this study, a climate impact analysis on agricultural yield has been carried out by using a curated experimental dataset which consists of 1,500 crop-year observations for five crops in 15 Indian states for the study period 2005–2024 using machine learning approaches. The modeling of the behavior of photoclimatic yield for four types of climate exposure, two seasonal structures, and five types of vegetation and vegetal status and management buffers was carried out. Well defined differences in temperature, precipitation, soil moisture, stress index and yield were found in the descriptive results. A negative correlation between yield and temperature and climate stress factors, and a positive correlation between yield and soil moisture, were found. Looking at moisture retention, stress timing and duration is important and not only rainfall depth for a season alone, as rainfall alone was weak.

The modelling results showed that the Gradient Boosting Regressor had highest test statistics results; RMSE (0.2325 t/ha), MAE (0.1826 t/ha), R2 (0.9712) and MAPE (5.31%). Ridge regression did not diverge much while random forest exhibited higher error. Some of the large factors were identified as baseline yield, soil moisture, climate stress index, NDVI, year, carbon dioxide, irrigation and temperature. Technically speaking, the result has no problem: “Climate does not effect one variable each other”. This is reflected in temperature, water and plant & regional capacity indicators.

The approach is descriptive and is limited in its informativeness. The data is simulated and should only be utilized as an experimental validation data set, and not field data. The next steps for the future would be to replace the experimental record with the Before where they use it for policy or insurance purposes, and to include district level mapping as well as uncertainty intervals and explainable output from the model. Also, with the increased capacity of this work, the exposition of the climate adaptation possibilities (expansion of irrigation, heat-tolerant cultivars, sowing date change and soil-water conservation) is also improved. These additions would move the model to more geographical, applied and disciplined assessment work.

REFERENCES

- [1] J. Jägermeyr, C. Müller, A. C. Ruane, J. Elliott, J. Balkovič, O. Castillo, *et al.*, “Climate impacts on global agriculture emerge earlier in new generation of climate and crop models,” *Nature Food*, vol. 2, pp. 873–885, 2021, doi: 10.1038/s43016-021-00400-y.
- [2] A. Ortiz-Bobea, T. R. Ault, C. M. Carrillo, R. G. Chambers, and D. B. Lobell, “Anthropogenic climate change has slowed global agricultural productivity growth,” *Nature Climate Change*, vol. 11, pp. 306–312, 2021, doi: 10.1038/s41558-021-01000-1.
- [3] E. E. Rezaei, S. Siebert, and C. Müller, “Integrated assessment of climate change impacts on crop yields and crop water productivity in Europe,” *Agricultural Systems*, vol. 193, Art. no. 103210, 2021, doi: 10.1016/j.agsy.2021.103210.
- [4] G. S. Malhi, M. Kaur, and P. Kaushik, “Impact of climate change on agriculture and its mitigation strategies: A review,” *Sustainability*, vol. 13, no. 3, Art. no. 1318, 2021, doi: 10.3390/su13031318.
- [5] D. Ozdemir, “The impact of climate change on agricultural productivity in Asian countries: A heterogeneous panel data approach,” *Environmental Science and Pollution Research*, vol. 29, no. 6, pp. 8205–8217, 2022, doi: 10.1007/s11356-021-16291-2.
- [6] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, “A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction,” *IEEE Access*, vol. 9, pp. 63406–63439, 2021, doi: 10.1109/ACCESS.2021.3075159.
- [7] L. Benos, A. C. Tagarakis, G. Doliias, R. Berruto, D. Kateris, and D. Bochtis, “Machine learning in agriculture: A comprehensive updated review,” *Sensors*, vol. 21, no. 11, Art. no. 3758, 2021, doi: 10.3390/s21113758.
- [8] K. Alibabaei, P. D. Gaspar, and T. M. Lima, “Crop yield estimation using deep learning based on climate big data and irrigation scheduling,” *Energies*, vol. 14, no. 11, Art. no. 3004, 2021, doi: 10.3390/en14113004.
- [9] M. G. Ziliani, M. U. Altaf, B. Aragon, R. Houborg, T. E. Franz, Y. Lu, J. Sheffield, I. Hoteit, and M. F. McCabe, “Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model,” *Agricultural and Forest Meteorology*, vol. 313, Art. no. 108736, 2022, doi: 10.1016/j.agrformet.2021.108736.
- [10] P. Muruganantham, S. Wibowo, S. Grandhi, N. H. Samrat, and N. Islam, “A systematic literature review on crop yield prediction with deep learning and remote sensing,” *Remote Sensing*, vol. 14, no. 9, Art. no. 1990, 2022, doi: 10.3390/rs14091990.
- [11] Y. Shen, B. Mercatoris, Z. Cao, P. Kwan, L. Guo, H. Yao, and Q. Cheng, “Improving wheat yield prediction accuracy using LSTM-RF framework based on UAV thermal infrared and multispectral imagery,” *Agriculture*, vol. 12, no. 6, Art. no. 892, 2022, doi: 10.3390/agriculture12060892.
- [12] J. Wang, H. Si, Z. Gao, and L. Shi, “Winter wheat yield prediction using an LSTM model from MODIS LAI products,” *Agriculture*, vol. 12, no. 10, Art. no. 1707, 2022, doi: 10.3390/agriculture12101707.
- [13] Y. Di, M. Gao, F. Feng, Q. Li, and H. Zhang, “A new framework for winter wheat yield prediction integrating deep learning and Bayesian optimization,” *Agronomy*, vol. 12, no. 12, Art. no. 3194, 2022, doi: 10.3390/agronomy12123194.
- [14] A. Nyéki and M. Neményi, “Crop yield prediction in precision agriculture,” *Agronomy*, vol. 12, no. 10, Art. no. 2460, 2022, doi: 10.3390/agronomy12102460.
- [15] F. Huber, A. Yushchenko, B. Stratmann, and V. Steinhage, “Extreme Gradient Boosting for yield estimation compared with deep learning approaches,” *Computers and Electronics in Agriculture*, vol. 202, Art. no. 107346, 2022, doi: 10.1016/j.compag.2022.107346.
- [16] A. K. Srivastava, N. Safaei, S. Khaki, G. Lopez, W. Zeng, F. Ewert, T. Gaiser, and J. Rahimi, “Winter wheat yield prediction using convolutional neural networks from environmental and phenological data,” *Scientific Reports*, vol. 12, Art. no. 3215, 2022, doi: 10.1038/s41598-022-06249-w.

- [17] A. Tripathi, R. K. Tiwari, and S. P. Tiwari, "A deep learning multi-layer perceptron and remote sensing approach for soil health-based crop yield estimation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 113, Art. no. 102959, 2022, doi: 10.1016/j.jag.2022.102959.
- [18] A. Oikonomidis, C. Catal, and A. Kassahun, "Deep learning for crop yield prediction: A systematic literature review," *New Zealand Journal of Crop and Horticultural Science*, vol. 51, no. 1, pp. 1–26, 2023, doi: 10.1080/01140671.2022.2032213.
- [19] K. Jhajharia and P. Mathur, "Machine learning approaches to predict crop yield using integrated satellite and climate data," *International Journal of Ambient Computing and Intelligence*, vol. 13, no. 1, pp. 1–17, 2022, doi: 10.4018/IJACI.300799.
- [20] S. Iniyar, V. Akhil Varma, and C. Teja Naidu, "Crop yield prediction using machine learning techniques," *Advances in Engineering Software*, vol. 175, Art. no. 103326, 2023, doi: 10.1016/j.advengsoft.2022.103326.
- [21] K. Jhajharia, P. Mathur, S. Jain, and S. Nijhawan, "Crop yield prediction using machine learning and deep learning techniques," *Procedia Computer Science*, vol. 218, pp. 406–417, 2023, doi: 10.1016/j.procs.2023.01.023.
- [22] M. Kuradusenge, E. Hitimana, D. Hanyurwimfura, P. Rukundo, K. Mtonga, A. Mukasine, C. Uwitonze, J. Ngabonziza, and A. Uwamahoro, "Crop yield prediction using machine learning models: Case of Irish potato and maize," *Agriculture*, vol. 13, no. 1, Art. no. 225, 2023, doi: 10.3390/agriculture13010225.
- [23] A. Morales and F. J. Villalobos, "Using machine learning for crop yield prediction in the past or the future," *Frontiers in Plant Science*, vol. 14, Art. no. 1128388, 2023, doi: 10.3389/fpls.2023.1128388.
- [24] M. R. Mahmood, M. A. Matin, S. K. Goudos, and G. K. Karagiannidis, "Machine learning for smart agriculture: A comprehensive survey," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 6, pp. 2568–2588, 2024, doi: 10.1109/TAI.2023.3345278.
- [25] M. A. Javed and M. A. A. Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," *Heliyon*, vol. 10, Art. no. e40836, 2024, doi: 10.1016/j.heliyon.2024.e40836.
- [26] U. V. Nikhil, A. M. Pandiyan, S. P. Raja, and Z. Stamenkovic, "Machine learning-based crop yield prediction in South India: Performance analysis of various models," *Computers*, vol. 13, no. 6, Art. no. 137, 2024, doi: 10.3390/computers13060137.
- [27] K. Meghraoui, I. Sebari, J. Pilz, K. Ait El Kadi, and S. Bensiali, "Applied deep learning-based crop yield prediction: A systematic analysis of current developments and potential challenges," *Technologies*, vol. 12, no. 4, Art. no. 43, 2024, doi: 10.3390/technologies12040043.
- [28] S. Ajith, M. K. Debnath, and R. Karthik, "Statistical and machine learning models for location-specific crop yield prediction using weather indices," *International Journal of Biometeorology*, vol. 68, pp. 2453–2475, 2024, doi: 10.1007/s00484-024-02763-w.
- [29] R. N. V. J. Mohan, P. S. P. S. Rayanoothala, and R. P. Sree, "Next-gen agriculture: Integrating AI and XAI for precision crop yield predictions," *Frontiers in Plant Science*, vol. 15, Art. no. 1451607, 2025, doi: 10.3389/fpls.2024.1451607.
- [30] M. Ashfaq, I. Khan, R. F. Afzal, D. Shah, S. Ali, and M. Tahir, "Enhanced wheat yield prediction through integrated climate and satellite data using advanced AI techniques," *Scientific Reports*, vol. 15, Art. no. 18093, 2025, doi: 10.1038/s41598-025-02700-w.