Development of Machine Learning based Models for Multivariate Prediction of Wheat Crop Yield in Uttar Pradesh, India

Sukirti[1]. Kamal Pandey[2*], Abhishek Danodia[3], . Harish Chandra Karnatak[2]
[1]Indian Institute of Remote Sensing (IIRS), Dehradun
[2]Department of Geoweb Services, IT and Distance Learning, IIRS, ISRO, Dehradun
[3] Agriculture and Soils Department, IIRS, Dehradun
*kamal@iirs.gov.in

(Received: 21 April 2023; in final form 27 November 2023)

DOI: *https://doi.org/10.58825/jog.2023.17.2.70*

Abstract: The consequences of climate change have a substantial impact on agricultural crop production and management. Predicting or forecasting crop yields well in advance would help farmers, agriculture corporations and government agencies manage risk and design suitable crop insurance plans. Ground survey is the traditional way of determining yield, which is subjective, time-consuming, and expensive. While Machine Learning (ML) techniques make yield prediction less expensive, less time taking and more efficient. In this study, thirteen years of meteorological parameters and wheat yield data (2001-2013) of Uttar Pradesh were used to train and analyze three Machine Learning Regression models viz. Support Vector Regression, Ordinary Least Squares, and Random Forest. Each model's performance was assessed using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Results revealed that the Random Forest model with a MAE of 0.258 t/ha, MSE of 0.096 t/ha and RMSE of 0.311 t/ha proved to be the best model in the yield prediction of wheat when results are statistically compared with others. Researchers and decision-makers can use the findings to estimate pre-harvest yields and to ensure food security.

Keywords: Meteorological parameters, Wheat, Multivariate, Yield prediction, Machine Learning, Random Forest.

1. Introduction

India has a long history of agriculture (Madhusudan, 2015) dating back to the Indus Valley Civilization. Agriculture sector is very important for India's socioeconomic structure as it is one of the largest economic sectors in terms of population diversity. The inability to integrate technology to provide the intended results is one of the biggest challenges Indian agriculture is facing (Nigam et. al., 2019). Due to the uneven climatic trends brought on due to the negative impacts of global warming, which has affected crop production as well, farmers find it challenging to precisely predict temperature and rainfall patterns for their agricultural practices. Farmers play important role to ensure food security for the human civilization. Food security is provided through agriculture, and is crucial for a nation's economic growth (Barman, 2020). Since the launch of Earth Observation (EO) satellites and the explorer missions that followed, satellite remote sensing has enabled us gather remote sensing data at improving scales (Karthikeyan et. al., 2020; Shetty et. al., 2021; Virnodkar et. al., 2020; Feizizadeh et. al., 2023; Jamali 2019; Chatziantoniou et. al., 2017; Ienco et. al., 2019.

Global temperature is rising as a result of the continuing rise in greenhouse gas emissions. Some effects of climate change include shifting seasons, melting glaciers, increasing precipitation, rise in extreme weather occurrences, and so on. World population and socioeconomic growth, along with climate change, represent a danger to food security. An extremely high rate of land degradation brought on by climate change is creating accelerated desertification and nutrient-deficient soils. The problem of land degradation is said to be a serious global threat and is getting worse day by day (Arora 2019). Rising temperature is also leading to gradual decline in annual crop yield. Crop failures are more likely to occur in the short term as a result of altered precipitation patterns, and diminished crop productivity \(Poudel and Shaw 2016).

Until recently, yield prediction was being done by considering the farmer's knowledge of a particular area and crop. Empirical and crop growth models (Jørgensen, 1994) were created using meteorological data to for the yield prediction, but these techniques also have a lot of problems owing to the spatial distribution of the weather stations. The outputs from these operations are accessible only after the crops have been harvested,. Therefore, it is essential to decrease expenses and longer waits involved in the traditional practices while also improving yield prediction accuracy. For this purpose, machine learning techniques provide an alternate option with improved and proven performance matrix. (Nigam et. al., 2019; Jaafar and Mourad 2021; Cunha et. al., 2018; Pantazi et. al., 2016).

Researchers across the globe have attempted and developed machine learning techniques for crop yield forecasting. Preseason forecasting was a technique developed that used a machine learning-based system to predict soybean yields before the start of the crop season (Cunha et. al., 2018). The system was based on a recurrent neural network (RNN) that was trained using historical data on municipal-level soybean and/or maize yields as well as parameters such as precipitation, temperature, and soil conditions. A reanalysis-based seasonal forecast product of temperature and precipitation, which enabled predicting up to seven months in advance, providing the meteorological data operationally (Cunha et. al., 2018). The findings were on par with, and in some cases better than, similar models that are limited to early season forecasting due to the need for remote sensing data across the farm. This study concentrated on the quantification of machine learning algorithms and their practical application. The method mentioned also considered the unpredictable rainfall and temperature in order to obtain a steady trend. On the basis of mean absolute error, the results of different algorithms were compared. The most accurate regressor for predicting yield was found to be the Random Forest Regressor. A sequential model called Simple Recurrent Neural Network performed better at forecasting rainfall than the LSTM for forecasting the temperature (Nigam et. al., 2019).

A technique to anticipate agricultural yield based on historical data was designed and implemented by Bondre et al (2019) Using agricultural data, machine learning methods like Support Vector Machine and Random Forest were used to determine the recommended fertiliser for a number of crops (Bondre and Mahagaonkar 2019). It concentrated on developing a prediction model that may be applied to crop yield forecasting in the future. Different sources were used to collect a variety of datasets, including those for crops, crop yield, geography, soil and crop nutrients, and fertiliser.

Haque et al (2020) suggested two distinct Machine Learning (ML) algorithms in their study to examine the crop yield (Haque et.al., 2020). With 140 data points collected, the Support Vector Regression (SVR) and Linear Regression (LR) algorithms proved to be quite effective for evaluating the performance of the parameters involved in the forecast. The Mean Square Error (MSE) and Coefficient of Determination were used to calculate the error rate (R^2) and the same dataset was used to directly compare the results obtained from the ML models. Pantazi et al (2016) attempted a study wherein online multi-layer soil data and crop growth parameters from satellite imagery were used to forecast within-field variation in wheat yield. An unsupervised learning system was combined with supervised self-organizing maps that could handle data from numerous soil and crop sensors. The effectiveness of XY-fused Networks (XY-Fs), Supervised Kohonen Networks (SKNs), and Counter-Propagation Artificial Neural Networks (CP-ANNs) for predicting wheat production in a field in Bedfordshire, UK, was compared during a single cropping season (Pantazi et. al., 2016).

In the field of crop yield analysis, machine learning (ML) is a new topic of research. By providing inputs for growing the best possible crop and anticipating the yields, ML has the potential to transform agriculture (Nigam et. al., 2019) to new dimension. In order to help farmers for selecting a suitable crop for cultivation to obtain maximum yield, ML model takes into consideration parameters like temperature, rainfall, area, etc. The ML based approaches has the potential to improve the expanding agricultural industry of countries like India and, taken together, raise the living standard of farmers. The objectives of the presented study are in line with the above argument viz. (i) To develop machine learning models to predict yield of wheat for the state of Uttar Pradesh, and (ii) To compare the performance of different models and to find out the best one for multivariate analysis of yield prediction. The study aims to promote wide spread use of ML models in decision making in a farming sectors for the countries like India where agriculture has a majoe share in economy.

2. Materials and Methods

2.1 Study Area

Uttar Pradesh (UP) is the fourth largest state of India with an area of 240,928 km^2. UP is the most populous state of India. It is located between latitude 24º to 31º North and longitude 77º to 84º East (Figure 1).

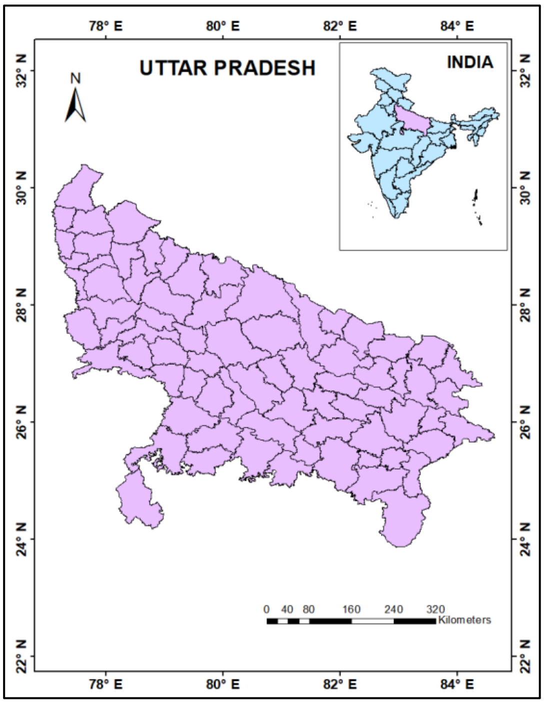

Figure 1. Map of Study Area (Uttar Pradesh)

There are 18 divisions and 75 districts in Uttar Pradesh. Due to the predominately agrarian economy, Uttar Pradesh is prominently dependent on the performance of industries like agriculture, horticulture, animal husbandry, dairy, fishery, etc. With four distinct seasons, Uttar Pradesh has a humid subtropical climate. The state receives an average of 650 mm of rainfall per year in the southwest and 1000 mm in the eastern and south eastern regions (Nihar et.al., 2022).

Approximately 47% of the population is directly dependent on agriculture for their livelihood, and climate is the primary factor affecting the production. Given the size of the state's geographic area and the access to the fertile Indo-Gangetic plains, UP makes a considerable contribution to the country's food security. The state produces about 12% of India's rice and 28% of it is wheat. Additionally, a significant amount of sugarcane is produced, making up 44% of the nation's total output (Gulati et.al., 2021). The major crops of the state are rice, wheat, maize, sugarcane, chickpea and pigeon pea. About 24% of the state's agricultural area is used to grow wheat throughout the state.. On an average the total geographical area under wheat crop is 9730.60 ha, total production is 32799.71 tons and total yield is 3371 kg/ha.

2.2 Data Used

Wheat yield was predicted using net sown area, and meteorological variables such as mean rainfall, mean temperature, mean relative humidity, mean wind speed, mean wind direction, and surface net solar radiation (for Rabi Season). The Indian government website (data.gov.in) provided historical district-level wheat yield data (2001-2013). The Climate Change Service (https://climate.copernicus.eu/), a website run by the ECMWF as part of The Copernicus Program, was used to collect all the meteorological data for the time period of 2001-2013. The datasets were gridded with hourly temporal resolution and a horizontal resolution of 0.5° x 0.5°. The ECMWF Copernicus services provide ever-changing datasets by continuously monitoring the climate and atmospheric composition and using climate and atmosphere reanalysis..

2.3 Data Pre-processing

The crop production data that were downloaded from data.gov.in were district- and season-specific, and they date back to 1997. Required yield data of wheat crop of Uttar Pradesh from 2001-2013 was extracted from the downloaded dataset. The collected data was organized in a Comma Delimited (CSV) format before being used with other meteorological parameters to form the dataset used for this study.

2.3.1 Principal Component Analysis

In order to study, model, and explain complex multivariate real-world systems, various statistical methods are available and Principal Component Analysis (PCA) is one of the widely used method to study the collinearity among the parameters . It is a useful tool for conducting data analysis on huge data sets and for identifying the main trends and the variables that influence them. It can quickly determine which primary components contribute the most to data variability. Instead of performing the whole process of attempting to estimate which of the variables might be of the most relevance, effort might instead be spent on trying to find and comprehend these few most critical components (Bloomer and Rehm 2014). In this research study, PCA was used to identify and address multicollinearity issues.

2.3.2 Data Splitting

The primary requirement of any ML model is data splitting, and the available dataset must be divided into three parts:

a. Using the **training set**, the model is trained or fitted. The optimal weights or coefficients for ML models can be found using the training set.
b. The **validation set** is utilised to improve the model performance by adjusting the the hyperparameters used in the model. For each potential setting of the hyperparameters, the model is fitted using the training set, and the performance is assessed using the validation set.
c. The **test set** is necessary for a fair assessment of the final model..

In cases where the dataset is less complicated and tuning of hyperparameters is not required , only the training and test sets can be used. Accordingly, the data array was divided into two subsets, one for training and the other for testing, using the Sklearn library's data splitting function known as train test split. By default, it divides the two subsets into random divisions. In this study, 80% of the data were used for training, with the remaining data being utilised to test the algorithm's performance..

2.4 Machine Learning Algorithms

The models were built using long-term datasets for the years from 2001 to 2013. The presented research work is focused on developing yield estimation models for wheat using the three ML models viz. Random Forest (RF), Ordinary Least Squares (OLS), and Support Vector Regression (SVR),.. The performance of these models were further compared to select the best performing model.

2.4.1 Random Forest

Random Forest is an ensemble based ML model wherein multiple decision trees are combined for producing aggregated output.. Ensemble learning is the process of combining the predictions or classifications of different models that have all been trained on the same data. The underlying principle of ensemble learning is that each model's errors—in this case, the errors of a decision tree—are separate and unrelated to one another. A Random Forest model's forecast is produced by averaging the predictions of various decision trees and hence there isn't a single equation that summarizes the RF model.
Similar study has been done in which Random Forests and Multiple Linear Regressions were compared, for crop production prediction, for wheat, maize, and potato at global and regional scales in response to meteorological and biophysical variables. They trained the RF models to predict crop yield using a variety of biophysical predictors (Jeong et. al., 2016).

2.4.2 Ordinary Least Squares

Ordinary Least Square is one of the popular regression model in ML. Building a model that lowers the overall squared deviations between the expected and actual data is an approach used to estimate unknown parameters. OLS tries to reduce this sum (Brinkhoff and Robson 2021). It is a linear regression method that aims to minimize the sum of squared residuals between the predicted values and the actual values. The equation (1) for OLS is:

$$y = w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n + \varepsilon \quad (1)$$

Here, y is the predicted output, $x_1, x_2,..., x_n$ are the input variables, $w_0, w_1, w_2,..., w_n$ are the coefficients (weights) to be learned, and ε represents the error term. The study carried out by Sharma and team (2013) utilized OLS for the yield prediction of rainfed maize and soyabeans crops and reported the outputs with high accuracy. (Sharma et. al., 2013).

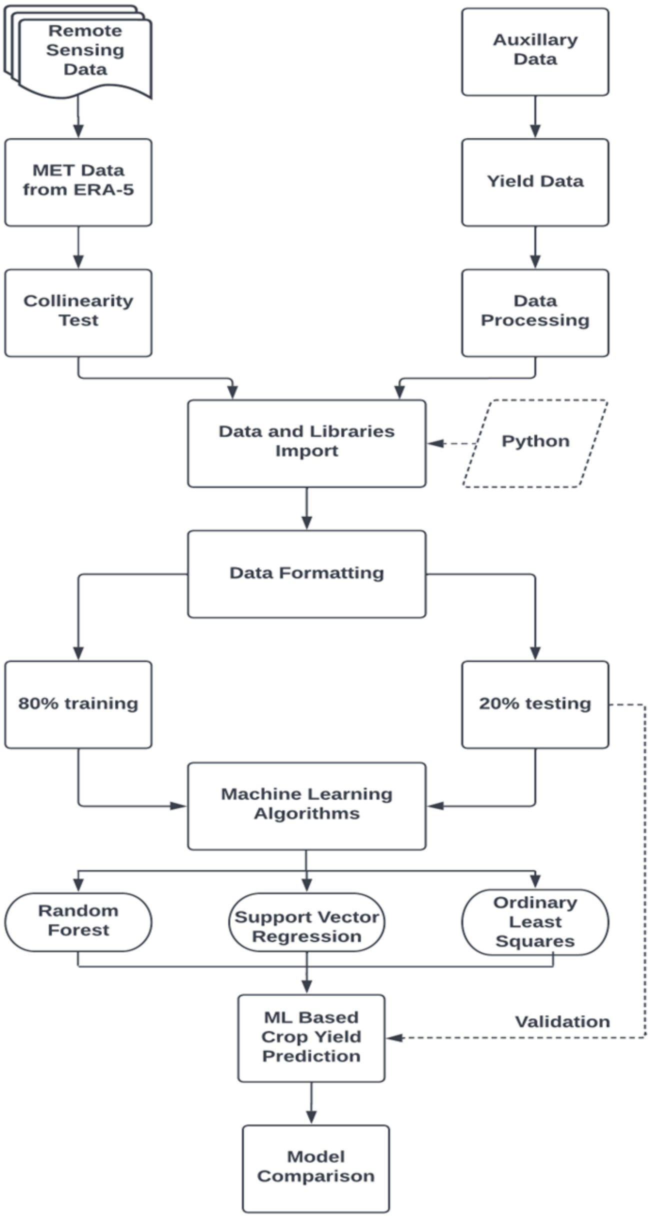

Figure 2. Flowchart of Methodology

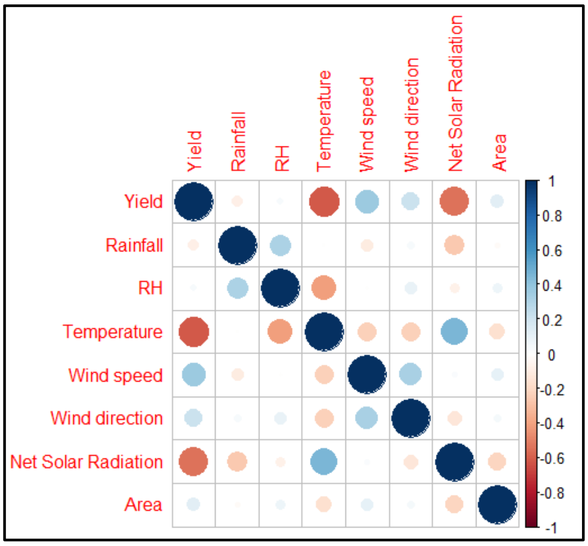

Figure 3. Correlation Matrix of Independent Variables

2.4.3 Support Vector Regression

A well-known machine learning approach called Support Vector Machine (SVM) is widely utilized in both classification and regression. The Support Vector Regression (SVR) aims to fit the best line within a threshold value, in contrast to other regression models that aim to minimize the difference between the actual and predicted value. Equation (2) is the generic representation for the line in support vector regression:

$$y = wx + b \tag{2}$$

SVR has been used in similar studies, a study in Andhra Pradesh, India which concentrated on the forecasting of important Kharif crops employed modular artificial neural networks to first anticipate the quantity of monsoon rainfall before applying support vector regression and rainfall data to calculate the potential yield of the main kharif crops. They concluded that the proposed strategy outperformed earlier machine learning algorithms in estimating kharif crop production after comparing their study with other ML methods (Khosla et.al., 2020).

3. Results and Discussion

3.1 Correlation Analysis

Collinearity check among the participating parameter is carried out before developing a predictive model in order to reduce the redundancy. If the correlation among different variables is high, it gives rise to multicollinearity problems and estimates are unstable. Therefore, to examine the co-linearity between the independent variables and prevent the multicollinearity issue, principal component analysis (PCA) is used in the present study. It was found that there was no significant correlation between the variables as shown in Figure 3..

3.2 Importance of Predictor Variables

The significance of these predictor variables was examined using the Random Forest model's characteristics. These factors were ranked according to their importance, and it was observed that each variable has its importance for yield estimation differently (Figure 4). Surface net solar radiation is observed to be the most crucial factor in the yield estimation, followed by mean temperature, mean wind speed, and area. Since wheat is a rabi (winter) crop, temperature and net solar radiation both have a significant impact on the crop's ability to grow. Strong winds can lead to lodging of the crop and hence mean wind speed also plays very important role in for the overall crop yield.

3.3 Model Comparison

The density scatter plots for each algorithm are shown in Figures 5. Among the models, RF demonstrated the highest predictive accuracy, as evidenced by its scatter plot exhibiting the strongest alignment between observed and predicted yield. The data points are closely clustered around a central line, with only a few outliers indicating a robust relationship between the two variables. This was further supported by the highest coefficient of determination (R^2) achieved by the RF model. On the other hand, SVR performed relatively poorly as compared to the other models and as evident in its scatter plot. The lower R^2 value associated with the SVR model indicated a weaker correlation and less accurate predictions. These findings highlight the superior performance of the RF model in predicting crop yield, while underscoring the limitations of SVR in this particular context.

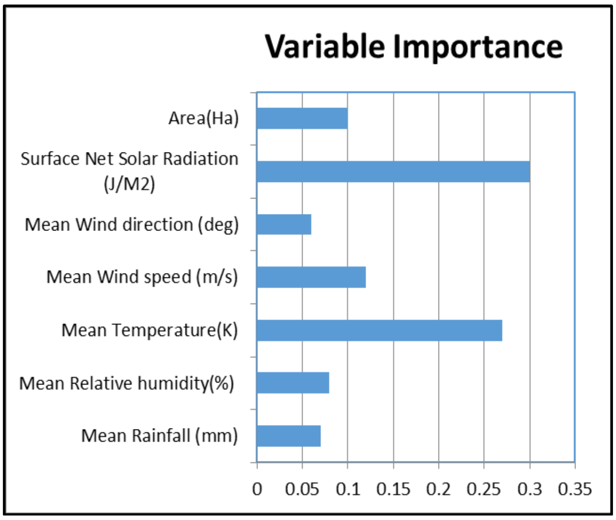

Figure 4. Importance of Predictor Variables

Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) of yield forecasts were estimated for the ML models and out of all three models, RF showed the best predictive capability for wheat yield ($R^2 = 0.72$).

3.3 Discussion

The ability of these ML models (RF, OLS, and SVR) to predict the yield of wheat in Uttar Pradesh was tested using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Results for the ML models are summarized in Table 1. Random forest successfully predicted the wheat yield and was compared against the test data, which was not used in training the model. Comparison between the observed and predicted yield (Figure 5a) produced an R^2 of 0.72. The model was able to explain with MAE value of 0.258 t/ha, MSE value of 0.096 t/ha and RMSE value of 0.311 t/ha.

The method we followed in this study was found similar to the method adopted by Schwalbert et al 2020. They found that combining meteorological information with multi-temporal satellite imagery could produce useful data, thus enabling the creation of more precise yield forecast models, which is consistent with our work. Using a regression model between historical meteorological data and yield data for food crops, the effects of climate change were investigated in the mountainous regions of Nepal. They came to the conclusion that different crops' yield is affected differently by climate variables. However, the current climate trends have a negative impact on winter crops (Poudel and Shaw 2016).

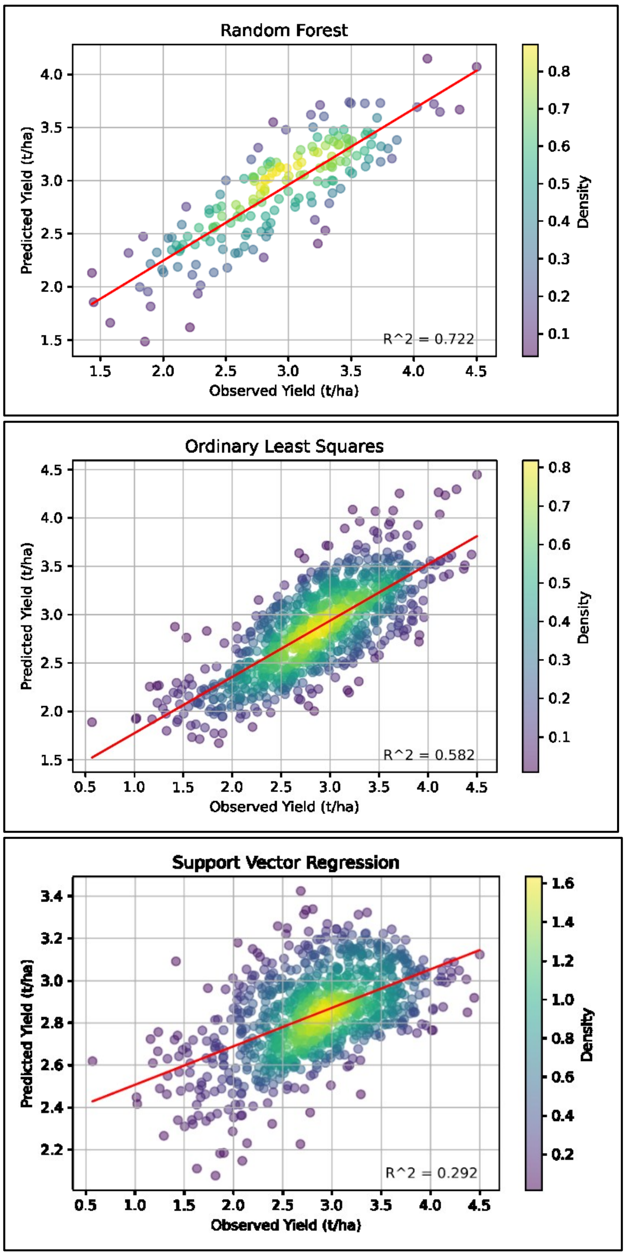

Figure 5. Scatter plot of Predicted yield against Observed Yield for (top) Random Forest, (middle) Ordinary Least Squares. And (bottom) Support Vector Regression

Table 1. Prediction Performance of Different Algorithms

Algorithm	MAE (t/ha)	MSE (t/ha)	RMSE (t/ha)
Random Forest	0.258	0.096	0.311
Support Vector Regression	0.439	0.304	0.552
Ordinary Least Squares	0.312	0.156	0.395

Cabas et al 2010 concluded in their study that the non-climatic variables had a relatively minor impact on the yield distribution, indicating that climatic factors should dominate this relationship (Cabas et. al., 2010). A system to forecast agricultural yield based on historical data has also been proposed by Bondre and Mahagaonkar (Bondre and Mahagaonkar 2019). They used agriculture data and machine learning techniques like Support Vector Machine and Random Forest and obtained the accuracy of 99.47% and 97.48% for . SVM and RF respectively. Therefore, SVM method was found to be good for agricultural yield prediction as opposed to our result. This might be due to the difference in predictor variables.

4. Conclusion

. With the help of machine learning algorithms and publicly available data, this study established a methodology for estimating winter wheat production on a regional scale that might possibly be used to estimate crop yield globally and in regions with low frequency of observed data. By merging crop models with additional data for crop yield estimation, forecasting, and catastrophe monitoring in vast areas, the framework can be further enhanced. In order to estimate winter wheat yield from 2001 to 2013 along with 07 parameter of winter season viz. mean temperature, mean relative humidity, mean pressure, mean wind speed, mean wind direction, mean precipitation and mean solar radiation, three ML models (Support Vector Regression, Ordinary Least Squares, and Random Forest) were used in this study. Using three error metrics—Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE)—accuracy was compared to assess each model's performance. Overall, all three models performed well but RF model performed the best. With MAE values of 0.258,

MSE values of 0.096, and RMSE values of 0.311, it was found that RF outperformed SVR and OLS. The poorest result was obtained by SVR. This study has the potential to open up new possibilities for productive research on agricultural production prediction utilizing a variety of meteorological factors. It will encourage and promote research into the use of machine learning techniques for yield forecasting.

Acknowledgements:
Authors express their sincere thanks to Copernicus (https://scihub.copernicus.eu/) for providing satellite data at no cost. The authors are also thankful to the Indian Institute of Remote Sensing for providing the laboratory facilities for the completion this work.

Competing Interests: We, the authors declare that we have no competing interests.

References

Arora N.K. (2019), "Impact of climate change on agriculture production and its sustainable solutions," *Environ. Sustain.*, vol. 2, no. 2, pp. 95–96, doi: 10.1007/s42398-019-00078-w.

Barman S. (2020), "The Political Economy of Food Security in India: Evolution and Performance, International Journal of management, 11(12), pp 1156-1162"

Bloomer C. and G. Rehm (2014), "Using principal component analysis to find correlations and patterns at diamond light source," *IPAC 2014 Proc. 5th Int. Part. Accel. Conf.*, pp. 3719–3721

Bondre D. and S. Mahagaonkar (2019), "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," International Journal of Engineering, Applied Science and technology vol. 4(5), pp. 371–376

Brinkhoff J. and A. J. Robson (2021). "Block-level macadamia yield forecasting using spatio-temporal datasets," *Agric. For. Meteorol.*, vol. 303, doi: 10.1016/j.agrformet.2021.108369.

Cabas J., A. Weersink and E. Olale (2010). "Crop yield response to economic, site and climatic variables," *Clim. Change*, vol. 101, no. 3, pp. 599–616, 2010, doi: 10.1007/s10584-009-9754-4.

Chatziantoniou A., G. P. Petropoulos and E. Psomiadis (2017). "Co-Orbital Sentinel 1 and 2 for LULC mapping with emphasis on wetlands in a mediterranean setting based on machine learning," *Remote Sens.*, vol. 9, no. 12,

Cunha R.L.F, B. Silva and M. A. S. Netto (2018). "A scalable machine learning system for pre-season agriculture yield forecast," *Proc. - IEEE 14th Int. Conf. eScience, e-Science* pp. 423–430, doi: 10.1109/eScience.2018.00131.

Feizizadeh B., D. Omarzadeh, M. Kazemi Garajeh, T. Lakes and T. Blaschke (2023). "Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine," *J. Environ. Plan. Manag.*, vol. 66, no. 3, pp. 665–697, doi: 10.1080/09640568.2021.2001317.

Gulati A., P. Terway and S. Hussain (2021) "Performance of agriculture in Uttar Pradesh", In: A. Gulati, R. Roy and S. Saini (eds) Revitalizing Indian Agriculture and Boosting Farmer Income. Indian Studies in Business and Economics, Springer

Haque F.F., A. Abdelgawad, V. P. Yanambaka and K. Yelamarthi (2020). "Crop Yield Analysis Using Machine Learning Algorithms," *IEEE World Forum Internet Things, WF-IoT 2020- Symp. Proc.*, doi: 10.1109/WF-IoT48130.2020.9221459.

Ienco D., R. Interdonato, R. Gaetano and D. Ho Tong Minh (2019). "Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 11–22, doi: 10.1016/j.isprsjprs.2019.09.016.

Jaafar H. and R. Mourad (2021). "Gymee: A global field-scale crop yield and et mapper in google earth engine based on landsat, weather, and soil data," *Remote Sens.*, vol. 13, no. 4, pp. 1–30, doi: 10.3390/rs13040773.

Jamali A. (2019). "Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran," *SN Appl. Sci.*, vol. 1, no. 11, doi: 10.1007/s42452-019-1527-8.

Jeong J.H, J.P. Resop, N.D. Mueller, D.H. Eleisher, K. Yun, E.E. Butler, D.J. Timlin, K. Shim, J.S. Gerber, V.R. Reddy and S. Kim (2016). "Random forests for global and regional crop yield predictions," *PLoS One*, vol. 11(6), doi: 10.1371/journal.pone.0156571.

Jørgensen S.E. (1994). "Models as instruments for combination of ecological theory and environmental practice," *Ecol. Modell.*, vol. 75–76, no. C, pp. 5–20, doi: 10.1016/0304-3800(94)90003-5.

Karthikeyan L., I. Chawla and A.K. Mishra (2020). "A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses," *J. Hydrol.*, vol. 586, 2020, doi: 10.1016/j.jhydrol.124905.

Khosla E., R. Dharavath and R. Priya (2020). "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environ. Dev. Sustain.*, vol. 22, no. 6, pp. 5687–5708, doi: 10.1007/s10668-019-00445-x.

Madhusudan L. (2015). "Agriculture role on Indian..Economy. - *Business and economics journal*, vol 6, no. 4

Nigam A., S. Garg, A. Agrawal and P. Agrawal (2019) "Crop yield prediction using machine learning algorithms," Fifth International Conference of Image Information Processing. Shimla, 125-130

Nihar A., N. R. Patel and A. Danodia, (2022). "Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India," *J. Indian Soc. Remote Sens.*, vol. 50, no. 8, pp. 1519–1530, doi: 10.1007/S12524-022-01549-0.

Pantazi X.E., D. Moshou, T. Alexandridis, R. L. Whetton and A. M. Mouazen (2016). "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agric.*, vol. 121, pp. 57–65, doi: 10.1016/j.compag.2015.11.018.

Poudel S. and R. Shaw (2016). "The relationships between climate variability and crop yield in a mountainous environment: A case study in Lamjung District, Nepal," *Climate*, vol. 4, no. 1, doi: 10.3390/cli4010013.

Schwalbert R.A., T. Amado, G. Corassa, L.P. Pott, P.V.V. Prasad and I.A. Ciampitti (2020). "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil," *Agric. For. Meteorol.*, vol. 284, doi: 10.1016/j.agrformet.2019.107886.

Sharma V., D. R. Rudnick and S. Irmak, (2013). "Development and evaluation of ordinary least squares regression models for predicting irrigated and rainfed maize and soybean yields," *Trans. ASABE*, vol. 56, no. 4, pp. 1361–1378, doi: 10.13031/trans.56.9973.

Shetty S., P.K. Gupta, M. Belgiu and S. K. Srivastav (2021). "Assessing the effect of training sampling design on the performance of machine learning classifiers for land cover mapping using multi-temporal remote sensing data and google earth engine," *Remote Sens.*, vol. 13, no. 8, doi: 10.3390/rs13081433.

Virnodkar S.S., V. K. Pachghare, V. C. Patil and S. K. Jha (2020). "Remote sensing and machine learning for crop water stress determination in various crops: a critical review," *Precis. Agric.*, vol. 21, no. 5, pp. 1121–1155, doi: 10.1007/s11119-020-09711-9.