International Journal of Applied Engineering & Technology

TO ASSESS THE EFFECT OF VARIOUS METEOROLOGICAL VARIABLES OF KHARIF & RABI YIELD IN INDIA

[1]Sompal and [2]Dr. Randeep Singh
[1]Research Scholar and [2]Professor, IEC University Baddi
sompal26@gmail.com, randeeppoonia@gmail.com

ABSTRACT
Agricultural productivity and crop yields in India are intricately linked to meteorological variables, making it essential to understand their relationship for effective agricultural resource management. This machine learning study examines the impact of various meteorological variables on Kharif and Rabi crop yields in India. Historical data on crop yields, encompassing the monsoon (Kharif) and winter (Rabi) seasons, is gathered from reliable agricultural databases. Potential predictors temp, humidity, sun radiation, and wind speed analyzed using machine learning to uncover complex interactions and non-linear relationships with crop yields. The research aims to identify critical meteorological conditions that significantly influence crop yields, aiding in informed decision-making for crop planning, resource allocation, and risk management by farmers, and researchers. Furthermore, this study strives to develop accurate predictive models based on meteorological variables, serving as valuable tools for crop production forecasting and assessing climate-related risks. Ultimately, the outcomes of this research can enhance agricultural practices, promote sustainable farming strategies, and bolster food security in India through climate-resilient agricultural systems.

Keywords: meteorological variables, crop yields, machine learning, Kharif, Rabi, India, agricultural productivity, predictive models, climate resilience.

1. INTRODUCTION

The agricultural sector in India is renowned for its remarkable diversity, stemming primarily from variations in resources, climate, geography, and historical and socio-economic factors. Serving as the indian economy relies on agriculture. However, a significant portion of Indian farmers currently faces challenges in achieving the anticipated crop yield. Multiple factors contribute to this situation, with weather conditions such as humidity, and temperature exerting a significant influence on agricultural productivity (**Mulik, Sujata. 2017, Burney et al., 2014**).

India's economy relies on agriculture, its main occupation. However, the success of agriculture is greatly influenced by rainfall. Therefore, the Early rain prediction boosts economic growth (**Poornima et al. 2023**).

Meteorological variables have been of interest since the 17th century, when the first technology was developed to accurately predict weather. Statistical, mathematical, and computer methods, are utilized for this goal, the majority of which are non-linear (**Ayala, M.F., 2017**). Climate conditions are changing due to a variety of factors. For example, as atmospheric pollution increases, climate change occurs and threatens the world (**Erdil et al., 2013**), which is why measuring meteorological variables have become more important since meteorological stations give vital climate change data (**Ruiz-Ayala et al., 2018**).

Over the years, predicting rainfall in advance has proven to be a highly challenging task worldwide (**Liyew et al., 2021, Gupta et al., 2022**). Machine learning types techniques have been employed to tackle this issue. In India, the crop seasons are categorized into two main periods known as "Rabi" and "Kharif." These terms originate from Arabic, with "Rabi" referring to spring and "Kharif" to autumn. The Kharif season takes place from July to October, coinciding with the southwest monsoon, while the Rabi season occurs from October to March (**Thirumalai et al., 2017**).

Each season is associated with specific crops that are harvested. In the Rabi season, crops such as wheat, oats, tomato, potato, peas, barley, linseed, mustard oil, and masoor are cultivated. On the other hand, the Kharif season involves the cultivation of rice, sorghum, peanut, millet, soya bean, bajra, jute, maize, cotton, hemp, tobacco, ragi millet, arhar among others (**Thirumalai et al., 2017**).

In recent years, researchers have recognized the significance of studying the impact of meteorological variables on specific crop seasons in India, particularly the Kharif (monsoon) and Rabi (winter) seasons.

Several studies have attempted to analyze the effect of meteorological variables on Kharif and Rabi crop yields in India. These studies employ various statistical models and methodologies to assess the relationship between weather parameters and crop productivity. The outcomes of these investigations provide crucial insights into the vulnerability of agricultural systems to changing climate patterns, aid in the development of climate-resilient farming strategies, and inform policymakers about the climate change's effects on food security.

2. LITERATURE REVIEW

Segovia et al. (2023) conducted a study on developing a meteorological variable forecasting utilizing machine learning and open-source software. The authors emphasize the significance of accurate meteorological forecasts for various applications such as renewable energy management, agriculture, health, engineering, and energy. They compare several machine learning models including MLP, random forest, decision tree, XGBoost, and many regression methods using evaluation metrics such as RMSE, MAPE, MAE, and R2. The results indicate that Most efficient methods rely on the variable being forecasted, but in general, random forest and XGBoost perform better. For example, random forest performs well in forecasting temperature with R2 0.8631, MAE 0.4728 °C, MAPE 2.73%, and RMSE 0.6621 °C. Similar for relative humidity, solar radiation, and wind speed, random forest is found to be the best performing technique based on the evaluation metrics. Overall, this research contributes to the advancement of meteorological forecasting systems by leveraging machine learning algorithms and open-source software.

Paritosh, et al. (2023) examined climate change risk factors on Indian agricultural production and productivity. The study focuses on India as an emerging economy heavily reliant on food and employment from agriculture. The authors highlight the unique role of India's RBI in supporting the agricultural sector and addressing climate change. Using a machine learning technique known as the sequential adaptive multivariate regression splines, the paper analyzes data from the 2010s, which was recorded as the warmest decade to date. The research finds CO2 emissions, precipitation, the use of irrigation water, and rainfall have the greatest impact on agricultural production. The findings provide valuable insights into the interactions among these factors, particularly the impact of CO2 emissions, on foodgrain and oilseed productivity. Overall, this study contributes a novel and unique modeling approach to understanding how climate change affects Indian agriculture.

Batool et al. (2022) predicted the crop using meteorological, soil, agro management, and crop data from 2016 to 2019 for the study region Shinkiari. MAE, MSE, and RMSE were used to assess the performance of the models after the author's added machine learning algorithms. At 0.15 t/ha RMSE, the XGBoost regressor outperformed all other applied models.

Hasegawa et al., (2022) explained the study of a global dataset of expected climate change consequences on four key crops. In this study, we create a worldwide dataset by combining information from existing meta-analyses and a newly conducted research study on current crop simulations. 8503 simulations from 202 1984–2020 studies compose the new global dataset. Several 21st-century emission scenarios with and without adaption measures, it describes the locations, present and forecast temperature and precipitation variations, and estimated yields of four major crops in 91 countries (maize, rice, soybeans, and wheat). This kind of data collection is essential for the fast advancement of data-driven machine learning systems and lays the groundwork quantitative analysis of climate change's effects on agriculture.

Pant et al., (2021) analysed that agriculture is a very important sector of India's economy. Because of the dramatic rise in the human population, crop productivity is the most important factor in determining whether or not there will be enough food. Predicting crop yields is one of the most important challenges facing the agricultural industry. The amount of produce that may be harvested from agricultural land is determined by a number of elements, including the current state of the climate (including rainfall, humidity, and temperature), as well as any relevant data on pesticides. Apart from these elements, having precise knowledge about the history of the crop production is a key notion for both generating forecasts and minimizing agricultural risk. In the past, predictions of yield were made by consider the farmer's prior experience with the specific field and crop. This study uses machine learning to estimate four common crop yields. that are widely grown in India. When a site-specific prediction of the crop yield has been made, the inputs such as fertilizers that are used may be applied in a variable manner according to the anticipated requirements of the crop and the soil. The use of machine learning is considered in this research in order to make a forecast about the four yields that are most often farmed in India. These crops are wheat, potatoes, maize, and paddy rice, respectively.

Pandhe et al. (2019) examined the uncertain climate and its consequences on crop yields, particularly in India, where decreasing yields of conventional crops have led to an alarming surge in farmer suicides. Predicting crop yields helps farmers choose crops based on local climate, according to the authors. This knowledge also aids policymakers in import-export, pricing, and marketing decisions. Due to climate neglect, crop yield predictions have been erratic. The authors suggest using machine learning, specifically the Random Forest technique, to build a crop yield prediction model. The study excludes soil quality, pests, and pesticide usage, which may differ between fields, and focuses on five meteorological characteristics to train the model. This research led to Smart Farm, an app that predicts agricultural yields based on climate.

Chandra et al. (2019) developed machine learning-based models for predicting kharif rice yield in the upland rainfed area of West Bengal counties Purulia and Bankura. It utilized ANN and Random Forest algorithms and integrated multi-temporal vegetation indices, weather variables, and block-level non-weather variables from 2006 to 2015. The developed models showed a correlation of 0.702 and mean squared error of 0.01, indicating satisfactory performance when weather variables were compared with vegetation indices (NDVI). However, when NDVI was correlated with kharif rice yield, the models exhibited a relatively lower correlation of approximately 0.6. This finding suggested the need for additional farmer-controlled inputs to enhance the accuracy of the models. The authors suggested that developing models specific to Different crop growth stages or a larger dataset with selectable weather and non-weather variables and NDVI would be better for improving the prediction of kharif rice yield.

Table 1: Comparison of reviewed technique

Author's & years	Methods	Focus	Key finding
Segovia et al. (2023)	Machine learning and open-source software	Meteorological variables forecasting system	Random forest and XGBoost perform better in forecasting meteorological variables.
Paritosh et al. (2023)	Sequential multivariate adaptive regression splines	Impact of global warming on India's agricultural output	CO2 emissions, Precipitation, irrigation water utilization, and rainfall dominate Indian agricultural productivity.
Batool et al. (2022)	XGBoost regressor	Crop prediction for the study region Shinkiari	XGBoost regressor outperformed other models in predicting crop yield with an RMSE of 0.15 t/ha.
Hasegawa et al. (2022)	Meta-analyses and crop simulations	Global climate change projections for main crops	Impact of global warming on India's agricultural output on maize, rice, soybeans, and wheat yields in different nations under various emission

			scenarios and with/without adaptation measures.
Pant et al. (2021)	Models of machine learning	Crop yield forecast for four common crops in India	Using machine learning models, one can predict the yields of wheat, potatoes, maize, and paddy rice based on various factors including climate, pesticide usage, and historical crop production.
Pandhe et al. (2019)	Random Forest technique	Predicting crop yields in India	Random Forest technique is used to predict crop yields based on meteorological characteristics, leading to the development of an app called Smart Farm.
Chandra et al. (2019)	Artificial Neural Network and Random Forest	Predicting kharif rice yield in Purulia and Bankura districts	Artificial Neural Network and Random Forest models show satisfactory performance in predicting kharif rice yield, but additional farmer-controlled inputs are needed to enhance accuracy.

3. RESEARCH METHODOLOGY

This study examines meteorological variables' relationships and Kharif and Rabi crop yields in India, and to develop accurate predictive models for crop production forecasting and risk assessment. Gather historical data on crop yields for the Kharif and Rabi seasons from reliable agricultural databases. Ensure the data spans enough time to capture fluctuations. in meteorological conditions and crop yields. Gather historical data on crop yields for the Kharif and Rabi seasons from reliable agricultural databases. Ensure that the data covers a sufficiently long time period to capture variations in meteorological conditions and crop yields. In this we use various machine learning techniques such as Random forest, KNN, MLR to assess the effect of various meteorological variables of kharif & Rabi yield in India.

In this section, we will delve into the employed prediction techniques and their design for studying and forecasting various meteorological variables. Specifically, this research focuses on utilizing machine learning methods to analyze temperature, wind speed, sun radiation, and humidity.

Random Forest: The random forest is a regression model that enhances predictive accuracy by dividing the dataset into subsets and applying the decision tree algorithm to each subset. Finally, the average of the individual predictions is taken into consideration.

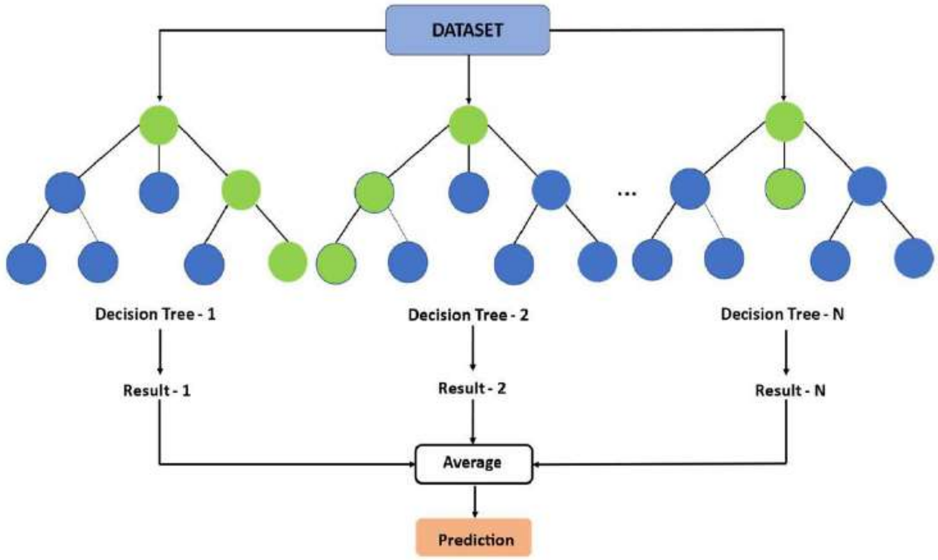

Fig 1: Random forest model

Neural Network KNN is a Versatile Classification Algorithm. and regression tasks, characterized by its non-parametric nature. During training, it creates a "lookup table" of labeled instances. To predict on new instances, KNN compares them to the training data using a distance metric, identifying the K nearest neighbors. For classification, majority voting determines the projected class, while regression uses the mean or median of K neighbors. Choosing the value of K is crucial, with smaller values emphasizing local variations and larger values smoothing predictions. KNN is powerful but computationally intensive for large datasets, often requiring preprocessing steps like feature scaling and handling imbalanced data.

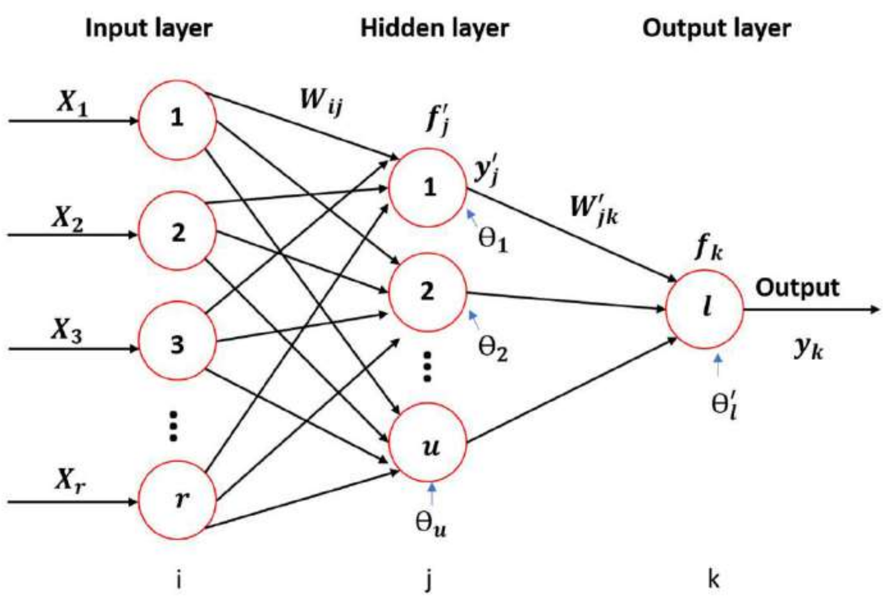

Fig 2: Structure of a multilayer perceptron neural network

MLR: Multiple linear regression models complex relationships between independent variables. and a single dependent variable. This technique proves beneficial when dealing with scenarios involving two or more x variables.

Models Evaluation

Regression models are evaluated using MAE, RMSE, and R-squared **(Van Klompenburg et al. 2020)**. MAE measured the average absolute difference between original and forecasted values across the dataset, providing a measure of the accuracy of the model's predictions.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

RMSE, which stands for Root Mean Square Error, quantifies the prediction errors in a model. It assesses residuals capture data point deviance from regression line and residual distribution. In essence, RMSE provides an indication of the proximity of the data to the line that represents the optimal fit.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

The determination coefficient, also known as R-squared, quantifies the extent to which values align with the original data. From 0 to 1, it's a percentage. The model fits better with a larger R-squared.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

4. RESULT AND DISCUSSION

4.1 Temperature Forecasting

The results of the evaluation metrics for temperature forecasting techniques are presented in Table 2 shows RMSE, MAE, and R2.

To determine root mean square, mean absolute percentage, and mean absolute error, the errors of the validation data (576 data points) were averaged. On the other hand, the computation of the coefficient of determination (R^2) utilized both the training and test sets, which consisted of a total of 93,780 data points.

Table 2: Temperature Forecasting

Technique	R²	MAE	RMSE
MLR	0.8244	0.6597	0.8453
Random forest	0.8631	0.4728	06621
Multilayer perceptron NN	0.8226	0.9124	1.2498

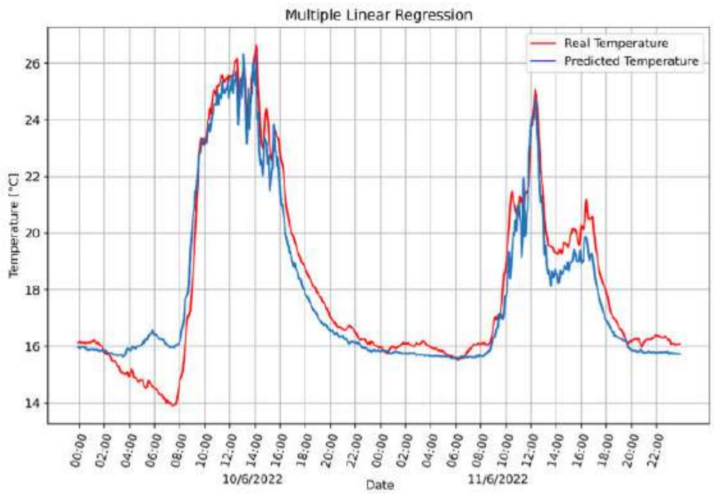

Fig 3: Forecasting temperature with multiple linear regression

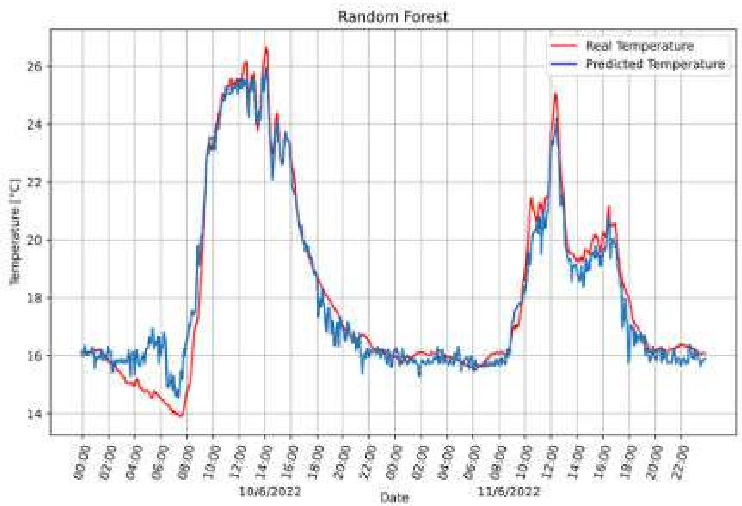

Fig 4: Temperature forecast techniques using Random forest

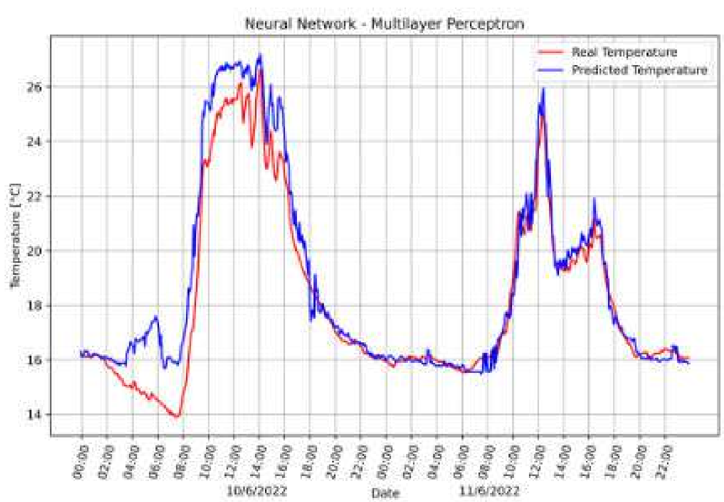

Fig 5: Temperature forecast techniques using Neural Network

4.2 Relative Humidity Forecasting

Table 3 demonstrates the convergence of R^2 values obtained from the implemented algorithms, indicating an accurate relative humidity prediction vs. reality. This convergence ensures the algorithm's effective performance and facilitates a comparison of forecast errors. Analyzing RMSE, MAPE, and MAE, and R2 is the coefficient of determination for each technique. employed, Random Forest predicts relative humidity best. Random Forest has 0.8583 R2, 2.1380 MAE, and 2.9003 RMSE.

Table 3: Humidity Forecast

Technique	R^2	MAE	RMSE
MLR	0.7815	3.0900	3.7475
Random forest	0.8583	2.1380	2.9003
Multilayer perceptron NN	0.8013	4.6055	5.5759

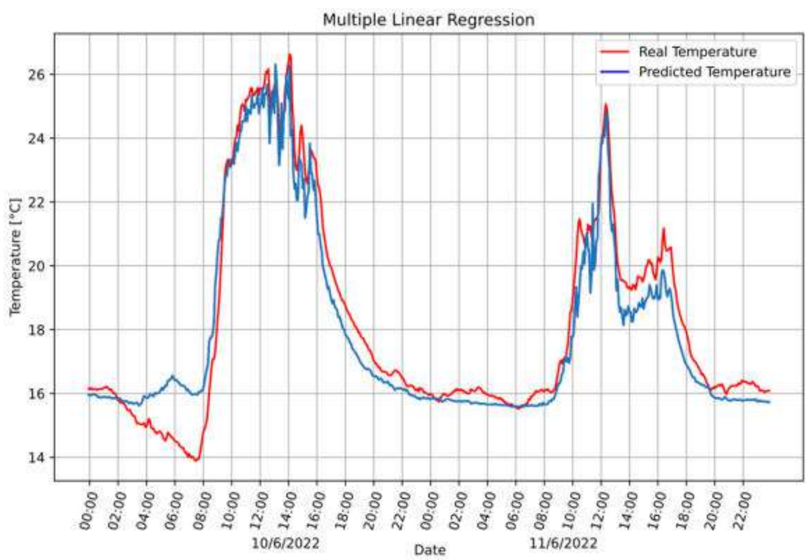

Fig 6: Humidity Forecasting using multiple linear regression

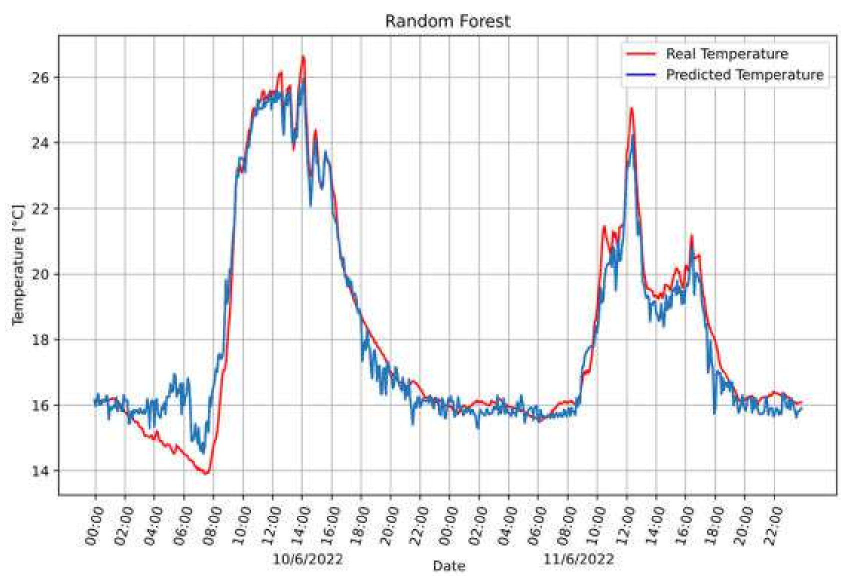

Fig 7: Humidity Forecasting using Random Forest

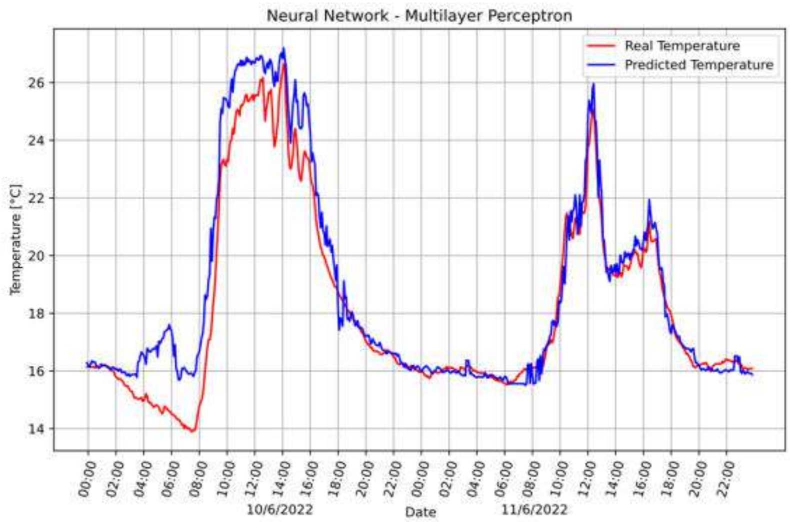

Fig 8: Humidity Forecasting using Neural Network

4.3 Solar Radiation Forecasting

In Table 4, the R^2 values obtained from the implemented algorithms demonstrate convergence towards appropriate values. This indicates a close approximation between solar radiation prediction and reality, ensuring the algorithm's strong performance. This convergence allows for a comparison of forecast errors, RMSE, MAE, and R2 are included. The analysis of the different implemented techniques reveals that the Random Forest method performs the best in forecasting the solar radiation variable. It has an R2 of 0.7333, an MAE of 65.8105 W/m2, and an RMSE of 105.9141 W/m2.

Table 4: Solar Radiation

Technique	R^2	MAE	RMSE
MLR	0.6689	106.9741	164.7435
Random forest	0.7333	65.8105	105.9141
Multilayer perceptron NN	0.7423	88.5897	140.0681

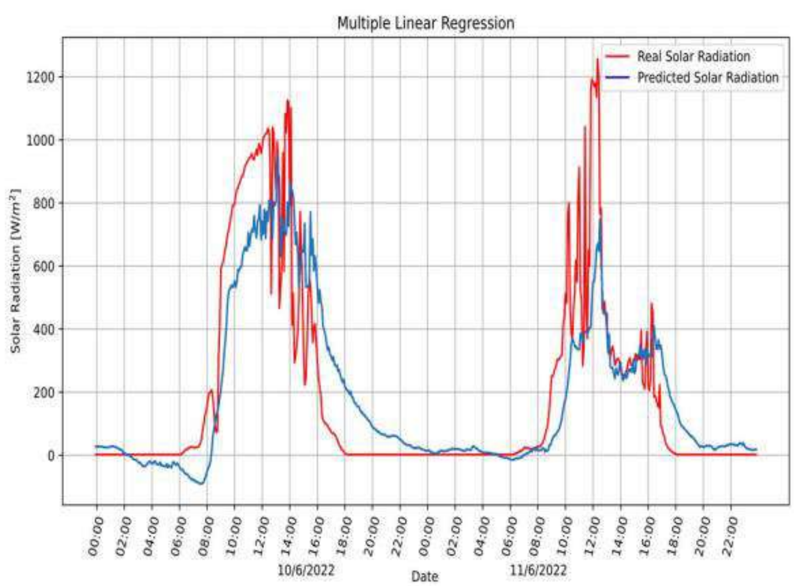

Fig 9: Solar Radiation using multiple linear regression

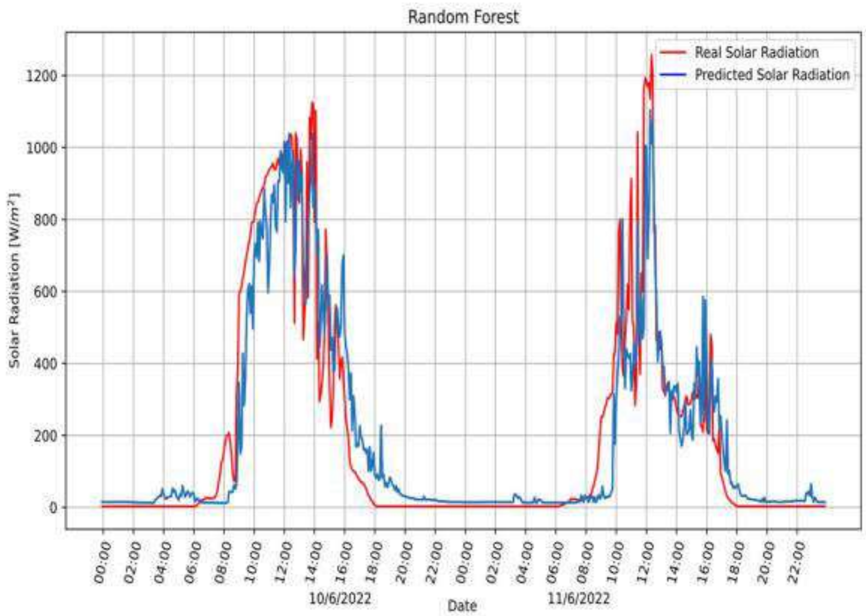

Fig 10: Solar Radiation using Random Forest

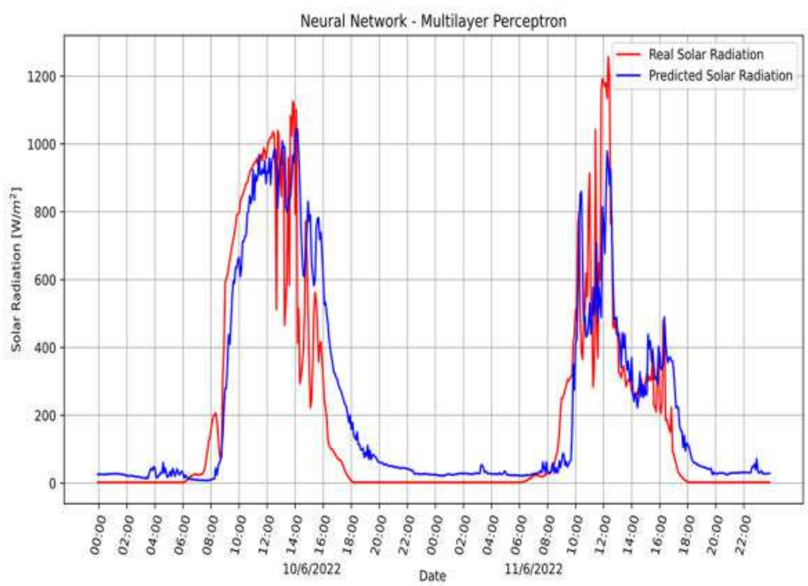

Fig 11: Solar Radiation using Neural Network

4.4 Wind Speed Forecasting

Evaluation metrics for wind speed forecasting techniques are presented in Table 5. Metrics include RMSE, MAE, and R2. The R2 calculation used both the training and test sets (93,780 data points), whereas the RMSE and MAE averaged the validation data errors (576 data points). Table 5 demonstrates that R2 values derived from implemented algorithms converge to acceptable levels. This indicates a reasonable approximation between the predicted and actual wind speed, ensuring the algorithm's effectiveness. Consequently, a comparison of forecast errors can be made. Upon comparing the RMSE, MAE, and R2 of the different implemented techniques, it is evident that the Random Forest technique performs the best in forecasting the wind speed variable. R2, MAE, and RMSE are respectively 0.3660, 0.1097, and 0.2136 m/s. It is important to note that the coefficient of determination for wind speed is relatively lower than for other variables due to its high variability. Nonetheless, the implemented techniques yield forecasts with acceptable errors.

Table 5: Wind Speed

Technique	R^2	MAE	RMSE
MLR	0.3428	0.1614	0.3354
Random forest	0.3660	0.1097	0.2136
Multilayer perceptron NN	0.3270	0.1654	0.3616

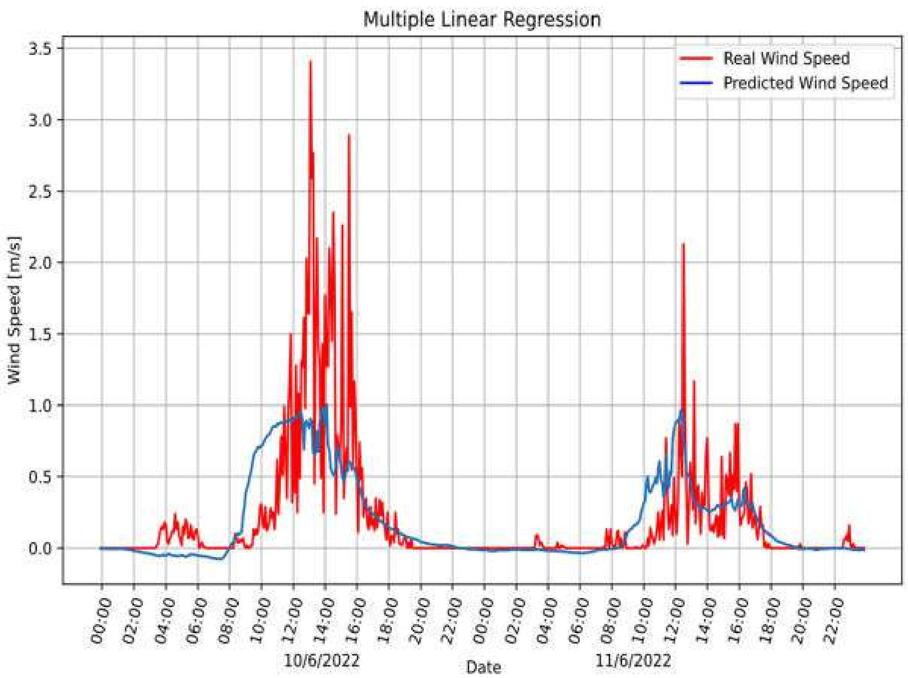

Fig 12: Wind Speed using multiple linear regression

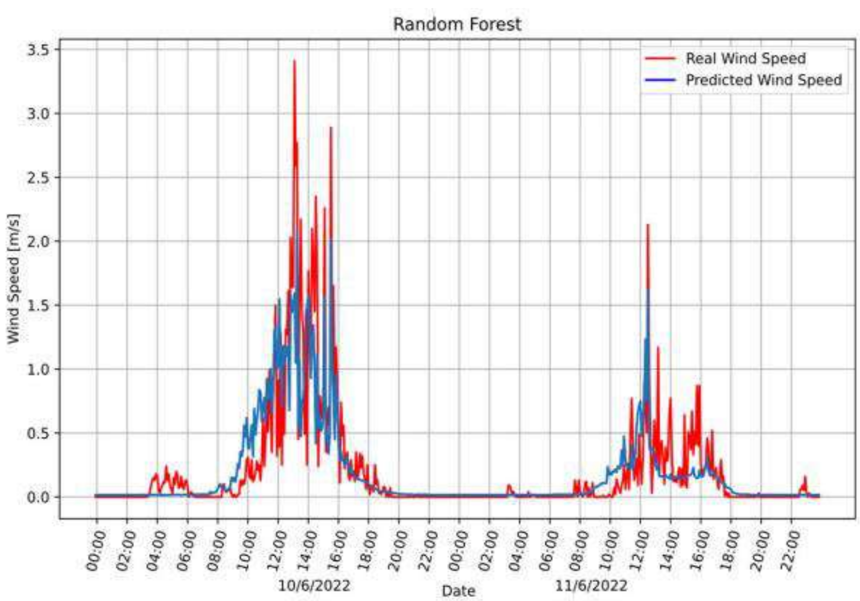

Fig 13: Wind Speed using Random Forest

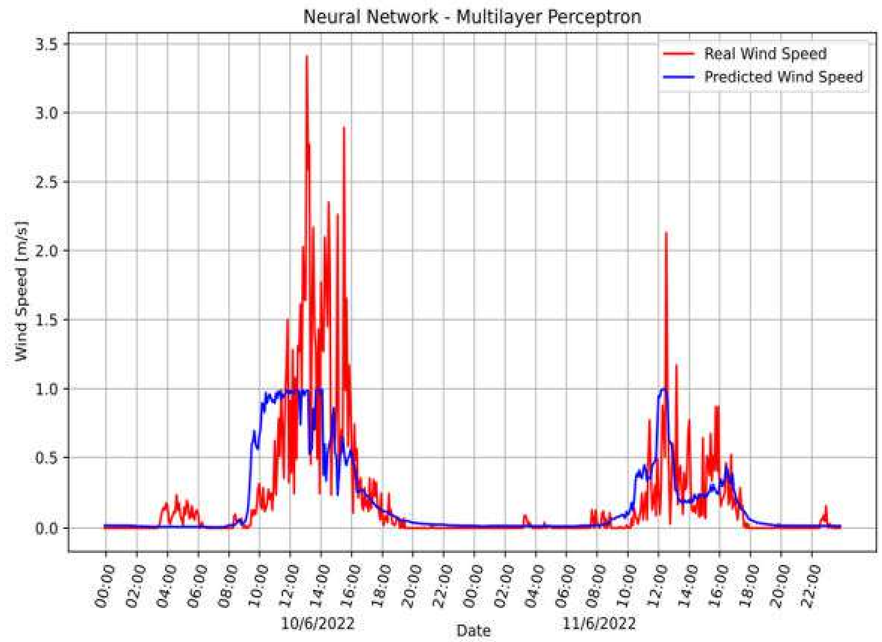

Fig 14: Wind Speed using Neural Network

CONCLUSION

Machine learning was used to assess the impact of different meteorological variables on Kharif and Rabi crop yields in India, aiming to identify critical weather conditions that significantly influence yields, develop accurate predictive models, and improve climate-resilient farming practices. Probabilistic models like random forests and neural networks, and multiple linear regression, were employed to forecast temperature, humidity, solar radiation, and wind speed. Model accuracy was assessed using R2, MAE, and RMSE. Random forest emerged as the most effective model for temperature and humidity forecasting, displaying high R2 values of 0.8631 and 0.8583, respectively. It also performed well in solar radiation forecasting with an R2 value of 0.7333. Although wind speed forecasting had higher variability, random forest still achieved an R2 value of 0.3660. These results show that machine learning, especially random forest, can properly predict meteorological variables and their influence on crop yields. The developed predictive models can serve as valuable tools for crop production forecasting, risk assessment, and climate resilience planning in the agricultural sector. By comprehending the relationships between weather variables and crop yields, Farmers and policymakers can make knowledgeable decisions concerning crop planning, resource allocation, and climate-related risks. This research contributes to advancing agricultural resource management and promoting sustainable farming strategies in India, harnessing the power of machine learning and meteorological data analysis to enhance agricultural productivity, food security, and alleviating climate change's effects on agriculture.

REFERENCES

1. Mulik, Sujata. (2017). Analysis of Crop Yield Prediction of Kharif & Rabi Jowar Crops Using Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 7. 79. 10.23956/ijarcsse.v7i11.468.

2. Burney, Jennifer, and V. Ramanathan. "Recent climate and air pollution impacts on Indian agriculture." *Proceedings of the National Academy of Sciences* 111.46 (2014): 16319-16324.

3. Poornima, S., et al. "Rainfall Forecast and Drought Analysis for Recent and Forthcoming Years in India." *Water* 15.3 (2023): 592.

4. Liyew, Chalachew Muluken, and Haileyesus Amsaya Melese. "Machine learning techniques to predict daily rainfall amount." *Journal of Big Data* 8 (2021): 1-11.

5. A. Gupta, H. K. Mall and S. Janarthanan., "Rainfall Prediction Using Machine Learning," *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, Hyderabad, India, 2022, pp. 1-5, doi: 10.1109/ICAITPR51569.2022.9844203.

6. Thirumalai, Chandrasegar, et al. "Heuristic prediction of rainfall using machine learning techniques." *2017 International Conference on Trends in Electronics and Informatics (ICEI)*. IEEE, 2017.

7. Segovia, Jenny Aracely, et al. "Meteorological Variables Forecasting System Using Machine Learning and Open-Source Software." *Electronics* 12.4 (2023): 1007.

8. Pandhe, Abhijeet, et al. "Crop yield prediction based on climatic parameters." *METHODOLOGY* 6.03 (2019).

9. D. Batool et al., ―A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning,‖ Plants, vol. 11, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/plants11151925.

10. Pant, Janmejay, R. P. Pant, Manoj Kumar Singh, Devesh Pratap Singh, and Himanshu Pant. "Analysis of agricultural crop yield prediction using statistical techniques of machine learning." *Materials Today: Proceedings* 46 (2021): 10922-10926.

11. Chandra, Aditi, et al. "Machine learning approach for kharif rice yield prediction integrating multi-temporal vegetation indices and weather and non-weather variables." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (2019): 187-194.

12. Hasegawa, Toshihiro, Hitomi Wakatsuki, Hui Ju, Shalika Vyas, Gerald C. Nelson, Aidan Farrell, Delphine Deryng, Francisco Meza, and David Makowski. "A global dataset for the projected impacts of climate change on four major crops." *Scientific data* 9, no. 1 (2022): 58.

13. Jha, Paritosh, et al. "A machine learning approach to assess implications of Climate Risk Factors on Agriculture: The Indian case." *Climate Risk Management* (2023): 100523.

14. Ayala, M.F. *Analisis de la Dinamica Caoticapara la Series Temporales de Variables Meteorologicas en la Estacion Climatica de Chone*; Universidad de las Fuerzas Armadas ESPE: Sangolquí, Ecuador, 2017; Available online: http://repositorio.espe.edu.ec/handle/21000/13629 (accessed on 24 November 2022).

15. Erdil, Ahmet, and Erol Arcaklioglu. "The prediction of meteorological variables using artificial neural network." *Neural Computing and Applications* 22 (2013): 1677-1683.

16. Ruiz-Ayala, Daniel Camilo, Carlos Arturo Vides-Herrera, and Aldo Pardo-García. "Monitoreo de variables meteorológicas a través de un sistema inalámbrico de adquisición de datos." *Revista de investigación, Desarrollo e Innovación* 8.2 (2018): 333-341.

17. Van Klompenburg, Thomas, Ayalew Kassahun, and Cagatay Catal. "Crop yield prediction using machine learning: A systematic literature review." *Computers and Electronics in Agriculture* 177 (2020): 105709.