IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# Machine Learning Models in Predicting Rabi and Kharif Crop Suitability in Rajasthan's Arid Environment: A Maxent and SDM Approach"

[1] Neelam Bohra, Research Scholar Department of Computer Science Engineering M.B.M University, Jodhpur, Rajasthan, neelam.bohra@gmail.com
[2] Dr. Anil Gupta, Professor Department of Computer Science and Engineering, M.B.M. University, Jodhpur, Rajasthan, anil.cse@mbm.ac.in

## Abstract

This study's main objective is to determine the impact of bioclimatic variables on rabbi and Kharif crops in western Rajasthan using the species distribution machine learning model MaxEnt. Species distribution models, often called SDMs, are valuable tools used to understand and predict where species are likely to live based on their relationships with the environment. The main challenge of using machine learning models for crop prediction in arid environments is the variability and complication of environmental data. It has been found that precipitation Seasonality(BIO15) contributes most to predicting the Isabgol crop, and the Mean Temperature of the wettest quarter (BIO 08) contributes most to the Mung crop. This work concludes that the AUC of mung crops (AUC=0.92) is better than the Isabgol crop.

**Keywords: SDM, MaxEnt, Bioclimatic variables, rabbi crop, Kharif crop.**

## 1. Introduction:

It has been observed that due to variations in climate, the quality of crop yield has declined all over the world, so now it is required that farmers use the new advanced tools for making agriculture-related decisions. The latest big data technology is directed to many intense analytical tools, and machine learning is one of them. To predict the annual yield of any crop, many components must be combined, like climate data, weather data, and agriculture yield, which help farmers make the correct decision. It has been seen that machine learning tools contribute a lot in making agriculture-related decisions, such as what crop to grow, and they also suggest habitat suitability for particular crops while considering various environmental variables [1].

Machine learning models have become increasingly crucial in predicting crop suitability and yield, particularly in challenging environments like Rajasthan's arid regions. The use of Species Distribution Models (SDM) offers promising approaches for assessing the suitability of Rabi and Kharif crops in arid areas.

This study's main objective is to determine the impact of bioclimatic variables on rabbi and Kharif crops in western Rajasthan using the species distribution machine learning model MaxEnt. This research paper is divided into five sections: section I highlights the characteristics of the arid zone and challenges of the arid region, defines SDM and Maxent, And Section II discusses related work done by another researcher. Section III elaborates on the adopted methodology. Section IV contains the experiment results, and section V concludes the outcomes of the obtained results.

### 1.1 Characteristics of Arid Zone

Arid regions are characterized by low annual rainfall, typically below 450mm, while evapotranspiration (ET) rates are significantly higher—often four to five times greater than precipitation levels. A primary challenge in these areas is the lack of water balance, leading to persistent water scarcity, particularly for drinking purposes. Additionally, the landscape in arid zones is highly fragmented in terms of land, water, and vegetation, making these regions prone to permanent land degradation and desertification.

Due to inadequate planning, agricultural productivity remains uncertain. In years with sufficient rainfall, a single crop may be cultivated, but historical patterns indicate that, over a five-year period, only one year is likely to result in a good harvest. Two years may witness moderate yields, while crop failure is common in the remaining two years. Furthermore, water harvesting and recycling opportunities are extremely limited.

In India, the hot arid zone spans approximately 31.7 million hectares. The largest portion, around 61%, is located in western Rajasthan, while Gujarat accounts for 20%. Additionally, Punjab and Haryana share 9% of the arid region, with the remaining 10% spread across Andhra Pradesh and Karnataka. [2].

### 1.2 Challenges in Western Rajasthan Arid Zones

Western Rajasthan comprises the largest portion of India's arid zone. Rainfall in this region is limited, typically ranging between 185 mm and 472 mm annually, with a short duration of precipitation. Harsh climatic conditions, including intense solar radiation, strong winds, and high temperatures, contribute to excessive evapotranspiration, leading to a persistent negative water balance.

The soil in this area is predominantly light sandy, with low organic matter content (0.03–0.30%). It exhibits an alkaline nature (pH 8.2–8.8), poor fertility, and limited water retention capacity (70–100 mm/m). Additionally, the soil profile shows weak to moderate development, is prone to crust formation, and is highly susceptible to wind erosion. Water erosion is also a notable concern in certain parts of the region.

Despite being naturally endowed with valuable vegetation, the arid zone faces rapid depletion of plant species due to overexploitation. Nutritious fodder species are gradually being replaced by less edible plants, further exacerbating ecological imbalances. The most pressing challenge in this region remains the sustainable management of land and rainwater resources.

[2].

### 1.3 SDM and MaxEnt Algorithm

This section defines the Species distribution models, often called SDMs, as valuable tools used to understand and predict where species are likely to live based on their relationships with the environment. There are two main ways to approach this. The first is a more detailed method, focusing on the biological processes—like how a species grows, reproduces, or adapts—determining its natural range. The second and far more common approach uses mathematical models to explore patterns between where species are observed and the environmental factors that influence their presence.

This work centers on the second method, offering insights into important ideas and resources within this broad and widely-used field. These models have applications that stretch across various disciplines, including ecology, conservation, biogeography, evolutionary biology, public health, biosecurity, and even computational science. The focus is mainly on the predictions these models can generate, which are often visualized through maps. While they're primarily referred to as species distribution models, alternatively, they might be called ecological niche models, habitat models, or bioclimatic envelope models [3].

When designing a species distribution model, it's essential to focus on three core elements: species data, environmental data, and the algorithm. Various algorithms are available for species distribution modeling, such as geographic models, profile models, statistical regression models, and machine learning models. However, these categories can overlap, as many machine learning techniques build on regression methods found in statistical models. Profile models are considered the simplest form of species distribution models. Like geographic models, they only use occurrence data, but they incorporate environmental data as well. Statistical regression models, however, require data's presence and absence to function. Machine learning models use ecological data, too, but they encompass a wide range of methods. Most of these methods rely on presence and absence data, except the widely used Maxent algorithm, which combines presence data with background data [4, 5].

### 1.3.1 MaxEnt

MaxEnt, also known as maximum entropy, is a popular algorithm for species distribution modeling, and it belongs to the realm of machine learning, meaning it creates multiple models through an iterative process. At its core, Maxent relies on two main ideas:

a. Entropy: The algorithm aims to spread the distribution across the study area, ensuring a balanced prediction.

b. Constraints: These rules shape the predicted distribution based on the environmental features found at the places where the species has been observed.

How it works: Maxent calculates two types of probability densities. One describes the environmental conditions where the species has been recorded (presence points), and the other represents the environmental conditions across the entire study region (background points). By comparing these two densities, the algorithm determines how suitable each location is for the species' presence.

The primary goal of the Maxent algorithm is to identify a distribution that closely matches the overall environmental conditions of the study region with those found at the species' locations. This initial result, the "raw output," is then transformed into a more user-friendly format called the logistic output. This version estimates the likelihood of a species being present at any given spot, considering how common or rare the species is. However, Maxent assumes a default prevalence value of 0.5, meaning the species is thought to occupy half of all possible locations. This may not always be accurate, especially for rare species, so it's essential to approach this default value with caution. An important feature of Maxent is regularization, which helps prevent the model from overfitting [6, 7].

## 2. Related Work

This section discusses the work done by various researchers using the MaxEnt algorithm and environmental factors to find habitats suitable for multiple species. In 2019, Benjamin Kipkemboi Kogo and their team used the MaxENT model to assess climate suitability for rainfed maize cultivation in Kenya under different climate scenarios. It identifies key bioclimatic factors influencing maize cultivation suitability, evaluates model performance with high accuracy, and maps habitat suitability zones for maize production in Kenya [8]. Saurabh Purohit & Neelam Rawat, in 2021, by using MaxEnt modeling, predicts the current and future distribution of Clerodendrum infortunatum L. under climate change scenarios in Dehradun district, India. The distribution of species habitats was influenced by environmental factors, with bioclimatic variables playing a key role in shaping these environments [9].In 2023, Shahzad Ali and others found that Bioclimatic variables Bio4, Bio12, Bio10, and Bio14 contribute significantly to the MaxEnt model, which accurately predicts changes in land suitability for wheat, soybean, and rice cultivation under climatic change scenarios in East Asia [10]. Ali and team 2024 utilize the MaxEnt model to analyze the impact of bioclimatic variables on the current and future potential land distribution dynamics of wheat, rice, and maize in South Asia under climate change scenarios by examining the contribution of specific bioclimatic variables like temperature and precipitation to the distribution of these crops [11]. Work done by various researchers concludes that the MaxEnt algorithm is widely used to assess habitat suitability for crops by analyzing bioclimatic variables. This approach helps predict how climate change might alter the geographic suitability of lands for various crops.

## 3. Methodology:

This section elaborates on the complete methodology used for experimenting with two crops, i.e., Kharif (rainy season) and Mung bean crop (Vigna radiata) also known as green gram, hold significant economic importance in Rajasthan due to their adaptability to arid conditions and their role in sustainable agriculture Mung beans are a considerable pulse crop in Rajasthan, especially in districts like Nagaur, Jodhpur, and Pali, which contribute over 50% of the state's mung bean cultivation. Their cultivation supports the livelihoods of farmers in these regions [12]. Another rabi crop chosen for this study is Isabgol (Plantago ovata), which plays a significant role in Rajasthan's economy, particularly in the arid and semi-arid regions like Barmer, Jodhpur, and Nagaur. Rajasthan contributes significantly to India's isabgol exports, which are in high demand globally for medicinal and dietary uses. Studies have shown that isabgol cultivation offers a favorable benefit-cost ratio, making it a profitable venture for farmers in Rajasthan. Its ability to thrive in Rajasthan's challenging climatic conditions makes it a reliable crop for farmers [13, 14].

### 3.1 Data Source

The species occurrence records of Mung and Isabgol were collected from the district statistics department of Jodhpur. Obtained locations were converted into geocoded data. For the mung crop, there are 1527 rows, and for Isabgol, 541 rows were obtained. The 19 raster layers of bioclimatic variables were available on the WorldClim website (worldclim.org) and used for an experiment. For this study19, bioclimatic variable raster layers were downloaded from the Worldclim website. This website provides data for four spatial resolutions between 30 seconds (~1 km2) and 10 minutes (~340 km2) and historical and future climate data. In this work, 19 bioclimatic variables for the years 1970-2000 and 30 arc-second resolutions were downloaded.

### 3.2 Data Pre-processing

Machine learning is a data-driven technique, so examining the data before using it to develop the target model is necessary. Pre-processing steps include several data cleaning techniques, and with the help of these techniques, those rows in the dataset with duplicate values, missing values, null values, and rows with special symbols must be removed. After data pre-processing, the resultant dataset obtained was in the desired format, which can be used for computational processing.

| **Species** | **Available record rows** | **After pre-processing rows** |
|---|---|---|
| Mung | 1527 | 450 |
| Isabgol | 541 | 206 |

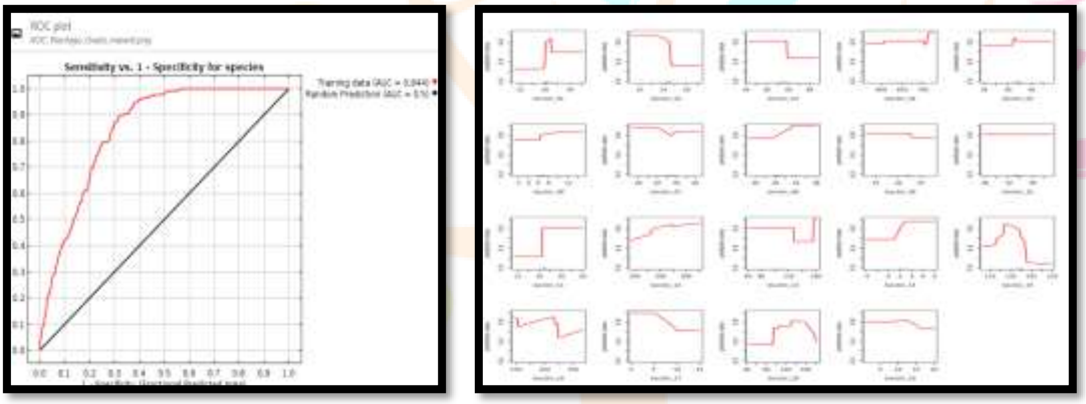
Table occurrence records of Mungh and Isabgol crops

Table 1 shows the occurrence data rows for crops before and after pre-processing. After this, computational data was obtained, and with 19 bioclimatic variables as input fed into the model, various outputs can be obtained, which helps to find out the habitat suitability of both crops considering the arid region climate condition.

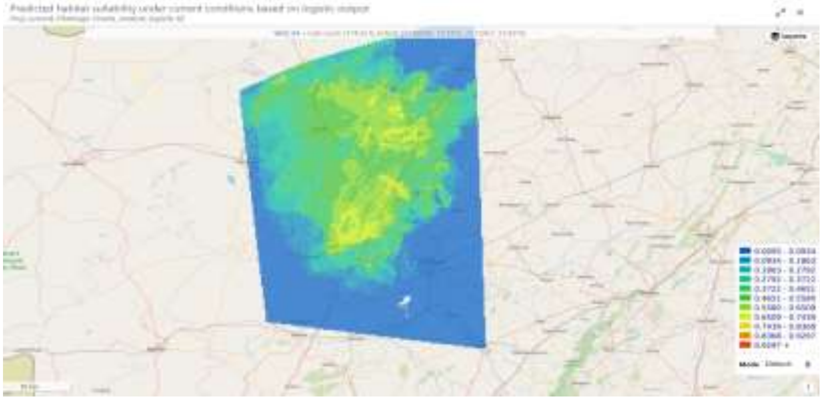
## 4. Results

In this section, results obtained for both crops are discussed. The experiment was conducted for both crops, considering current climate conditions for the arid regions. As mentioned, for conducting species distribution modeling, two inputs required for this19 bioclimatic variables and geocoded occurrence data were input in the MaxEnt algorithm. In this work, separate experiments were performed for individual crops. The main objective was to determine each bioclimatic variable's contribution and find the most suitable habitat for both crops, specifically in Rajasthan's arid region. The ROC plot is also used as a performance measure of various models; it is a graph with the False Positive Rate on the x-axis and the True Positive Rate on the y-axis plotted across the range of possible thresholds. The value for ROC is the area under the curve (AUC) and is calculated by summing the area under the ROC curve. A value of 0.5 thus represents a random prediction, and values above 0.5 indicate predictions better than random. The closer the ROC curve follows the y-axis, the larger the area under the curve and, thus, the more accurate the model. Generally, AUC values of 0.5–0.7 are considered low and represent poor model performance, values between 0.7 and 0.9 are considered moderate, and values above 0.9 represent excellent model performance.

**4.1 Result for Isabgol (Plantago ovata) Crop**: For the Isabgol crop, the total training sample was 206, and the number of iterations was 500. The graph shows the contribution of each bioclimatic variable. Figure (a) shows the ROC plot, and it shows that the model has an AUC of 0.84; figure (b) shows the graph showing the contribution of each bioclimatic variable, and these graphs show that precipitation Seasonality(BIO 15) contributes most means that variation in precipitation throughout the year consider as important factor for determining habitat suitability of Isabgol crop or it can be concluded that hence Isabgol growth depend on distribution of rainfall across seasons. The model also revealed that precipitation of the Driest Quarter (BIO 17)had less contribution. It shows that the extent of rainfall during the driest quarter has minimal effect on crop distribution. Figure (c) shows the predicted habitat suitability under conditions based on logistic output, and it shows that Bheekam Kor has higher predicted habitat suitability, and Jodhpur and Baroi have the lowest. Cherai has medium habitat suitability for the Isabgol crop under the current climate conditions.

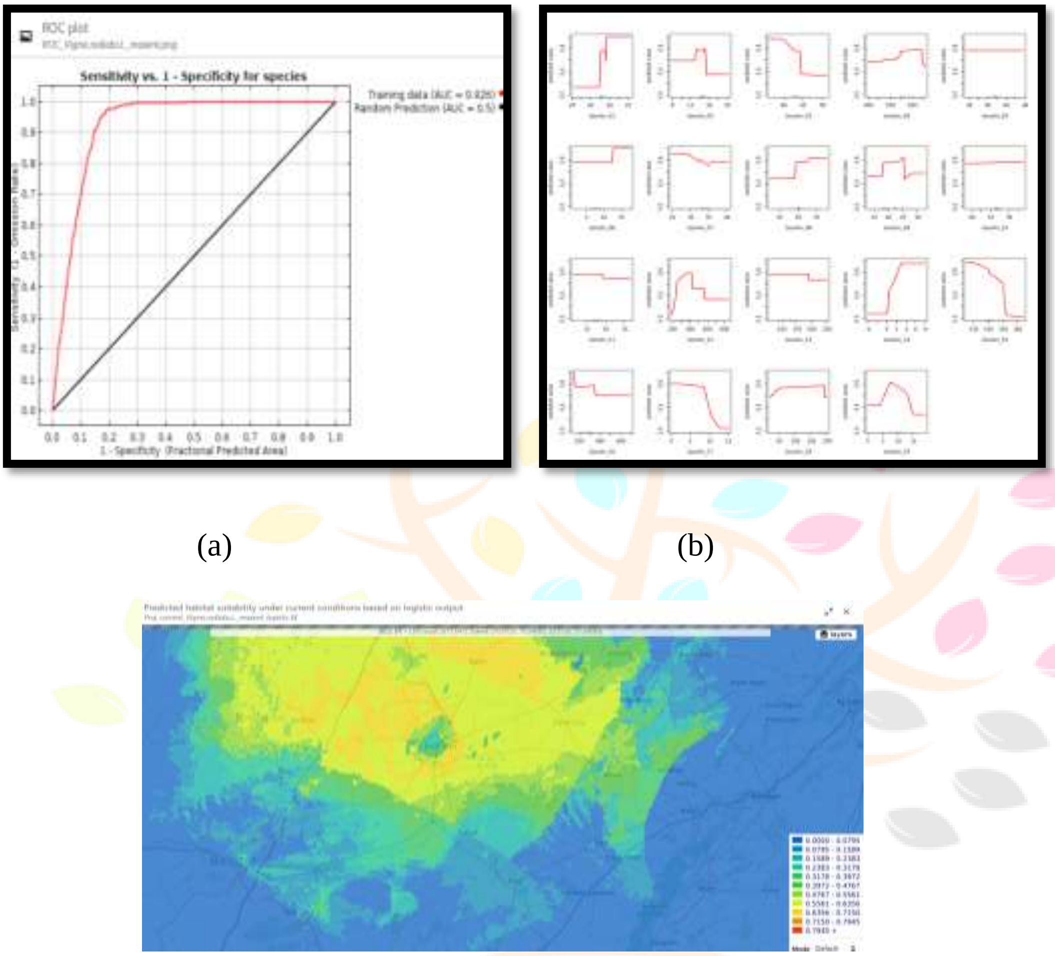


(a) (b)



(c)Figure1 Results for Isabgol crop (a) ROC Plot (b) contribution of environmental variables(c) predicted habitat suitability under conditions based on logistic output

**4.2 Result for Mungh (Vigna radiata) crop:** For the Mungh crop, the total training sample was 445, and the number of iterations was 500. The graph shows the contribution of each bioclimatic variable. Figure (a) shows the ROC plot, and it shows that the model has an AUC of 0.92; figure (b) shows the graph showing the contribution of each bioclimatic variable and from these graphs shows that the Mean Temperature of the Wettest Quarter(BIO 08) contributes most means that those regions with optimal temperature during the wettest quarter are most suitable for mung bean farming. Figure (c) shows the predicted habitat suitability under conditions based on logistic output, and it shows Baroi and Balesar have maximum probability suitability, piper, Tiwari, and sheath have medium suitability, and Jodhpur have the lowest habitat suitability



(a) (b)

(c)

Figure Results for Mung crop (a) ROC Plot (b) contribution of environmental variables

(c) predicted habitat suitability under conditions based on logistic output

## 5. Challenges and Future Directions

The main challenge of using machine learning models for crop prediction in arid environments is the variability and complication of environmental data. Model accuracy can be increased by continuously updating it with new data sets. Future research should focus on integrating more varied datasets and refining models to enhance prediction accuracy [15, 16].

## Conclusion

This work concludes that the AUC of the mung crop(AUC=0.92) is better than that of the Isabgol crop(AUC=0.84), which means the Maxent model works better for Kharif crop and bioclimatic variable also becomes a deciding factor for growth of crops, it has been found that that precipitation Seasonality(BIO15) contribute most for predicting Isabgol crop and Mean Temperature of the Wettest Quarter(BIO 08) contributes most for Mung crop. These studies also conclude that Bheekam kor has higher predicted habitat suitability for the Isabgol crop. Baroi and Balesar have the most suitable areas for Mung crops. Machine learning models, including Maxent and SDM, provide valuable tools for predicting crop suitability and yield in Rajasthan's arid environment. These models help optimize agricultural practices by identifying suitable crops and improving yield predictions, ultimately supporting sustainable farming and food security in challenging climates [17, 18].

## REFERENCES

[1]. Lontsi Saadio Cedric a, Wilfried Yves Hamilton Adoni b, Rubby Aworka, Jérémie Thouakesseh Zoueuc,d, Franck Kalala Mutombo,e, Moez Krichenf,g, Charles Lebon Mberi Kimpolo" Crops yield prediction based on machine learning models: Case of West African countries." Published by Elsevier B.V. Tin 2022.

[2]. Bhati TK, K Shalander, H Amare and AM Whitbread. 2017. Assessment of Agricultural Technologies for Dryland Systems in South Asia: A Case Study of Western Rajasthan, India. Patancheru 502 324. Telangana, India: International Crops Research Institute for theSemi-Arid Tropics. 68 pp.

[3]. Elith, J. Species Distribution Modeling. Ecology. 2019 https://doi.org/10.1093/obo/9780199830060-0226.

[4]. Elith J & Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40(1), 677.

[5]. Guisan A & Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecological modeling, 135(2), 147-186.

[6]. Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., Yates, C. J. (2011). A Statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1): 43-57.

[7]. Phillips, S. J., Anderson, R. P., Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. Ecological modelling, 190(3): 231-259.

[8]. Kogo, B., Kumar, L., Koech, R., & Kariyawasam, C. Modelling Climate Suitability for Rainfed Maize Cultivation in Kenya Using a Maximum Entropy (MaxENT) Approach. Agronomy. 2019 https://doi.org/10.3390/agronomy9110727.

[9]. Saurabh Purohit, Neelam Rawat "MaxEnt modeling to predict the current and future distribution of Clerodendrum infortunatum L. under climate change scenarios in Dehradun district, India" in Springer 2021

[10]. Ali, S., Makanda, T., Umair, M., & Ni, J. MaxEnt model strategies to studying Current and future potential land suitability dynamics of wheat, soybean, and rice cultivation under climatic change scenarios in East Asia. *PLOS ONE.* 2023; 18. https://doi.org/10.1371/journal.pone.0296182.

[11]. Ali, S., Umair, M., Makanda, T., Shi, S., Hussain, S., & Ni, J. Modeling Current and Future Potential Land Distribution Dynamics of Wheat, Rice, and Maize und ClimateChangeScenariosUsingMaxEnt. *Land.* 2024 https://doi.org/10.3390/land13081156

[12] .Rekha Rani1*, Uma Sah1, Hemant Kumar1, Rajesh Bishnoi2, Jitendra Ojha1 and Mohit Katiyar. Mungbean in Rajasthan: Growth performance and decomposition analysis. Journal of Food Legumes 78 36(1): 78-83, 2023 DOI: 10.59797/jfl.v36.i1.13.

[13]. Lokesh Kumar Jain. International Journal of Agricultural Extension ECONOMICS AND GAP ANALYSIS IN ISABGOL CULTIVATION THROUGH FRONTLINE DEMONSTRATIONS IN WESTERN RAJASTHAN. 2009 ISSN: 2311-6110 (Online), 2311-8547 (Print) http://www.escijournals.net/IJAE.

[14]. Rathore, R. and Mathur, A. (2020). An Economic Analysis of Production of Isabgol And Constraints Faced by Farmers in Rajasthan. Economic Affairs, 65(4): 491-497.

[15]. Jhajharia, K., & Mathur, P. Prediction of crop yield using satellite vegetation indices Combined with Machine Learning Approaches. *Advances in Space Research.* 2023 https://doi.org/10.1016/j.asr.2023.07.006.

[16]. Jatav, M., Sarangi, A., Singh, D., Sahoo, R., & Varghese, C. Advanced machine learning-based kharif maize evapotranspiration estimation in semi- climate. Water science and technology : a journal of the International on Water Pollution Research. 2023; 88 4. https://doi.org/10.2166/wst.2023.253.

[17]. Choudhary, N., Chukkapalli, S., Mittal, S., Gupta, M., Abdelsalam, M., & Joshi, A. YieldPredict: A Crop Yield Prediction Framework for Smart Farms. *2020 IEEE* International Conference on Big Data (Big Data). 2020. .https://doi.org/10.1109/BigData50022.2020.9377832.

[18]. Tirth, V., Singh, R., Islam, S., Badruddin, I., Abdullah, R., Algahtani, A., Mahmoud, E., Arabi, A., Shukla, N., & Gupta, P. Kharif Crops Selection for Sustainable Farming Practices in the Rajasthan-India Using Multiple Attribute-Based Decision-Making. *Agronomy*. 2020 https://doi.org/10.3390/agronomy10040536.