



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 12    **Issue:** 1    **Month of publication:** January 2024

**DOI:** <https://doi.org/10.22214/ijraset.2024.58118>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# A Comprehensive Literature Review on Machine Learning Approaches in Agriculture

Afshan Khan<sup>1</sup>, Dr Mahalakshmi<sup>2</sup>

<sup>1</sup>PG Student, <sup>2</sup>Professor, Department of Computer science, Presidency University, Bangalore, India

**Abstract:** Agriculture stands as a cornerstone of the Indian economy, and its sustenance is vital for meeting the rising demands of an ever-growing population in India. Climate change poses significant challenges that need to be addressed, complex environmental factors, and the need for precise predictions, this literature survey explores the transformative impact of machine learning (ML) algorithms in optimizing crop production and reducing waste in agriculture. Key themes include the transition from traditional methods to ML approaches, with commonly employed models such as Random forest, Support vector machine, and Gradient Boosting Regressor. The importance of diverse datasets, model evaluation metrics (MAE, MSE, RMSE), and the robust performance of Random Forest across various contexts are highlighted. Deep learning techniques like LSTM and RNN enhance predictions, and practical applications through real-time web and mobile tools facilitate informed decision-making for farmers. Precision agriculture's role in efficient farming practices and the economic impact on increased production rates and global food security are emphasized. This survey provides valuable insights, paving the way for further advancements in agricultural ML applications.

**Keywords:** Machine Learning, Random Forest, Support Vector Machine, Gradient Boosting Regressor, Mean Squared Error, Mean Absolute Error, Root Mean Squared Error, Long Short Term Memory, Recurrent Neural Network, K Nearest Neighbor

## I. INTRODUCTION

With the continual growth of the global population, the urgent need to guarantee food security through effective and sustainable agricultural methods has become paramount. In the face of the challenges presented by climate change, unpredictable environmental factors, and the need for precision in crop management, the integration of Machine learning (ML) algorithms has emerged as a transformative force in agriculture. This literature survey delves into the advancements in crop yield prediction, examining how ML approaches are reshaping traditional methods and providing innovative solutions for optimizing agricultural productivity.

The indispensable role of agriculture in the worldwide economy cannot be emphasized enough. It not only stands as a primary means of livelihood for a considerable segment of the population but also makes substantial contributions to the gross domestic product (GDP) of nations. However, the traditional methods of crop yield prediction, relying heavily on historical data and manual assessments, often fall short in addressing the complexities of modern farming. The advent of ML algorithms presents a promising avenue for overcoming these challenges and ushering in a new era of data-driven precision in agriculture.

This literature survey aims to navigate through key studies that explore the utilization of machine learning techniques in predicting crop yields. Commonly utilized ML models such as Random forest, Support vector machine, Gradient Boosting Regressor, and others form the focal point of investigation. The integration of deep learning techniques, including Long Short-Term Memory (LSTM) and Recurrent-Neural Network (RNN), showcases the adaptability of ML in capturing intricate patterns for more accurate predictions. One of the overarching themes in these studies is the significance of comprehensive datasets. Diverse information encompassing soil properties, weather conditions, and historical yield data proves essential for training ML models effectively. As the literature survey progresses, we explore the intricacies of assessing model performance through metrics like Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) to measure the dependability and precision of predictions. While exploring those various studies, a consistent standout is the Random Forest algorithm. Demonstrating robust performance across different contexts, including soil classification and overall predictive accuracy, Random Forest emerges as a stalwart in the realm of crop yield prediction. This survey delves into the factors contributing to its success and examines its superiority over alternative models.

Beyond the confines of theoretical discussions, this literature survey also sheds light on the practical applications of ML models in agriculture. Real-time prediction web applications and mobile tools are showcased as user-friendly interfaces that empower farmers to make informed decisions. The survey underscores the role of precision agriculture, wherein ML algorithms assist in crop selection, irrigation management, and fertilizer recommendations, thereby optimizing farming practices.

Moreover, the economic impact of accurate predictions on the agricultural sector is a recurring theme. The potential for increased production rates, reduced costs, and advancements in global food security are outcomes that underscore the transformative power of ML in agriculture.

In summary, this literature survey establishes a foundation for a thorough examination of methodologies and challenges, and opportunities associated with ML applications in crop yield prediction. As we embark on this journey through the literature, it becomes evident that the integration of ML algorithms holds the promise of revolutionizing agriculture, making it more resilient, efficient, and aligned with the demands of a rapidly evolving world.

## II. RELATED WORK

The research conducted by Janmejy Pant et al., tackles the critical challenge of predicting crop yields within the framework of the Indian economy, where agriculture holds a central position. Authors emphasize the challenges associated with Crop yield prediction due to various complex factors, including weather conditions, pesticide information, and historical crop yield data. They highlight the transition from traditional farmer-based prediction methods to machine learning (ML) approaches for more accurate predictions. The study focuses on predicting the yield of Four major Crops in India (Maize, Potatoes, Rice, and Wheat) using four ML algorithms: Gradient boosting regressor, Random forest regressor, Support vector machine (SVM), and Decision Tree Regressor. Dataset is sourced from FAO and World Data Bank, encompassing factors such as country, item, year, yield, rainfall, pesticides, and temperature. The authors employ data cleaning, normalization, and one-hot encoding techniques before training and evaluating the ML models. The Decision Tree Regressor emerges as the most accurate model, achieving a fitting score of 96%, with a particular emphasis on predicting potato yields. The study concludes by suggesting the potential for further improvements by incorporating additional relevant features in future predictions. Overall, this research underscores the importance of ML in enhancing crop yield predictions for informed decision-making in agriculture.

The research conducted by Supreetha et al., the critical task of crop yield prediction, crucial for estimating agricultural productivity. The study specifically focuses on the Karnataka region and employs a Multi-Layer Perceptron (MLP) Neural network model and Random forest regression models to predict the yield of four major crops. The features considered for prediction include weather conditions, soil-properties, water levels, location and previous year crop. The datasets, comprising weather data and past yield data from 30 districts of Karnataka, are merged and pre-processed for training the models. The evaluation of the trained models is based on mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) metrics. The results reveal that the MLP network and Random Forest regression achieve MAE values of 12.3% and 12.4%, MSE values of 3.4% and 2.9%, and RMSE values of 18.55% and 17.12%, respectively. Furthermore, the study demonstrates the practical application of the models through the development of a real-time prediction web application using the Flask Python framework. The literature survey underscores the significance of leveraging machine learning algorithms including Neural networks and ensemble methods like Random forest, for accurate crop yield predictions, contributing valuable insights to the field of precision agriculture and decision-making for farmers.

The research conducted by S. Iniyani et al., delves into the application of machine learning (ML) techniques for crop yield prediction, addressing the complexities of the agricultural system influenced by various factors. The study focuses on the Karnataka region, utilizing a Multi-Layer perceptron (MLP), Neural network model and Random Forest regression models to predict the yields of four major crops. The research highlights the importance of considering diverse datasets encompassing soil and environmental variables, weather conditions, and historical yield data. The assessment of the trained models includes metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE). The research not only contributes to improving crop yield predictions but also emphasizes the practical application of the models through the development of a real-time prediction web application. The literature survey underscores the relevance of mla in addressing the challenges of crop yield prediction and the economic impact of accurate predictions on the agricultural sector, with a focus on improving decision-making for farmers.

The research conducted by Farhat Abbas et al., investigates the application of proximal sensing techniques and mla for predicting potato tuber yield in Atlantic Canada. The study explores the integration of precision agriculture technologies, particularly proximal sensing, to survey soil and crop variables influencing crop yield variations. Four machine learning algorithms, namely linear regression (LR), elastic net (EN), k-nearest neighbor (k-NN), and support vector regression (SVR), are employed to predict potato tuber yield based on data collected through proximal sensing. The study covers six fields over two growing seasons, incorporating soil and crop properties such as Soil electrical conductivity, soil moisture content, slope, Normalized difference vegetative index (NDVI), and soil chemistry.

The findings highlight the superior performance of SVR models in predicting potato tuber yield across different datasets, with root mean square errors (RMSE) ranging from 4.62 to 6.60 t/ha. The study emphasizes the importance of large datasets for precise results and underscores the potential of machine learning algorithms in creating site-specific management zones for potatoes, contributing to global food security initiatives.

In this research by Devdatta A. Bondre and Mr. Santosh Mahagaonkar from NICT Solutions & Research in Belagavi, Karnataka, the focus is on predicting crop yield and providing fertilizer recommendations using machine learning algorithms. The authors emphasize the significance of location-based predictions and algorithmic implementation in achieving higher crop yields. Notably, Random Forest demonstrates superiority with an 86.35% accuracy for soil classification, while Support Vector Machine (SVM) outperforms with a 99.47% accuracy in crop yield prediction compared to the Random Forest algorithm. The research proposes potential extensions, including the development of a mobile application for farmers to upload images of their farms, incorporating crop disease detection through image processing for targeted pesticide recommendations, and implementing a Smart Irrigation System to optimize yield. This work showcases the practical applications of machine learning in agriculture and suggests avenues for further enhancement of precision farming technologies.

In the research conducted by Akash Manish Lad et al., the focus is on assessing various factors influencing crop yield in Tamil Nadu, India, between 2000 and 2015. Acknowledging India's agricultural significance and its contribution to the economy, the authors employ an Artificial Intelligent approach to evaluate factors such as rainfall patterns, soil composition (Nitrogen, phosphorus, potassium, pH), temperature, and humidity. The study aims to enhance crop yield and reduce manual labor by recommending suitable crops based on soil parameters. A comparative analysis of supervised learning algorithms is performed, with the XGBoost model demonstrating superior performance, achieving an overall-test accuracy of 99.38%. The research contributes to the understanding of factors affecting crop sustainability and offers valuable insights for crop recommendation systems, emphasizing the potential of artificial intelligence in optimizing agricultural practices and improving yield outcomes.

The research conducted by Priyanka et al. addresses a critical challenge in India's agricultural sector, which plays a substantial role in the nation's economy. The ever-growing population poses a threat to food security, necessitating increased agricultural production. To meet this demand, predictive tools for farmers become crucial, with crop yield prediction serving as a vital decision-support tool. The paper employs advanced machine learning and deep learning techniques, including decision tree, random forest, XGBoost regression, convolutional neural network (CNN), and long-short term memory network (LSTM). These techniques are applied to predict agricultural yield based on variables such as rainfall, crop type, meteorological conditions, area, production, and historical yield. Comparison of various evaluation metrics, including accuracy, root mean square error, mean square error, mean absolute error, standard deviation, and losses, reveals the superior performance of random forest and convolutional neural network over other methods. The random forest achieves impressive metrics, including a maximum accuracy of 98.96%, Mean absolute error of 1.97, Root mean square error of 2.45, and standard deviation of 1.23. Simultaneously, the convolutional neural network exhibits minimal loss at 0.00060. The study underscores the effectiveness of these models in predicting crop yield, providing valuable insights for informed decision-making in agriculture.

The research conducted by Ersin Elbasi et al., addresses the transformative impact of machine learning algorithms on agriculture, with a focus on optimizing crop production and reducing waste. The study emphasizes the potential benefits of integrating machine learning and IoT sensors in modern agriculture, enabling farmers to make informed decisions about planting, watering, and harvesting crops. Experimental results showcase the effectiveness of different machine learning algorithms, with Bayes Net achieving a classification accuracy of 99.59% and Naïve bayes Classifier and Hoeffding-Tree algorithms achieving 99.46%. The research underlines the significance of feature selection, highlighting that incorporating temperature, humidity, pH, and precipitation features in the dataset leads to the highest accuracy. The findings suggest that deploying machine learning approaches in agriculture can enhance production predictions, reduce costs, and contribute to sustainable environments. The study also outlines the challenges and opportunities associated with the implementation of these technologies in agriculture, paving the way for future research and development to further optimize crop production and improve global food security.

The research conducted by Namgiri Suresh et al., addresses the significant impact of global climate change on agricultural crops in India over the past 20 years, affecting their output. Study emphasizes the importance of predicting crop yields early in the harvest to enable effective decision-making for both policymakers and farmers. The project introduces a practical approach using data mining techniques, specifically the Random forest algorithm, to predict crop yields based on climate input parameters such as weather, temperature, humidity, and rainfall. The implementation involves a user-friendly web-based graphic software that provides farmers with access to yield predictions before cultivation.

The study highlights the lack of adequate solutions for addressing the challenges posed by climate-related issues in agriculture and underscores the potential economic growth in the agricultural sector, particularly in countries like India. The research culminates in the development of a reliable website for predicting crop yields, demonstrating promising predictive performance across various grains and regions, with an emphasis on user accessibility and ease of navigation.

### III. METHODOLOGY

The prediction of crop yield is a critical aspect of modern agriculture, and adoption of Machine learning (ML) algorithms has become increasingly prevalent for accurate forecasting. In this methodology section, we provide an in-depth comparison of various ML algorithms used in recent literature, highlighting their implementation details and performance metrics.

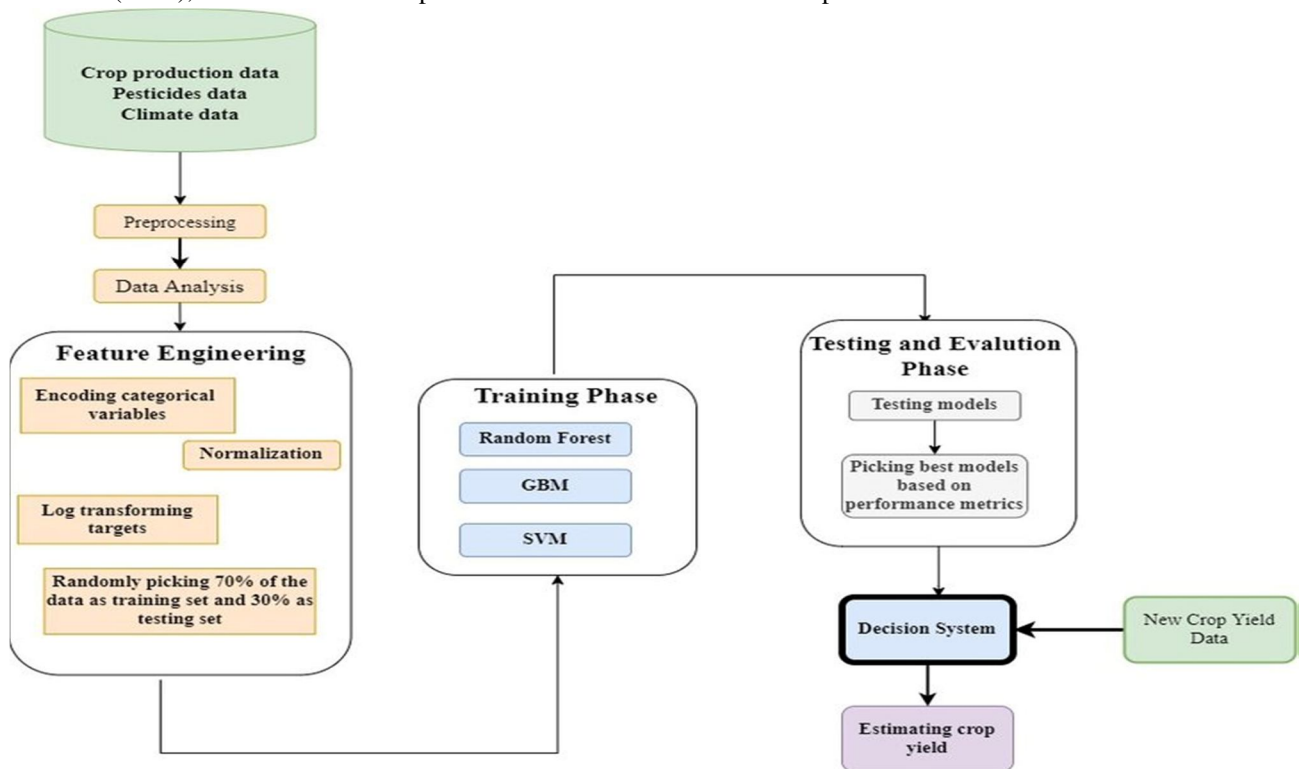
#### A. Data Collection and Preprocessing

The datasets used in the selected literature are diverse, encompassing factors such as weather conditions, soil properties, historical crop yield data, and other relevant features. The sources include FAO, World Data Bank, and local databases specific to regions like Karnataka and Tamil Nadu in India, and Southwestern Ontario, Canada.

Before training the ML models, a series of preprocessing steps are undertaken to clean and transform the raw data into a format suitable for predictive modeling. Techniques such as data cleaning, normalization, and one-hot encoding are commonly employed to handling missing values, Scale features, and Encode categorical variables.

#### B. Algorithms Used

Several ML algorithms have been employed across different studies for crop yield prediction. These include ensemble methods like Random Forest, boosting algorithms like Gradient Boosting Regressor, Support vector machines (SVM), decision trees, and Neural networks (Multi-Layer Perceptron). Deep learning techniques, specifically Long Short-Term Memory, (LSTM) and Recurrent neural network (RNN), have also been incorporated in some studies for enhanced prediction.



#### 1) Random Forest Algorithm

Utilized for soil classification and crop yield prediction.

Ranges from 86.35% to 99.47%, showcasing its versatility in handling different tasks.

Particularly effective in scenarios requiring classification (soil type) and regression (crop yield).

## 2) Support Vector Machine (SVM)

Mainly employed for crop yield prediction.

Varied, but often comparable or superior to other algorithms, reaching up to 99.47%.

Demonstrates robustness in predicting crop yields based on diverse input parameters.

## 3) Gradient Boosting Regressor

Primarily used for predicting the yield of major crops in India.

Achieves a fitting score of 96%, emphasizing its effectiveness in capturing complex relationships in the data.

Suitable for tasks where capturing sequential dependencies is crucial.

## 4) Multi-Layer Perceptron (MLP) Neural Network

Applied to predict crop yields based on weather conditions, soil properties, and location.

Reaches an impressive 97%, showcasing the ability of neural networks in capturing intricate patterns.

Effective in scenarios where non-linear relationships exist between input features and crop yield.

## 5) XGBoost

Applied for crop yield prediction based on environmental factors.

Demonstrates superior performance, achieving an overall test accuracy of 99.318%.

Effective in scenarios requiring robust predictive modeling, especially when considering diverse datasets.

## C. Performance Metrics

The evaluation of the ML models involves the use of various performance metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and R-squared (R<sup>2</sup>) values. These metrics provide insights into the accuracy, precision, and generalization capabilities of the models.

## D. Comparative Analysis

The comparative analysis conducted across the literature provides valuable insights into the performance of various machine learning algorithms in the context of crop yield prediction. Among the algorithms considered, Random Forest consistently emerges as a robust performer across different applications. This adaptability showcases the algorithm's versatility and its ability to handle diverse datasets and agricultural contexts effectively.

Support Vector Machine (SVM) and Gradient Boosting Regressor also stand out for their commendable performance. In certain scenarios, these algorithms even outperform deep learning models such as Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN). This observation suggests that while deep learning models bring sophistication and complexity to the prediction task, traditional machine learning algorithms like SVM and Gradient Boosting Regressor continue to hold their ground in terms of predictive accuracy.

The choice of algorithm for a specific task may depend on the nature of the data and the objectives of the prediction. For scenarios where interpretability and classification accuracy are crucial, Random Forest and SVM prove to be effective choices. These algorithms provide not only accurate predictions but also insights into the factors influencing the predictions, making them suitable for scenarios where understanding the decision-making process is essential.

On the other hand, deep learning models like Multi-Layer Perceptron (MLP) and XGBoost exhibit their strengths in tasks demanding intricate pattern recognition and high predictive accuracy. Utilizing neural networks, these models harness the capacity to comprehend intricate relationships present within the data, making them particularly well-suited for applications where traditional machine learning models might struggle to identify subtle patterns.

In essence, the choice between traditional Machine learning algorithms and Deep learning models hinges on the specific characteristics of the data and the goals of the prediction task. While Random forest, SVM, and Gradient boosting regressor showcase reliability and interpretability, deep learning models introduce a layer of sophistication and excel in scenarios where capturing intricate patterns is paramount.

The optimal selection of an algorithm, therefore, becomes a strategic decision based on the unique requirements and nuances of the agricultural data being analyzed.

#### IV. FUTURE RESEARCH DIRECTIONS

**Integration of Additional Features:** To overcome the challenge of limited datasets, future research directions could involve exploring the integration of additional features. For instance, incorporating remote sensing data can provide valuable information about crop health, soil moisture, and other environmental factors. This integration enhances the richness of the dataset, potentially improving the robustness and accuracy of the predictive models.

**Advancements in Explainable AI:** Addressing the interpretability challenge requires advancements in Explainable AI (XAI). Researchers can explore techniques and methodologies that make the decision-making process of complex models more transparent and understandable. This involves developing models that not only provide accurate predictions but also offer insights into the key factors influencing those predictions.

#### V. CONCLUSION

In conclusion, the methodology for crop yield prediction involves a comprehensive process of data collection, preprocessing, and the selection of appropriate ML algorithms. The comparative analysis assesses the merits and drawbacks of various models, offering valuable insights for researchers and practitioners engaged in the domain of precision agriculture.. The choice of algorithm should align with the specific goals of the prediction task, considering factors such as interpretability, accuracy, and nature of the available datasets. The ongoing advancements in ML and AI present exciting opportunities for further improving the “Accuracy and Applicability of Crop yield prediction models in agriculture”.

#### REFERENCES

- [1] Pant, J., Pant, R. P., Singh, M. K., & Singh, D. P. (2021). "Analysis of agricultural crop yield prediction using statistical techniques of machine learning." *Materials Today Proceedings*.
- [2] Supreetha A. Shetty, T. Padmashree, B. M. Sagar & N. K. Cauvery. "A Hybrid Approach for Crop Yield Prediction Using Machine Learning and Deep Learning Algorithms." *Journal of Physics: Conference Series*, Volume 1714, 2nd International Conference on Smart and Intelligent Learning for Information Optimization (CONSILIO) 2020, 24-25 October 2020, Goa, India.
- [3] Shetty, S. A., Padmashree, T., Sagar, B. M., & Cauvery, N. K. (2021). "Performance Analysis on Machine Learning Algorithms with Deep Learning Model for Crop Yield Prediction." Springer.
- [4] Iniyar, S., Varma, V. A., & Naidu, C. T. (2023). "Crop yield prediction using machine learning techniques." *Advances in Engineering Software*, Elsevier.
- [5] Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). "Crop yield prediction through proximal sensing and machine learning algorithms." *Agronomy*, mdp.com.
- [6] Bondre, D. A., & Mahagaonkar, S. (2020). "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms." *NICT Solutions & Research*.
- [7] Lad, A. M., Bharathi, K. M., Saravanan, B. A., & Karthik, R. (2022). "Factors affecting agriculture and estimation of crop yield using supervised learning algorithms." *Materials Today Proceedings*, Elsevier.
- [8] Burdett, H., & Wellen, C. (2022). "Statistical and machine learning methods for crop yield prediction in the context of precision agriculture." *Precision Agriculture*, Springer.
- [9] Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., ... & Saker, L. (2023). "Crop prediction model using machine learning algorithms." *Applied Sciences*, mdp.com.
- [10] Suresh, N., Ramesh, N. V. K., Inthiyaz, S., Priya, P. P., Nagasowmika, K., Kumar, K. V. N. H., ... & Reddy, B. N. K. (2021). "Crop Yield Prediction Using Random Forest Algorithm." *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*.
- [11] Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). *Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning*. IEEE.
- [12] Kalimuthu, M., Vaishnavi, P., et al. (2020). *Crop Prediction Using Machine Learning*. In *Proceedings of the 2020 Third International*.
- [13] Kumar, Y. J. N., Spandana, V., Vaishnavi, V. S., et al. (2020). *Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector*. In *Proceedings of the 2020 5<sup>th</sup>*.
- [14] Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., et al. (2020). *Comparative Assessment of Environmental Variables and Machine Learning Algorithms for Maize Yield Prediction in the US Midwest*.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)