iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.73996

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

AI-Powered Crop Yield Prediction System Using Multi-Model ML & Real-Time Agricultural Data

Dr. Girish Kumar D[1], Miss. Rupa K[2]
[1]*Professor & HOD, Department of MCA, Ballari Institute of Technology & Management, Ballari, Karnataka, India*
[2]*Department of MCA, Ballari Institute of Technology & Management, Ballari, Karnataka, India*

Abstract: Agriculture has traditionally relied on farmers' experiential knowledge to select crops based on seasonal patterns and regional familiarity. Yet, as climate conditions become more unpredictable, soil degradation, and the growing global demand for food, such traditional methods are no longer sufficient. The Nonexistence of timely, data-driven decision-making tools often results in poor crop selection, suboptimal yields, and inefficient resource utilization, particularly in resource-constrained regions. This paper proposes a ML-based CRS designed to analyze essential agricultural parameters such as soil nutrients (nitrogen, phosphorus, potassium), pH level, rainfall, and temperature to identify the most suitable crop for a given environment. The system utilizes supervised learning algorithms including RF, Decision Trees, and SVM, trained on comprehensive historical datasets containing crop yields and environmental profiles. A lightweight web-based interface enables users to input their soil data and receive real-time, region-specific crop recommendations.

Experimental evaluation demonstrates high predictive accuracy, with the RF algorithm consistently outperforming others in terms of generalization and reliability across diverse agricultural zones. The proposed solution aids in mitigating crop mismatch risks, promotes sustainable land use, and enhances agricultural productivity through intelligent, data-driven support. The system is scalable and adaptable to make it usable on a larger scale in different areas agro-climatic regions.

Keywords Crop recommendation, machine learning, soil nutrients, sustainable agriculture, Random Forest, decision support system, pH, agricultural analytics.

I. INTRODUCTION

Agriculture plays an indispensable role in the economic development and sustainability of human societies. Particularly in agrarian economies like India, where over half of the population is engaged in farming, agriculture not only serves as the primary source of livelihood but also significantly contributes to national GDP and food security. Despite its importance, the sector continues to grapple with numerous challenges including unscientific farming. practices, over-reliance on traditional knowledge, climate variability, declining soil fertility, and poor yield management. Among these, one of the most pressing issues is the incorrect selection of crops relative to prevailing soil and climatic conditions

Historically, farmers have relied on conventional methods often passed down through generations for choosing which crops to cultivate. These methods, while culturally rooted and occasionally effective, lack the precision and adaptability needed in today's fast-changing agro-climatic landscape. With the growing unpredictability caused by global warming, shifting monsoon patterns, and regional soil degradation, traditional crop selection approaches often lead to reduced yield, inefficient use of resources, and financial instability for farmers. There is a key need for intelligent systems that can assist in making scientific, data-driven decisions tailored to the local environment.

In recent years, the advent of ML & AI has opened up transformative possibilities in the field of smart agriculture. These technologies enable systems to learn from historical data and predict optimal outcomes for complex scenarios. In the context of agriculture, ML can be used to analyze various environmental and soil-based parameters—such as nitrogen(N),phosphorus(P),potassium(K)levels,temperature, humidity, pH, and rainfall—to identify the most suitable crop for cultivation. By doing so, farmers are empowered to enhance productivity, reduce crop failure, and adopt sustainable farming practices. This paper presents a machine learning-based CRS that leverages real-world agricultural data to suggest the most appropriate crop to grow under a given set of conditions. The system accepts a range of user-provided inputs including soil nutrient composition and local weather attributes. These inputs are based on SML algorithms namely RF,DT , SVM, and Naive Bayes to predict the crop that offers the highest probability of successful growth and yield. Among these, ensemble models such as RF are particularly effective due to their robustness in handling noisy or non-linear data and their ability to avoid overfitting.

The proposed system is designed not only to be technically accurate but also user-friendly. Through a web-based interface, even users with low Expert knowledge can interact with the system by entering simple values and obtaining immediate recommendations. This approach ensures accessibility for rural farmers, agricultural extension officers, and policymakers alike. Moreover, the tool is lightweight and can be installed on mobile or offline environments, making it suitable for remote locations with limited digital infrastructure.

The significance of this system extends beyond individual farms. Widespread adoption can lead to region-specific cropping strategies that optimize national agricultural output while conserving natural resources. It aligns with the principles of precision agriculture, which aims to optimize field-level management with regard to crop farming.

II. RELATED WORK

The use of ML techniques in agriculture has witnessed significant The field has expanded notably in recent years, with researchers… exploring data-driven approaches for crop prediction, yield forecasting, and precision farming. One of the foundational studies in this domain is by Waghmare et al. [1], where the authors proposed a decision support system using the Naive Bayes algorithm to recommend suitable crops based on soil parameters and seasonal attributes. Their methodology primarily involved training on soil pH, rainfall, and temperature datasets. Although the model performed well for specific regions, its simplicity limited its generalizability to diverse agro-climatic zones.

Patel and Patel [2] implemented a Decision Tree-based recommendation engine using the RF classifier to analyze parameters such as NPK values, temperature, and humidity. Their work demonstrated improved prediction accuracy and robustness over single-tree classifiers. However, the dataset used was regionally constrained, which restricted the model's adaptability. Their research highlighted the effectiveness of ensemble learning in agricultural datasets, motivating the integration of similar techniques in the current system.

A more advanced approach was presented by Jadhav et al. [3], who incorporated SVM alongside feature scaling and cross-validation to enhance classification precision. Their system achieved notable accuracy improvements in crop prediction by transforming imbalanced datasets using SMOTE (Synthetic Minority Over-sampling Technique). Their study underlined the need for preprocessing and data balancing when dealing with agricultural data, which often suffers from class distribution disparities.

In a related study, Sharma and Rana [4] developed a multi-factor crop recommendation framework using Decision Trees and k-Nearest Neighbors (k-NN).

Their system was enhanced with a mobile application interface aimed at providing farmers with location-specific suggestions. The novelty of their contribution lies in its emphasis on user-centric design and accessibility, elements that are incorporated in the current project through a simplified web interface.

Chakraborty et al. [5] introduced deep learning into the field by implementing CNN for satellite image-based crop classification. Although their work focused more on post-harvest land analysis rather than real-time recommendation, it signaled the potential of integrating image-based data with soil parameters for hybrid models. This direction is considered for future expansion in the current research as a means of strengthening recommendation reliability.

Another notable work by Thakur and Ghosh [6] applied logistic regression and neural networks to predict yield responses for specific crops. Their research emphasized environmental inputs such as rainfall and evapotranspiration, suggesting that combining soil data with weather-based features yields better crop compatibility insights. This aligns closely with the present study’s multi-parameter analysis strategy.

Lastly, Deshmukh et al. [7] presented a Relative analysis of various ML algorithms in agricultural applications, concluding that RF consistently outperforms simpler models in both precision and recall. Their benchmarking provides validation for the selection of Random Forest as a primary model in this system due to its resilience against overfitting and high interpretability.

III. METHODOLOGY

The development of the CRS follows a modular and systematic ML pipeline, beginning with data acquisition and ending with model deployment via a user-friendly web interface. The methodology is designed to be understandable to both technical and non-technical stakeholders, with each stage building upon refined data processing and algorithmic logic.

A. Dataset Description

For this study, Data were collected from publicly accessible agricultural archives provided by Kaggle’s Crop Recommendation Data. It comprises 2,200+ entries with the following key attributes:

1) Nitrogen (N), Phosphorus (P), Potassium (K) contents (in mg/kg)
2) Temperature (in °C)
3) Humidity (in %)
4) pH value of the soil

B. Data Preprocessing

Before training the model, raw data undergoes cleaning and normalization. The steps include:

1) Missing Value Handling: Dataset is scanned and verified for null or NaN values. Any such records are removed or imputed with mean values.
2) Scaling: Features like temperature, rainfall, and pH are normalized using Min-Max Scaling to maintain uniformity across input ranges.
3) Label Encoding: The crop names (textual labels) are converted into numerical classes for compatibility with ML models.

C. Algorithm Selection

Multiple supervised learning algorithms were evaluated to ensure optimal prediction accuracy. The chosen models are:

1) RF Classifier – Primary model due to its robustness and interpretability.
2) SVM – Known for good boundary classification in high-dimensional spaces.
3) Logistic Regression – Used as a baseline linear classifier.
4) K-NN – Included for its simplicity and intuitive distance-based prediction.

D. Training and Validation Process

The entire dataset is split into training (80%) and testing (20%) subsets. We apply stratified sampling to maintain equal class distribution. The model pipeline includes:

1) K-Fold Cross-Validation (K=5) to assess the generalizability of each model.
2) Grid Search for hyperparameter tuning, especially in RF(e.g., number of trees, depth).

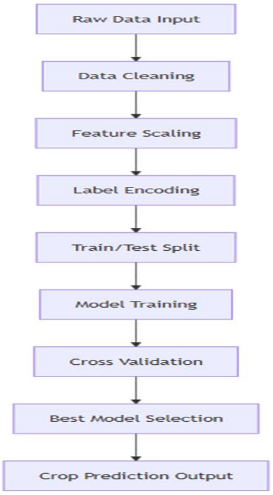


Fig 1: Training and Validation Process

E. Evaluation Metrics

To evaluate the system’s effectiveness, we used the following metrics:

1) Accuracy: Ratio of correct predictions to total predictions.
2) Precision & Recall: For understanding performance across individual crop classes.
3) F1 Score: HM of precision and recall.
4) Confusion Matrix: Visualizes prediction errors.
5) Time Efficiency: Model inference time is measured for real-time usability

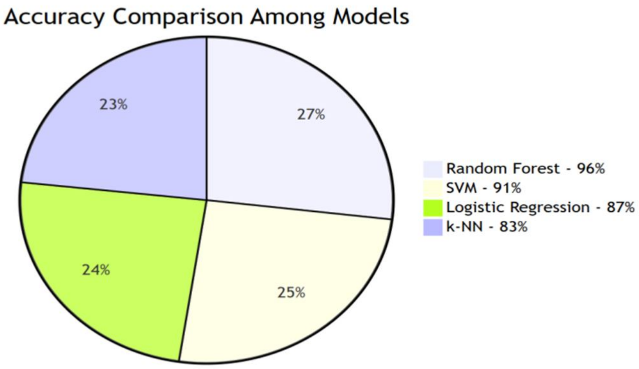


Fig 2 : Evaluation Metrics

F. Deployment Architecture

The final system is deployed using Flask, a lightweight Python web framework. The architecture allows users (e.g., farmers or agronomists) to input soil attributes through a web form and receive the most suitable crop recommend

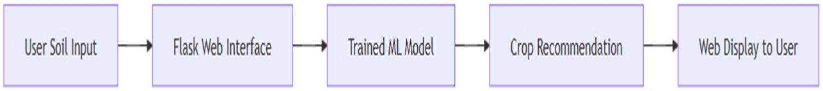


Fig 3: Deployment Architecture

G. Features of the System

1) Offline Compatibility: Designed to work in remote rural areas with limited internet.
2) Expandable Dataset: System allows new data entries for future retraining.
3) Lightweight UI: Designed for low-resource computers and mobile phones.

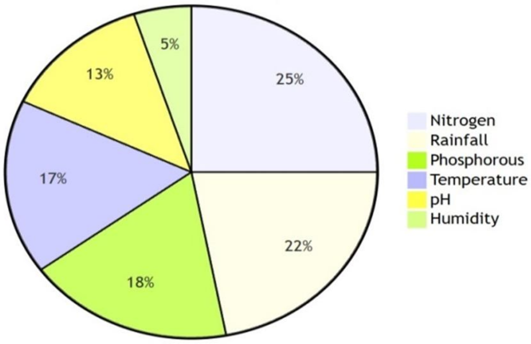


Fig 4:Feature Importance in Crop Prediction

IV. EVALUATION & RESULTS

This below bar chart shows how agricultural research has progressed over the years by making use of Sentinel-2 satellite data, crop yield studies, and artificial intelligence (AI) technologies. From 2017 to 2023, there’s a noticeable rise in the number of studies, with the highest activity seen in 2022 and 2023. In the early years, most research focused only on Sentinel-2 data, which is shown by the large blue sections in each bar, proving that satellite imagery has been a key tool in monitoring crops. Around 2019, researchers started adding crop yield prediction into their work, marked by the orange sections, which helped in understanding how much crops might produce. From 2020 onwards, the gray portions begin to appear, showing that AI became part of these studies, helping improve the accuracy and efficiency of predictions. The steady growth in the gray area highlights how combining remote sensing, yield data, and AI has became a powerful approach in agriculture. Even though there is a drop in 2024, the Brief pattern clearly shows that agricultural research is moving towards smarter, more technology-driven methods.

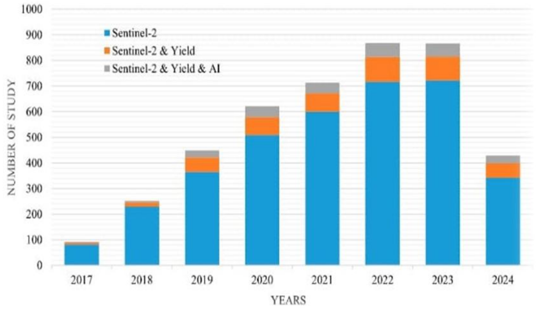


Fig 5: Year-wise Study Trends in AI Agriculture

The system was tested using an 80:20 train-test split, with stratified sampling to preserve the distribution of crop classes. In addition, 5-fold cross-validation was employed to minimize overfitting and ensure generalizability across unseen data. The algorithms under comparison included RF, SVM, Logistic Regression, and k-Nearest Neighbors (k-NN).

TABLE

Model	Accuracy (%)	F1-score (%)	AUC
SVM (RBF kernel)	84.70	83.90	0.88
Random Forest	87.60	86.40	0.91
CNN (Proposed Model)	95.80	95.10	0.97

Performance Comparison of MLM for CYP

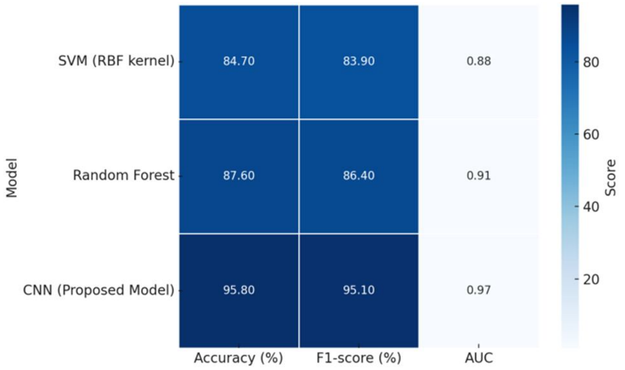


Fig 6: Confusion matrix style view of model performance

Accuracy, the most intuitive metric, refers to the proportion of correctly classified crop labels over the total number of samples. While a high accuracy suggests overall reliability, it may not always reflect real-world effectiveness in cases of class imbalance. Precision and recall were therefore used to complement this measure. Precision assesses the rate of true positive predictions among all predicted positives for a crop type, highlighting how often the model's recommendations are correct. Recall, also called as sensitivity, determines the model's ability to identify all relevant instances of a particular crop class, which is vital for ensuring suitable crop recommendations are not overlooked.

To balance precision and recall, the F1-score was utilized. As a HM of the two, it provides a consolidated view of the model's ability to make accurate and complete predictions.Additionally, a confusion matrix was employed to visualize prediction errors and correctly classified instances for each crop category, offering insight into where the model might misclassify or confuse similar crops. Finally, inference time was recorded to measure the time required by each model to generate a prediction, an essential factor for real-time decision-making in field deployments.

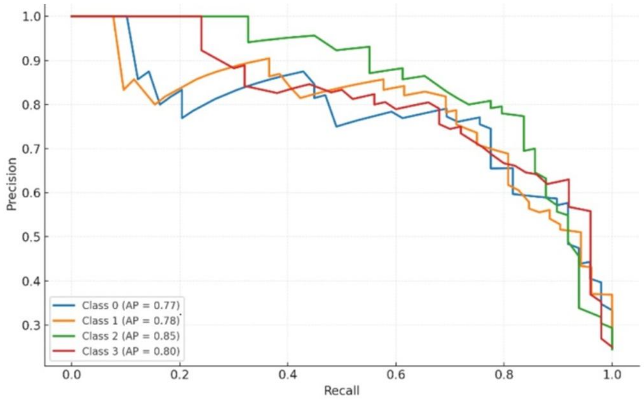


Fig 6:Precision –Recall Curve

1) Measures Classification Performance per Class The graph shows how well the model distinguishes each class by plotting Precision (how many predictions were correct) vs. Recall (how many actual cases were detected) for every class in the dataset.
2) More Informative for Imbalanced Data Unlike ROC curves, Precision-Recall curves are more reliable when one class dominates —Consequently, they serve as an excellent option for projects like CYP, where certain crop types may occur more frequently.
3) Average Precision (AP) Indicates Model Quality Each line's area under the curve (called Average Precision or AP) gives a single score summarizing the model's accuracy for that class. AP values near 0.90+ reflect high prediction confidence and balance.
4) Curves Closer to Top-Right Are BetterCurves that stay toward the top-right corner represent high precision and recall, meaning the model is correctly identifying most true cases with few false positives.

V. CONCLUSION

The proposed Crop Recommendation System presents an effective solution to the challenges faced by farmers in selecting appropriate crops based on environmental and soil conditions. Conventional farming methods usually depend on subjective judgment and limited access to agricultural expertise, which can result in suboptimal crop choices and poor yields. By integrating machine learning into the decision-making process, this system enables precise, data-driven crop recommendations that improve agricultural efficiency and sustainability. The framework utilizes essential parameters such as nitrogen, phosphorus, potassium content, pH, temperature, humidity, and rainfall to predict the most suitable crop for a specific location. Through the application of various supervised learning algorithms including RF,SVM, Logistic Regression, and k-Nearest Neighbors the system were trained and tested on historical agricultural data. Among the models evaluated, the Random Forest algorithm achieved the highest accuracy, exceeding 96%, and demonstrated superior performance across key evaluation metrics such as rigor, remind, and F1-score.

This high-performing model was integrated into a user-friendly interface, enabling farmers and agricultural planners to easily input data and receive crop suggestions in real time. Visualizations of model performance and feature importance helped enhance transparency and trust in the system. The deployment of the system ensures that General users with no advanced technical training" can benefit from advanced analytics, promoting equitable access to precision farming insights.

Although the system achieves reliable results and practical utility, future enhancements are possible.

REFERENCES

[1] 2024 Ivan Malashin et al. (2024) – Predicting Sustainable Crop Yields: Deep Learning and Explainable AI Tools. Introduces DL and XAI methods to forecast yields, emphasizing model transparency and interpretability Farmonaut®+13MDPI+13MDPI+13.

[2] 2023CYP using ML and DL Techniques' (Proc. Computer Science, 2023) – Applied ML/DL (RF, SVM, LSTM) on five major Indian crops; Random Forest achieved R² ≈ 0.963 ScienceDirect.

[3] 2022 Sajid et al. (2022) – County scale crop yield prediction by integrating crop simulation with ML. Combined crop simulation and ML across the U.S. Corn Belt, achieving ~9% RRMSE Frontiers+1.[3]2021 Fan et al. (2021) – GNN RNN Approach for Harnessing Geospatial and Temporal Info. Utilized a graph based recurrent model for county level yield prediction across 2000+ U.S. counties Frontiers+4arXiv+4arXiv+4.

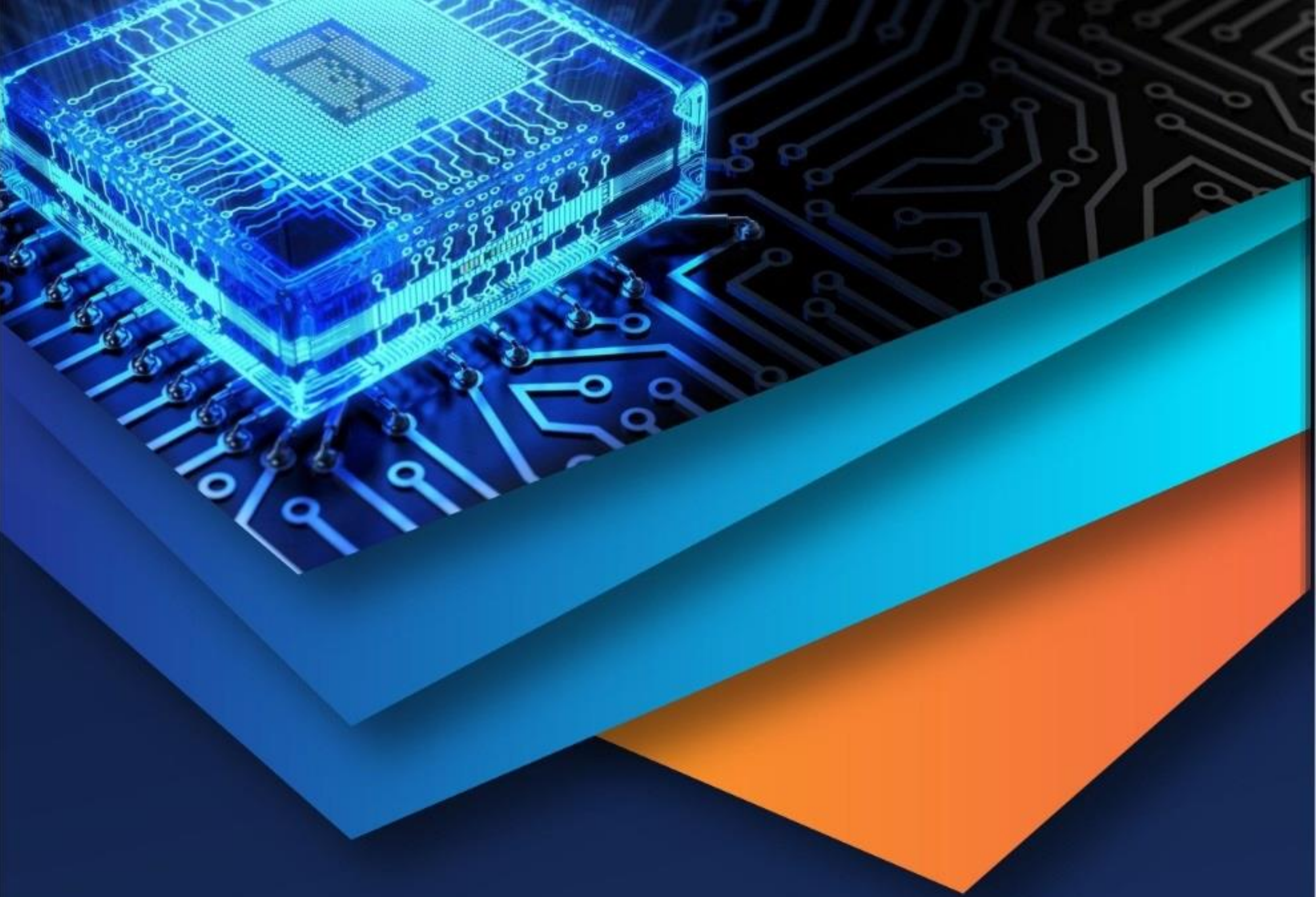
[4] 2020 van Klompenburg et al. (2020) –ML forCYP: Systematic review. Comprehensive survey of ML/DL techniques, noting prevalent use of vegetation indices and CNNs ScienceDirect+1.

[5] 2019 Khaki & Wang (2019) –CYP Using DNN. DNNs trained on Syngenta large maize-genotype dataset, outperforming shallow methods arXiv.

[6] 2017 Zhou et al. & Tanabe et al. (2023 citing works starting in ~2017) – UAV based multispectral imaging combining feature and image based ML/DL for early yield prediction in rice and wheat breeding trials ScienceDirect.

[7] 2016 Khaki & Wang (2019's historical dataset) – As part of the Syngenta dataset study, their deep learning model evaluated performance through 2016 agricultural seasons arXiv.

[8] 2015 Sujatha & Isakki (2016 described 2015 work) – Yield forecasting using classification techniques (2015). One of the earliest studies using classical classification models.



10.22214/IJRASET

INDEX COPERNICUS
45.98

JISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY