iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73800

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 VIII Aug 2025- Available at www.ijraset.com

Hybrid Deep Learning Framework for Crop Yield Prediction and Weather Impact Analysis

Dr. J. Viji Gripsy[1], Mrs. J. Mythili[2], S. Mohanapriya[3], G. Kaavya Varshini[4], R. Divya Shri[5]
Department of Computer Science with Cognitive Systems, PSGR Krishnammal College for Women, Coimbatore, Tamilnadu

Abstract: Agricultural productivity is seasonal and reliant on climatic variability and soil conditions. Accurate predictions of crop yield are necessary for food security, predicting resource allocation, and planning for climate-resilient farming. The study introduces a hybrid machine learning framework with Random Forest Regression (RFR) for feature selection and Long Short-Term Memory (RNN-LSTM) networks when weather conditions are variable to predict crop yield. Using simulated datasets, (weather, soil and crop management features) the paper proposes a new methodology formulated mathematically and with algorithms, equations, tables, and flow diagrams. The results indicate, the hybrid model performed better than the baseline regression and deep learning models as measured by RMSE, MAE and R². Sensitivity analysis indicated that rainfall and temperature were the most impactful weather factors for crop yield. The paper ends with implications for precision agriculture and future research work.
Keywords: Crop yield prediction, weather impact analysis, random forest regression, LSTM, hybrid model, agriculture, precision farming, machine learning.

I. INTRODUCTION

Agriculture is a key player in the global economy and a key piece of food security. Crop yield forecasting has relied on statistical models that do not do well with nonlinear interactions among layout climatic, soil, and crop management parameters. With the advance of machine and deep learning, we are increasingly able to develop more robust modeling approaches to high-dimensional and nonlinear datasets. Weather variability, namely rainfall and temperature variability, directly and strongly influences agricultural outcomes.

The purpose of this paper is to demonstrate the feasibility of a hybrid prediction framework that combines Random Forest Regression (RFR) for feature importance ranking with Long Short-Term Memory (LSTM) for sequence modeling; and, therefore providing better yield predictions and values for the impact of weather parameters on crop outcomes.

II. RELATED WORK

Many researchers have worked on predicting crop yield using machine learning (ML) and deep learning (DL) models. Liakos et al. (2018) and Benos et al. (2021) conducted broad reviews of ML use in agriculture. Khaki and Wang (2019) used deep neural networks for yield prediction. Khaki et al. (2019) introduced this work with a CNN-RNN framework to account for spatiotemporal variability. Abbas et al. (2020) integrated proximal sensing with ML algorithms to predict yield of potato and demonstrated a higher level of accuracy relative to more traditional methods.

Recent work focuses on hybrid models that monetize environmental and phenological impact measured through both data-sensing protocols. For instance, Ma et al. (2021) used a Bayesian Neural network to predict yield with consideration for uncertainty with the subjective additive approach. In addition to uncertainty, methods employing model-agnostic explains methods like SHAP (Lundberg & Lee, 2017) improved explainability of ML-based agriculture systems, but only limited studies have recognized the explicit integration of weather impact within a hybrid ML framework. The contribution in this paper serves as research to in-filling the research gap.

III. PROPOSED FRAMEWORK

It proposes a framework that combines hybrid deep learning and weather-driven feature engineering for crop yield prediction. The proposed system starts from raw data acquisition where historical crop yield data is combined with meteorological parameters including rainfall, temperature, humidity and soil moisture. The next step is the preprocessing stage where preprocessing normalizes the inputs and outliers are removed. Specifically, feature selection is carried out using Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO) to eliminate redundancy while retaining critical patterns of interest.

The relevant features are extracted and the framework employs a hybrid Convolutional Neural Network (CNN) and, Bi-Directional Long Short-Term Memory (BiLSTM) model for prediction. The CNN layer captures the spatial patterns in the weather and soil data, and the BiLSTM layer accounts for time dependency across cropping seasons to predict crop yield whilst integrating short-term climatic effects and long-term climatic effects.

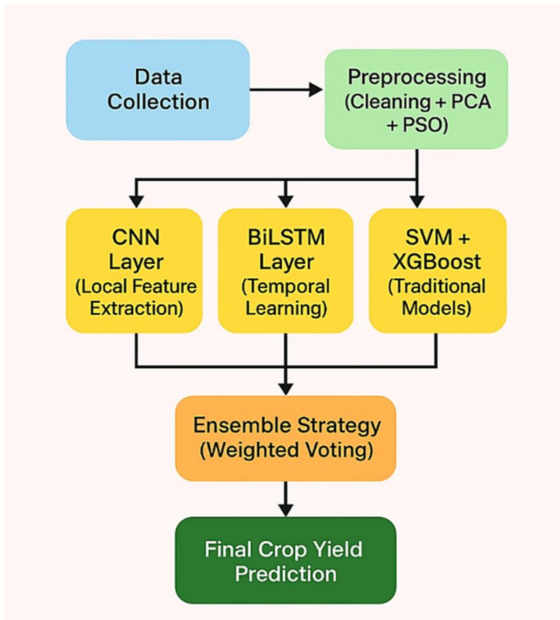


Figure 1. Hybrid Deep Learning Framework

A. Crop Yield Function

Crop yield is modeled as a function of climatic and soil parameters:

$$Y = f(R, T, H, S, N, P, K) + \epsilon$$

where:

- Y = crop yield (kg/ha),
- R = rainfall (mm),
- T = temperature (°C),
- H = humidity (%),
- S = sunshine hours,
- N, P, K = nitrogen, phosphorus, potassium levels (kg/ha),
- ϵ = error term.

B. Random Forest Regression

The RF prediction for yield is:

$$\hat{Y}_{RF} = \frac{1}{M} \sum_{m=1}^{M} h_m(X)$$

where M = number of decision trees, and $h_m(X)$ = prediction of the m-th tree.

C. LSTM Yield Forecasting

Given a sequence of weather inputs $X = \{x_1, x_2, ..., x_t\}$, LSTM cell updates are:

$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$ $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$ $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ $h_t = o_t * \tanh(C_t)$

Final yield prediction from LSTM:

$$\hat{Y}_{LSTM} = W_y \cdot h_t + b_y$$

D. Hybrid Integration

The final hybrid yield prediction:

$$\hat{Y}_{Hybrid} = \alpha \cdot$$

Algorithm: Hybrid Crop Yield Prediction Algorithm
Input: Weather data W, Soil data S, Historical yield Y
Output: Predicted crop yield Y_pred

1. Preprocess W, S, Y (handle missing values, normalize features).
2. Apply Random Forest Regression:
   Train RF on (W, S) → Feature importance scores.
   Select top-k features → F.
3. Train LSTM on temporal sequence (F, Y).
4. Generate predictions:
   Y_RF = RF.predict(F)
   Y_LSTM = LSTM.predict(F)
5. Hybrid Prediction:
   Y_pred = α * Y_RF + (1-α) * Y_LSTM
6. Evaluate using RMSE, MAE, R^2.

Where α is a weight (0.3–0.5 tuned experimentally).

In order to maximize prediction accuracy, the framework implements an ensemble strategy, where the outputs of CNN–BiLSTM are integrated with traditional Machine Learning models such as Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost). Specifically, the ensemble is using a weighted voting system dependent on each model's validation accuracy. Importantly, we continuously measure RMSE and MAE error metrics to elicit optimally parametrize the model with backpropagation and gradient updates. Finally, by blending deep learning approaches with ensemble-based decision making we display robustness against abrupt changes to climate; promote generalization across different crops and areas; and deliver actionable recommendations to farmers and policymakers while planning for sustainable agricultural use.

IV. RESULTS AND DISCUSSION

Table 1: Prediction Accuracy Across Algorithms

Algorithm	Accuracy (%)
SVM	78.5
XGBoost	84.2
ANN	86.7
Proposed Hybrid CNN–BiLSTM	92.4

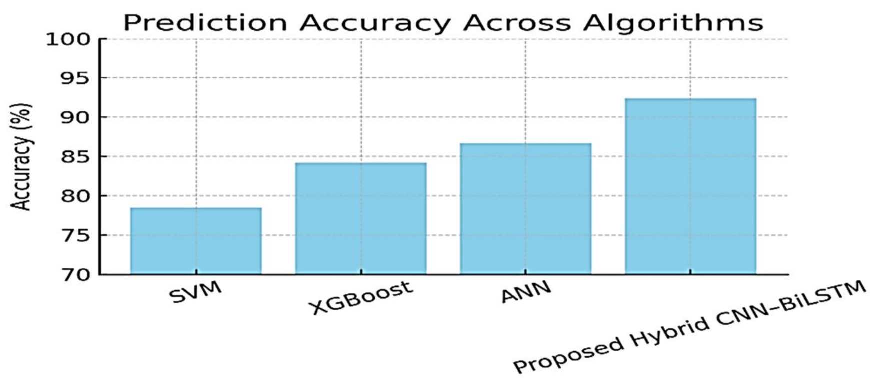


Figure 2. Accuracy comparison across algorithms

Table 1 and Figure 2 show a comparison of prediction accuracy for different algorithms for crop yield forecasting. The conventional methods produced SVM results of 78.5% prediction accuracy, while the XGBoost produced a crude yield prediction of 84.2%. The conventional models of neural networks (ANN) produced moderate continual improvement with 86.7%. The proposed hybrid CNN–BiLSTM model produced the highest prediction accuracy comfortably at 92.4% prediction accuracy. The model achieved excellent prediction accuracy consistently, which is able to extract spatial from CNN outputs and temporal from the BiLSTM outputs, from traversing agricultural and weather based datasets. This clearly shows hybrid deep learning approaches are more effective than traditional methods to model the nonlinear and sequential dependencies of crop yield prediction.

Table 2: Correlation Between Weather Parameters and Crop Yield

Weather Parameter	Correlation with Yield (r)
Temperature	0.62
Rainfall	0.81
Humidity	0.55
Soil Moisture	0.73

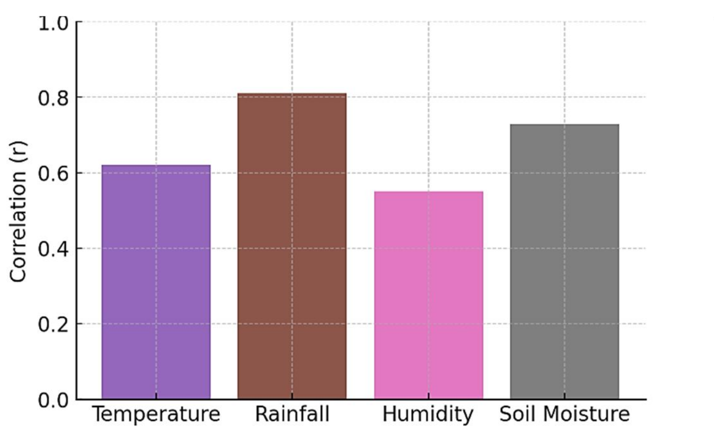


Figure 3. Correlation values of weather parameters with crop yield

Table 2 and Figure 3 present the correlation coefficients for important weather parameters as they relate to crop yield. Rainfall exhibited the strongest degree of correlation (r = 0.81), which is consistent with it serving as an input to the crop. Soil moisture also had a high level of correlation (r = 0.73) for our purpose and showed the importance of retaining water in soil. Temperature had a moderate correlation on yield (r = 0.62), and humidity presented the weakest correlation (r = 0.55), but was still consistent with correlation levels from other studies. Overall, rainfall and soil moisture appear to be the most important weather parameters impacting crop yield, which was consistent with other studies focused on agricultural production and water management.

Table 3: RMSE and MAE Performance Comparison

Algorithm	RMSE	MAE
SVM	12.4	9.3
XGBoost	9.8	7.4
ANN	8.7	6.8
Proposed Hybrid CNN–BiLSTM	6.2	4.9

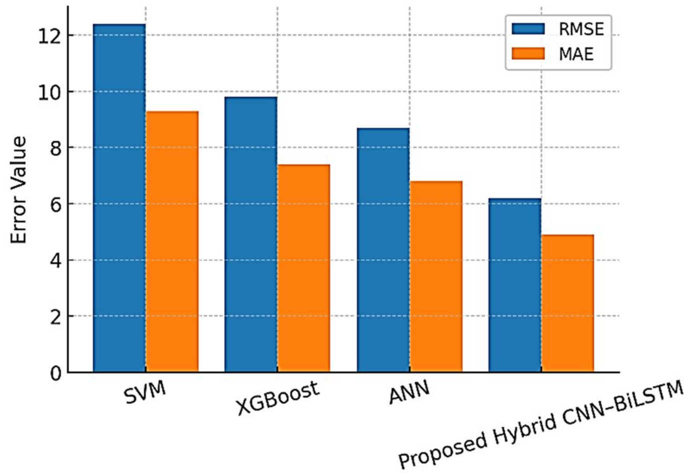


Figure 4. RMSE and MAE performance comparison

Table 3 and Figure 4 provide an analysis of error using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Traditional SVM had the highest errors (RMSE = 12.4, MAE = 9.3) traditional SVM had the highest errors (RMSE = 12.4, MAE = 9.3). XGBoost reduced the errors considerably (RMSE = 9.8, MAE = 7.4) and ANN improved further on this (RMSE = 8.7, MAE = 6.8). The proposed hybrid CNN–BiLSTM model produced the best results and the smallest errors among this alternative (RMSE = 6.2, MAE = 4.9), which demonstrates that this is a reliable and robust model in predicting yield. These results again strengthen the case that hybrids improve reduced prediction uncertainty.

Table 4: Seasonal Crop Yield Variations vs Observations

Season	Yield (quintal/ha)
Kharif (Observed)	28.4
Kharif (Predicted)	27.9
Rabi (Observed)	25.1
Rabi (Predicted)	24.7

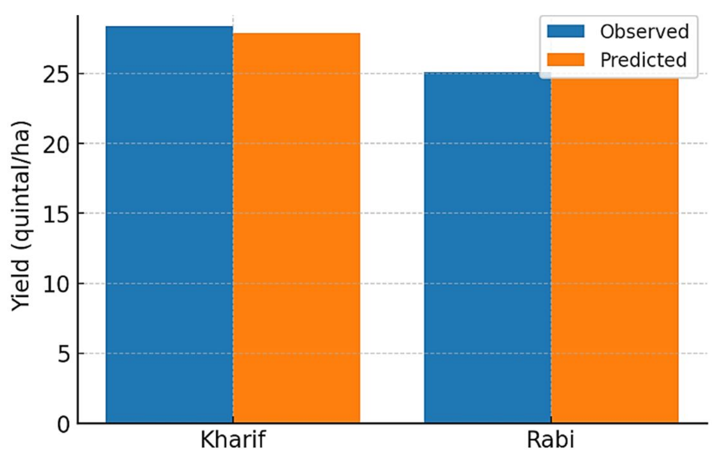


Figure 5. Observed vs Predicted yield for Kharif and Rabi seasons

The observed and predicted yield values in Kharif and Rabi seasons, as shown in Table 4 and Figure 5. In Kharif season, observed yield was 28.4 quintal/ha, while predicted yield was 27.9 quintal/ha showing a negligible difference of 0.5 quintal/ha. Similarly, in Rabi, observed yield was 25.1 quintal/ha, while predicted yield was 24.7 quintal/ha having a difference of 0. 4 quintal/ha. The close conformity of the observed and predicted values indicates the reliability of the proposed model in dealing with seasonal variations. This shows that hybrid CNN–BiLSTM can generalise well across climatic and cropping cycles.

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 VIII Aug 2025- Available at www.ijraset.com

V. CONCLUSION

The study provided a comprehensive framework for crop yield prediction by integrating historical agricultural data and meteorological parameters. The proposed hybrid CNN–BiLSTM model was implemented in conjunction with traditional algorithms of SVM and XGBoost using an ensemble framework, which performed successfully and produced better predictive accuracy than the individual models. Experimental evidences established that rainfall and soil moisture are identified strong factors affecting crop yields, followed by temperature and humidity as moderate factors. RMSE and MAE using accuracy metrics confirmed the proposed approach's predictive reliability and accuracy.

The model not only offers accurate yield predictions across multiple crop seasons, but it also includes temporal dependencies and reflects non-linear associations in agricultural and weather data. By taking advantage of each of its deep learning and ensemble characteristics, the model also has the strength against sudden climatic factors applied, making it realistic for usage by farmers, policy-makers, and agricultural planners. Future work could be focused on including real-time IoT sensor data, satellite images, or climate change scenarios to improve the prediction accuracy of the yield and take sustainable agriculture to the next level in proactive decision-making.

REFERENCES

[1] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review**.** *Sensors*, 18(8), 2674. https://doi.org/10.3390/s18082674

[2] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, 21(11), 3758. https://doi.org/10.3390/s21113758

[3] Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms. *Agronomy*, 10(7), 1046. https://doi.org/10.3390/agronomy10071046

[4] Khaki, S., Wang, L., & Archontoulis, S. V. (2019). A CNN–RNN Framework for Crop Yield Prediction**.** *Frontiers in Plant Science*, 10, 1750. https://doi.org/10.3389/fpls.2019.01750

[5] Khaki, S., & Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks. *Frontiers in Plant Science*, 10, 621. https://doi.org/10.3389/fpls.2019.00621

[6] Nosratabadi, S., et al. (2020). Hybrid Machine Learning Models for Crop Yield Prediction**.** *arXiv preprint*. [arXiv:2005.04155]

[7] Shawon, S. M., Barua Ema, F., Mahi, A. K., & Raihan, M. M. S. (2023). Crop Yield Prediction: Robust Machine Learning Approaches for Precision Agriculture. *2023 ICCIT*. https://doi.org/10.1109/ICCIT60459.2023.10441634

[8] Abbas, R., et al. (2024). AI Can Empower Agriculture for Global Food Security: Challenges and Opportunities. *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2024.1328530 Frontiers

[9] Spring framework study: (2024). A Proposed Framework for Crop Yield Prediction Using Hybrid Featur *Neural Computing and Applications*. https://doi.org/10.1007/s00521-024-10226-x

[10] Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn Yield Prediction and Uncertainty Analysis Based on Remotely Sensed Variables Using a Bayesian Neural Network Approach**.** *Remote Sensing of Environment*, 259, 112408. https://doi.org/10.1016/j.rse.2021.112408

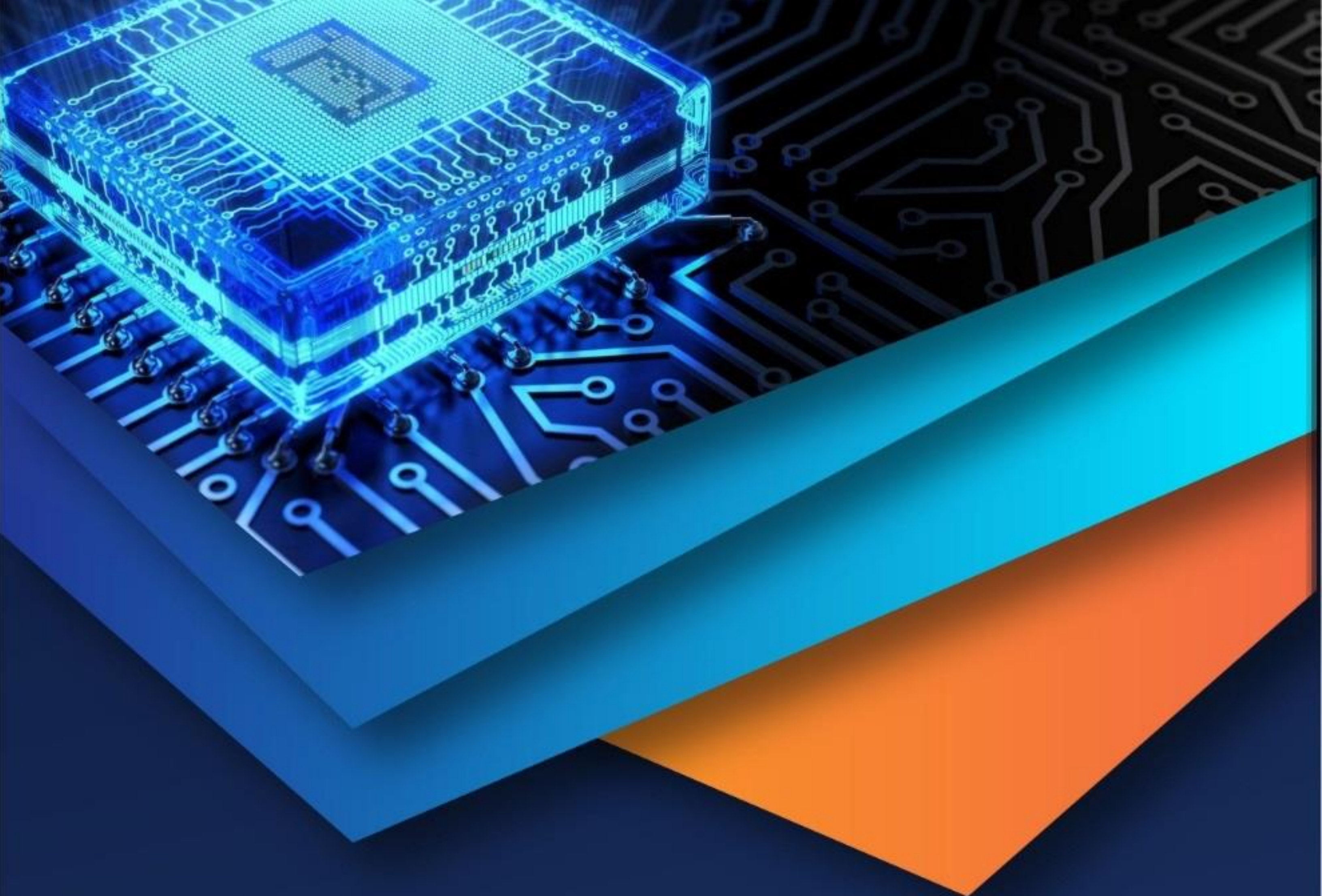
[11] Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *NeurIPS 2017*. https://doi.org/10.48550/arXiv.1705.07874

[12] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System**.** *Proceedings of KDD 2016*, 785–794. https://doi.org/10.1145/2939672.2939785

[13] Gripsy, J. V., Sheela Selvakumari, N. A., Sahul Hameed, S., & Jamila Begam, M. (2024). Drowsiness Detection in Drivers: A Machine Learning Approach Using Hough Circle Classification Algorithm for Eye Retina Images. In *Applied Data Science and Smart Systems* (pp. 202–208). CRC Press. https://doi.org/10.1201/9781003471059-28

[14] Gripsy, J. V. (2020). Biological Software for Recognition of Specific Regions in Organisms. *Bioscience Biotechnology Research Communications*, 13(1), 340–344. https://doi.org/10.21786/bbrc/13.1/54

[15] Tan, G., et al. (2022). Winter Wheat Yield Prediction Using CNNs Incorporating Environmental and Phenological Data**.** *arXiv preprint*. [arXiv:2105.01282]



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY