iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.83095

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Preliminary Study on Turmeric Yield Prediction Using Machine Learning with Soil and Weather Parameters in Erode District

Ms.C.Gohila[1], Dr. C. Premavathi[2]
[1]Research Scholar, [2]Associate Professor, Department of Computer Science, Vellalar College for Women (Autonomous), Thindal, Erode.

Abstract: The rhizome of Turmeric is a major agricultural commodity in Erode, contributing significantly to trade and production, as the region accounts for nearly 60 percent of India's turmeric output and is widely known as the "Turmeric Capital of the World".
However, while such importance is well acknowledged, farmers in Erode have not yet to switch from traditional guess work and subjective evaluation of future crop yield. The purpose of this preliminary study is to determine whether there is a way to predict turmeric yield through analysis of data that is easily obtainable and available to framers and agricultural experts working in Erode district. For this purpose, a limited yet comprehensive database was compiled containing values for eight major factors influencing turmeric yield: soil pH, soil nitrogen, phosphorus and potassium levels, total seasonal rainfall, average temperature, relative humidity and the final crop yield measured in tons per hectare. Subsequently, two machine learning models were trained and compared: Linear regression, selected as a baseline model for its simplicity and interpretability and random forest, chosen for its robustness and higher predictive capability. The results indicated that random forest significantly outperformed linear regression, achieving lower RMSE and MAE along with a higher R^2 score on test dataset. Among all input variables, rainfall and potassium availability were identified as the most influential factors affecting turmeric yield.
Keywords: Turmeric Yield Prediction-Linear Regression-Random Forest-Soil Parameters-Weather Data-Erode

I. INTRODUCTION

Erode plays a crucial role in India's spice economy, particularly through the activities of the erode turmeric merchants' market, one of the world's largest trading hubs for turmeric. The market attracts farmers, traders, and brokers who collectively influence the pricing of a crop widely used in Indian cuisine and traditional medicine. Erode contributes nearly 60% of Indian annual turmeric production, and the region's turmeric variety, recognized for its high Curcumin content often exceeding 3.5%, commands premium value in both domestic and export markets.

Despite its economic significance, turmeric yield estimation in the region is still largely based on traditional practices such as former experience, field observation, and approximate judgment. Official yield assessment conducted by the state agriculture department primarily rely on crop-cutting results, market supply forecasts are often inaccurate, creating challenges in post-harvest planning, inventory management and agriculture price policy formulation. Machine learning provides an alternative and data-driven solution for crop yield prediction. Instead of depending solely on manual field surveys, machine learning models can utilize existing datasets such as soil test reports, meteorological observation from nearby IMD station and frame management records to generate yield forecasts with reduces dependence on costly satellite imaginary and advanced sensing technologies, requiring mainly structured agricultural and environmental data along with suitable predictive models. This study presents a preliminary investigation based on a small, locally collected data set from Erode district. The methodology intentionally focuses on simplicity and practicality by employing two machine learning models: Linear Regression, selected for its simplicity and interpretability and Random Forest, chosen for its stronger predictive capability while remaining sufficiently interpretable. The primary objective is to demonstrate the feasibility and practical applicability of crop yield production within Erode's agro ecological conditions using accessible local data sources.

II. RELATED WORK

A. *Machine Learning in Crop Yield Forecasting*

The use of data-driven approaches for crop yield prediction has existed for several decades through traditional statistical regression techniques. However, significant advancement in computational power and the increasing availability of digital agricultural datasets over the past decade have greatly improved the scale and sophistication of predictive models [4]. Studies such as those by Jeong et al.

Demonstrator that random forest models could predictively corn yield across American counties with high accuracy, outperforming conventional regression approaches, particularly when weather-related anomalies where include as input variable [5]. Similarly, Gandhi and Armstrong applied multiple machine learning algorithms for crop yield production across various Indian status and reported that ensemble-based methods constantly performed better than individual predictive models [6] for rise yield predication, Cao et al, showed that XGBoost with optimized hyper parameters reduced prediction error compared to random forest models in several Chinese provinces [7]. Although XGBoost is not evaluated in the present study, these finding highlights effectiveness of ensemble learning approaches in handling the non-linear and interactive characteristics commonly observed in agriculture datasets, including those associated with turmeric cultivation in Erode

B. *Linear Regression as a Baseline*

Despite the increasing probability of advanced machine learning techniques linear regression remains and imported benchmark model in agricultural studies because of its transparency computational efficiency, and ease of interpretation. Sehgal et al. Utilized linear regression as a baseline model for what yield prediction in Haryana and observed that all through non-linear models achieved higher production accuracy, the linear model provided meaningful insights through interpretable feature coefficient that wear useful for agricultural stakeholders [8].

For this reason, linear regression is included in the present study as a reference model to better understand the strengths and limitations of simple linear relationships in explaining turmeric yield variation.

C. *Soil and weather in yield modeling*

Existing literature consistently identifies rainfall as one of the most influential climatic variables affecting crop productivity in tropical agricultural system [9].

In turmeric specific studies, Verma et al. Reported that potassium availability, monsoon onset timing and, minimum temperature collectively explained a substantial portion of yield variation in single, Maharashtra [10]. Based on those findings the present study selected soil pH, nitrogen, phosphorus, potassium, rainfall, temperature, and humidity as input variables for model development. These parameters were chosen to reflect the ago-climatic conditions of Erode, which is characterized by red lateritic soils and a bimodal monsoon system.

D. *Research gap addressed by the Study*

Although machine learning applications in crop yield prediction have expanded rapidly, studies specifically focused on turmeric yield prediction in Erode remain extremely limited [11].

The district possesses unique agricultural characteristics, including loamy soils, bimodal rainfall patterns and cultivar-specific cultivation practices, making it difficult to directly apply models developed for other turmeric growing regions without proper local validation these steady addresses. That gap by establishing preliminary machine learning framework based on locally collected agricultural and environmental data from the Erode ago ecological zone.

III. STUDY AREA DESCRIPTION

Erode district is in the western region of Tamil Nadu, situated the Nilgiri hills to the west and the border Cauvery delta planes to the east. The district extends between Latitudes 10°36'N and 11°58'N and longitudes 76°49'E and 77°58'E, covering an area of approximately 5,714 km^2. The Cauvery River and Bhavani River flow through the district, providing essential irrigation support for agricultural activities, particularly during dry periods. The climate is classified as tropical semi-arid with annual rainfall ranging between 600 and 900mm, primarily received through south-west monsoon (June -September) and northeast monsoon (October-December). The present study focuses on the taluks of Erode, Bhavani and Gobi chettipalayam, which collectively represents the major turmeric growing regions of the district.

The soils in those areas predominantly red loamy in nature, with slight acidic to natural pH levels ranging from 5.5 to 7.2 moderate organic carbon content and variable nutrient composition influenced by cultivation history and management practices. The primary cultivar of turmeric grown in the region is “Erode local”, also known as “Duggirala”, which is highly valued for its elevated Curcumin content and superior drying properties. Planning is typically undertaken during April-May or July-August with harvesting generally completed after a growth period of 8 to 10 month. Figure 1 presents the study area map of Erode districts, showing the special extent of the selected taluks (Erode, Bhavani and Gobi chettipalayam) along with the major river system that support agricultural activity in the region.

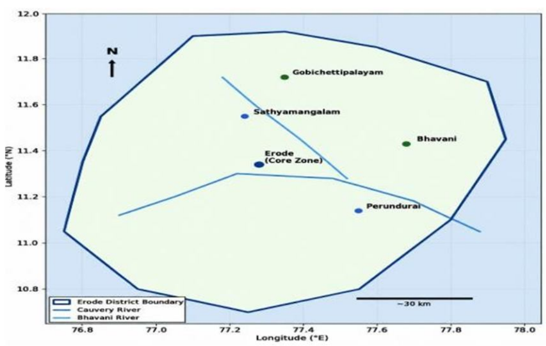


Fig.1 Study Area Map Erode District, Tamil Nadu, India (Major Turmeric-Growing Taluks and River Systems)

IV. EXISTING METHODOLOGY

A. Data collection

1) Soil parameter data

Soil-related data were obtained from two complementary source, the first source was the soil health Card (SHC) Portal Maintained by the government of India Department of Agriculture and farmers welfare, which prove guide-based measurements soil pH and key macronutrients, including (N), phosphorus (P_2O_5)and potassium (K_2O), for agricultural lands in Erode [12]. The second source involved structured field survives conducted across 47 agricultural plots in the three selected Talukas Where farmers provided Recent soil test reports along with corresponding crop production records soil moisture information was Supplemented using India Meteorological department (IMD) Agro-metrological Bulletins Validated Through direct field observation.

2) Weather Data

Meteorological variable wear collected from two Primary sources. The first was the India Meteorological Department IMD Agro-metrological advisory service, which provides station level daily records of rainfall temperature and relative humidity for station located in and around Erode. The second source was publicly available Kaggle agricultural weather data set for Tamil Nadu, which was used to fill missing historical value and ensuring continuity in long term records for model development. Daily weather observation where aggregated into season summaries, including total rainfall (mm), mean maximum and a minimum temperature (^{o}C) and average relative humidity(%). The season and crop compiled three primary sources-The season and crop reports published by the Department of economics and statistics, Tamil Nadu [13]; Production registrar maintained by the Erode turmeric farm cooperative society and the structured former surveys described earlier. The final merged dataset consisted of 80 plot-season observations spanning 10 crop season (2015-2024). This dataset represents a relatively small but carefully validated sample making it suitable for a preliminary machine learning analysis of turmeric yield prediction in Erode. Figure 2 presents the structure of the dataset used for turmeric yield prediction in Erode district. The dataset integrates soil parameters such as pH, nitrogen, phosphorus, and potassium with weather variables including rainfall, temperature, and humidity, along with the corresponding turmeric yield values. This combined dataset forms the foundation for training and evaluating the machine learning models used in the study.

3) Data preprocessing

Before model training significant effort was devoted to data cleaning and preparation as the quality of input data is considered more critical than model complexity in achieving reliable predictions. Missing values, which accounted for approximately 4% of the dataset were addressed using K-Nearest neighbor (KNN) imputation with k=5.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

This method estimates missing values by identifying the five most similar observations in the dataset and using their corresponding values, providing a more robust alternative to simple mean imputation [14]. All continues variables were standardized using z-score normalization, where features were transformed by subtracting the mean and dividing by the standard deviation computed from the training dataset only.

The step ensured that variables measured on different scales such as rainfall in millimeter and soil pH did not disproportionality influence the model. Outlier detection was performed using the inter quartile range (IQR) method. Six extreme observations, associated with documented flood event in 2019 and a severe drought in 2023, were retired as valid data those were considered genuine climate induced variations rather than errors, and excluding them would have reduced the model's ability to learn from real-world extremes affecting turmeric production.

Finally, the dataset was split into training and testing sets using 70:30 ratios, resulting in 56 training samples and 24 testing samples. A chronological split strategy was applied, where earlier seasons were used for training and later seasons for testing ensuring a realistic forecasting setup that reflects how future yield prediction would operate in practice within Erode.

Sample ID	pH at 25°C	Total Hardness (mg/L)	Calcium Hardness (mg/L)	Magnesium Hardness (mg/L)	Alkalinity (mg/L)	Chloride (mg/L)	TDS (mg/L)	EC (dS/m)
S1	6.9	247	78	169	162	15.4	67	0.4
S2	7.2	287	92	195	180	17	74	0.5
S3	6.7	214	67	147	148	11.6	54	0.4
S4	8.1	375	128	247	224	23	112	0.7
S5	7.5	299	98	201	189	18	81	0.5
S6	7.1	331	115	216	205	20.4	97	0.6
S7	8.2	412	142	270	250	26	124	0.8
S8	6.6	265	86	179	153	14.2	63	0.4
S9	7.4	276	90	186	171	16	71	0.4

Figure 2: Sample Dataset Structure Soil, Weather & Turmeric Yield Parameters (Erode District)

B. Machine Learning Models

1) Linear Regression

Liner Regression is one of the most fundamental predictive techniques in statistics. It estimates the best-fitting straight line through the data by minimizing the sum of squared errors between observed and predicted values [15].

The method assumes a linear and additive relationship between inputs features and output, which is a simplification but often effective in capturing dominant trends.

In this study, Linear Regression was implemented using the Scikit-learn Library (v1.3.0). The model required no complex hyper parameter tuning and was trained using all eight input features, then evaluated on the held-out test dataset. Although it was expected to perform weaker than ensemble methods, it was intentionally included as a transparent baseline model that can be easily replicated using basic analytical tools, including spreadsheet-based approaches.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

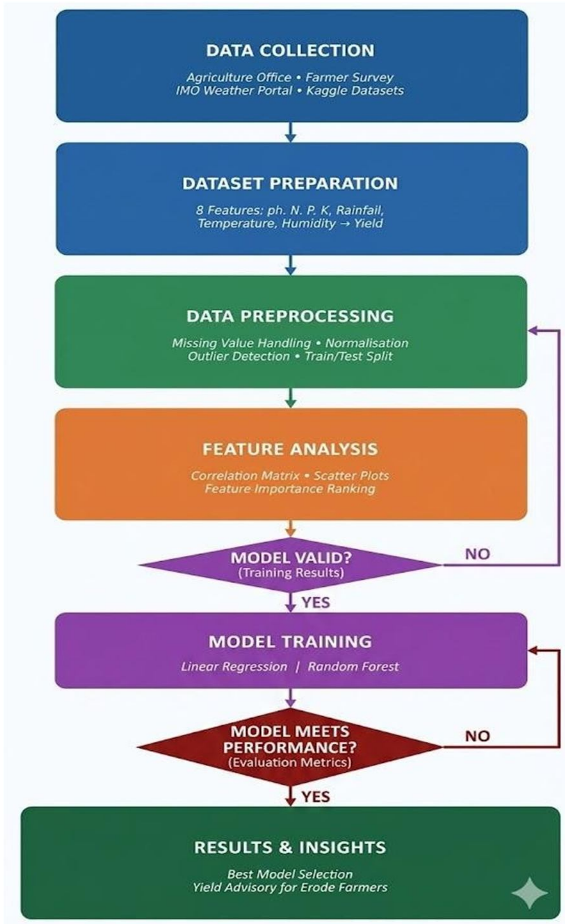


Figure 3: Research Methodology Flow Chart-From Data Collection Yield Prediction

2) Random Forest

Random Forest, proposed by Leo Breiman in 2001[16], is an ensemble learning method that constructs Multiple decision trees using different random subsets of the data and aggregates their predictions through averaging. This "wisdom of the crowd" approach significantly improves robustness compared to a single decision tree, which is often prone to over fitting. The built-in randomness in both data sampling and feature selection ensures diversity among trees, enabling the model to capture complex, non-linear relationships effectively. Hyper parameter tuning for the random forest model was carried out using 5-fold cross-validation combined with grid search across different parameter setting: number of trees (n estimators) = {100,200,300}, maximum depth (max, depth) = {5,10,None} and minimum sample per leaf (min sample leaf) = {1,2,5}. The best performing configuration was found to be n estimators = 200, max depth = 10 and min sample leaf = 2. Feature importance was calculated using mean decrease in impurity (MDI), which measures the contribution of each feature in reducing overall prediction error across all trees in the ensemble. Evaluation metrics model performance was assessed using three widely accepted metrics that collectively provide x comprehensive evaluation of predictive accuracy:

- Root mean Square Error (RMSE): Measures the average magnitude of predictive error, expressed in tons per hectare. It penalizes large errors more heavily, making it suitable for evaluating extreme deviations.
- Mean Absolute Error (MAE): Represents the average absolute difference between predicted and actual yield values. It is more interpretable, indicating typical deviation in t/ha
- Coefficient of Determination (R^2): Indicates the proportion of variance in yield explained by the model. For example, an $R^2$of 0.94 implies that 94% of yield variability is captured by the model.
-

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

V. RESULT AND ANALYSIS

A. Yield Trend Analysis (2015-2024)

Before applying predictive modeling, the temporal variation in yield across Erode was examined. The trend shows an overall upward trajectory in turmeric productivity over the study period, despite inter-annual fluctuations driven by monsoon variability. A notable decline was observed during the drought year of 2019 (6.4t/ha), while subsequent years showed recovery and steady improvement. By 2024, average yield reached approximately 9.5 t/ha, respectively a 32% increase compared to 2015 (7.2 t/ha).this improvement is likely associated with enhanced agronomic practices, increased adoption of soil health management under the soil health card scheme and gradual expansion of irrigation infrastructure such as drip systems.

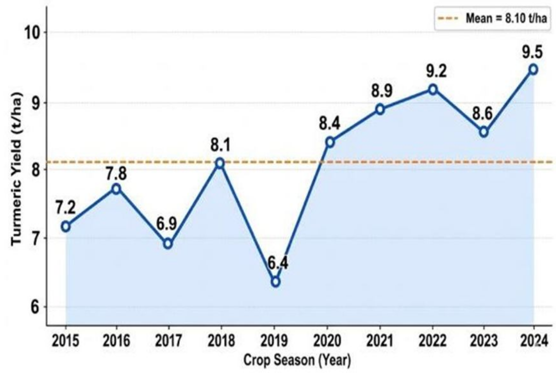


Fig 4: Turmeric Yield Trending Erode District(2015-2024) Dashed Line Indicates 10-year mean(8.00t/ha)

B. Correlation Analysis

Figure 5 presents the correlation matrix for all eight variables in the dataset. The value in each cell represent person's correlation coefficient, where values close to +1 indicate a strong positive relationship, values close to -1 indicate a strong negative relationship, and values near 0 indicate weak or no linear relationship. Several clear patterns emerge from the analysis. Rainfall exhibits the strongest positive correlation with yield (r = 0.82), which aligns well with agronomic expectations, as turmeric is highly dependent on adequate soil moisture, particularly during the rhizome development stage. Potassium also shows a strong positive correlation (r = 0.78), reflecting its essential role in rhizome bulking and Curcumin formation. Nitrogen is another important contributor (r = 0.73), supporting overall vegetative growth and biomass development. In contrast, temperature shows a moderate negative correlation with yield (r = -0.61), suggesting that hotter conditions often associated with moisture stress, tend to reduce productivity. Overall, these correlation patens provide strong empirical support for the selected feature set and confirm their relevance in explaining yield variability within Erode.

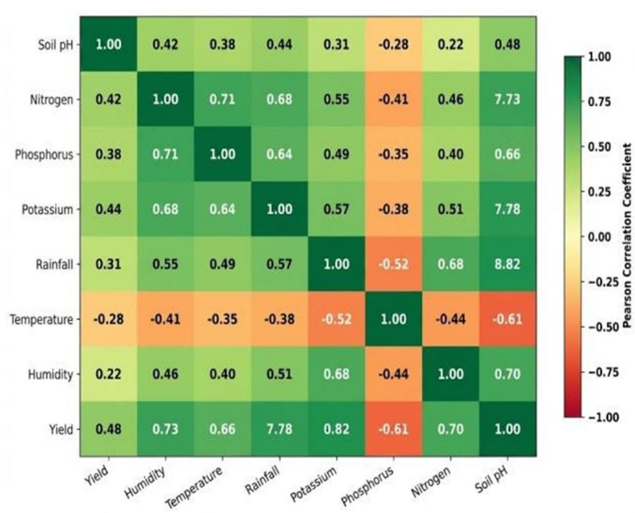


Figure 5: Correlation Matrix-soil, weather and yield variables (Erode Data; Red = Negative, Green= Positive Correlation)

C. Scatter Plot Analysis-Key Predictors vs. Yield

Figure 6 illustrates the two most influential predictors of yield in the dataset: rainfall and potassium. Each point represents a single plot-season observation of turmeric cultivation. Both scatter plots exhibit clear positive trends, indicating that higher rainfall and greater potassium availability are consistently associated with increased yield.

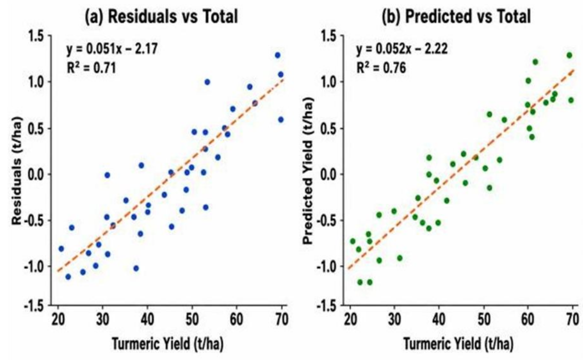


Figure 6 : Scatter Plots of Key Predictors. Turmeric yield (a)Rainfall vs Yield (b) Potassium vs Yield

The fitted trend lines further confirm this relationship, with rainfall showing an R^2 of 0.67 and potassium showing an R^2 of 0.071 while both variables demonstrate strong individual influence on yield variation, neither is sufficient on its own to fully capture the complexity of yield dynamics. This highlights the necessity of using a multi-variable machine learning approach that integrates soil and climatic factors for more accurate predictions within Erode.

D. Actual vs Predicted Yield

Figure 7 compares the predicted yield from each model against the actual observed yields in the test dataset. In an ideal scenario, all points would lie exactly on the 45-degree diagonal line, where predicted values perfectly match actual yields. The random forest model shows a strong alignment with this ideal pattern, with most points closely clustered around the diagonal across the full yield range. This indicates high predictive accuracy and stable performance for turmeric yield estimation.

In contrast, the linear regression model also follows the general trends but exhibits greater dispersion from the diagonal line. The deviation is especially noticeable at higher yield values, where the model tends to underestimate actual productivity. This systematic under prediction reflects a known limitation of linear models, which often fail to capture complex interaction effects between variables such as soil nutrients and rainfall that contribute to exceptional yield conditions within Erode.

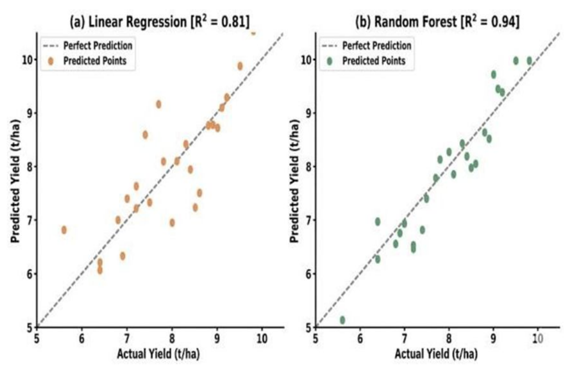


Figure 7: Actual vs Predicted Turmeric Yield (a) Linear Regression,(b) Random Forest [Points near Diagonal = Beter Accuracy]

E. Model Performance Comparison

Table 1 compares the performance of the Linear Regression and Random Forest models for turmeric yield prediction using RMSE, MAE, R^2 score and training time. The Random Forest model achieved better predictive accuracy with lower RMSE (0.34 t/ha) and MAE (0.27 t/ha) values, along with a higher R^2 score (0.94), compared to Linear Regression. Although Random Forest required slightly higher training time, the results indicate that it is more effective in capturing the complex relationships between soil, weather, and turmeric yield parameters.

Model	RMSE(t/ha)	MAE(t/ha)	R^2 Score	Training Time
Linear Regression	0.68	0.54	0.81	<1sec
Random Forest	0.34	0.27	0.94	~45sec

Table 1: Performance Comparison Liner Regression vs Random Forest (test set)

Figure 8 indicates that the Random Forest model provides better predictive performance than Linear Regression due to its lower error values and higher R^2 score.

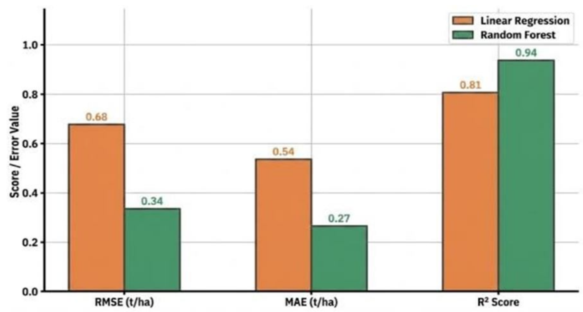


Figure 8: Performance Comparison Bar Chart RMSE,MAE and R2 for Both Models (Lower RMSE/MAE and higher R2 Indicate better performance)

F. *Feature Importance Analysis*

One of the key advantages of the random forest model is its ability to rank input variable based on their contribution to prediction accuracy (Figure 9). This ranking is typically derived by measuring the increase in prediction error each feature is randomly shuffled, thereby breaking its relationship with yield. Feature that causes a large increase in error are considered more important for predicting.

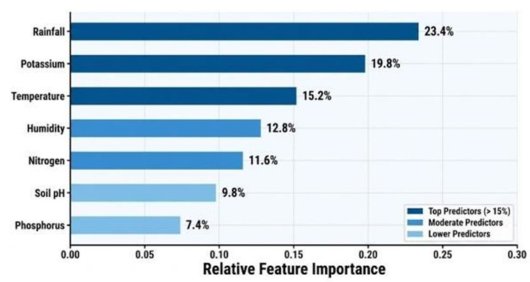


Figure 9: Feature Importance Ranking from Random Forest Model Darker Blue = Higher Importance for Turmeric Yield Prediction

Rainfall emerges as the most influential variable, contributing 23.4% of the overall importance, reinforcing the strong dependence of turmeric on adequate water availability. Potassium ranks second with 18.7% important, highlighting is essential role in rhizome development and crop quality. Temperature follows at 15.2%, reflecting the impact of thermal stress during the growing season. Humidity (12.8%) and Nitrogen (11.6%) also play meaningful roles in influencing yield variability. Although soil pH shows a comparatively lower concentration (9.8%), it remains an important indirect factor because it regulates nutrient availability in the soil. In region such as Erode, slightly acidic conditions can reduce the accessibility of key nutrients like phosphorus and potassium, thereby indirectly affecting overall crop productivity.

VI. DISCUSSION

The performance difference between linear regression and random forest reflects the inherent complexity of turmeric yield formation. Regression assumes a strictly additive and independent relationship between predictors and yield, limiting its ability to capture real-world agricultural interactions, whereas random forest effectively models non-linear relationships and feature interactions through ensemble decision trees. Although linear regression achieved a reasonable $R^2$of 0.81, random forest significance improved predictive accuracy with an R^2 of 0.94, while the reduction in RMSE from 0.68 to 0.34t/ha further confirms its superiority for practical yield forecasting in Erode. The feature important result align strongly with agronomic understanding, where rainfall during the Southwest monsoon period (June-September) plays a critical role in rhizome development occurring between the fourth and seventh months after planting, making it a dominant yield driver, while potassium emerges as another key factor influencing rhizome expansion, water-use efficiency and Curcumin synthesis. The interaction between adequate rainfall and sufficient potassium explains a major portion of yield variability, particularly in the red loamy soils of Erode where potassium loss during heavy rainfall can reduce nutrient availability during critical growth stage. Therefore, spilt application of potassium fertilizer around the rhizome initiation stage an important management practice, which is implicitly supported by the model findings.

VII. CONCLUSION

This study addresses a practical question: whether turmeric yield in Erode can be predicated using a small, locally collected dataset combined with basic machine learning models. The findings provide a qualified but meaningful affirmative answer. The random forest model, train on 56 training observations and eight input feature, achieved strong predictive performance with an $R^2$of 0.94 and an RMSE of 0.34t/ha. These results indicate that even with limited data reliable yield forecasting is achievable. The linear regression model, while less accurate (R^2=0.81),still serves as a useful and interpretable baseline that can be applied in low-resource or extension-level contexts.

Rainfall and potassium availability emerged as the most influential predictors of yield, aligning strongly with established agronomic understanding and reinforcing their central role in turmeric production system. These findings have direct implications for irrigation scheduling and fertilizer management in the Erode region. This study demonstrates that effective yield predictions do not necessarily require advanced sensing infrastructure or large-scale datasets. With careful data collection, appropriate pre-processing and suitable modeling techniques, practical forecasting system can be developed even in data-limited agriculture environments. This work is intended as an initial step toward more advanced, scalable and farmer-centric decision-supported tools.

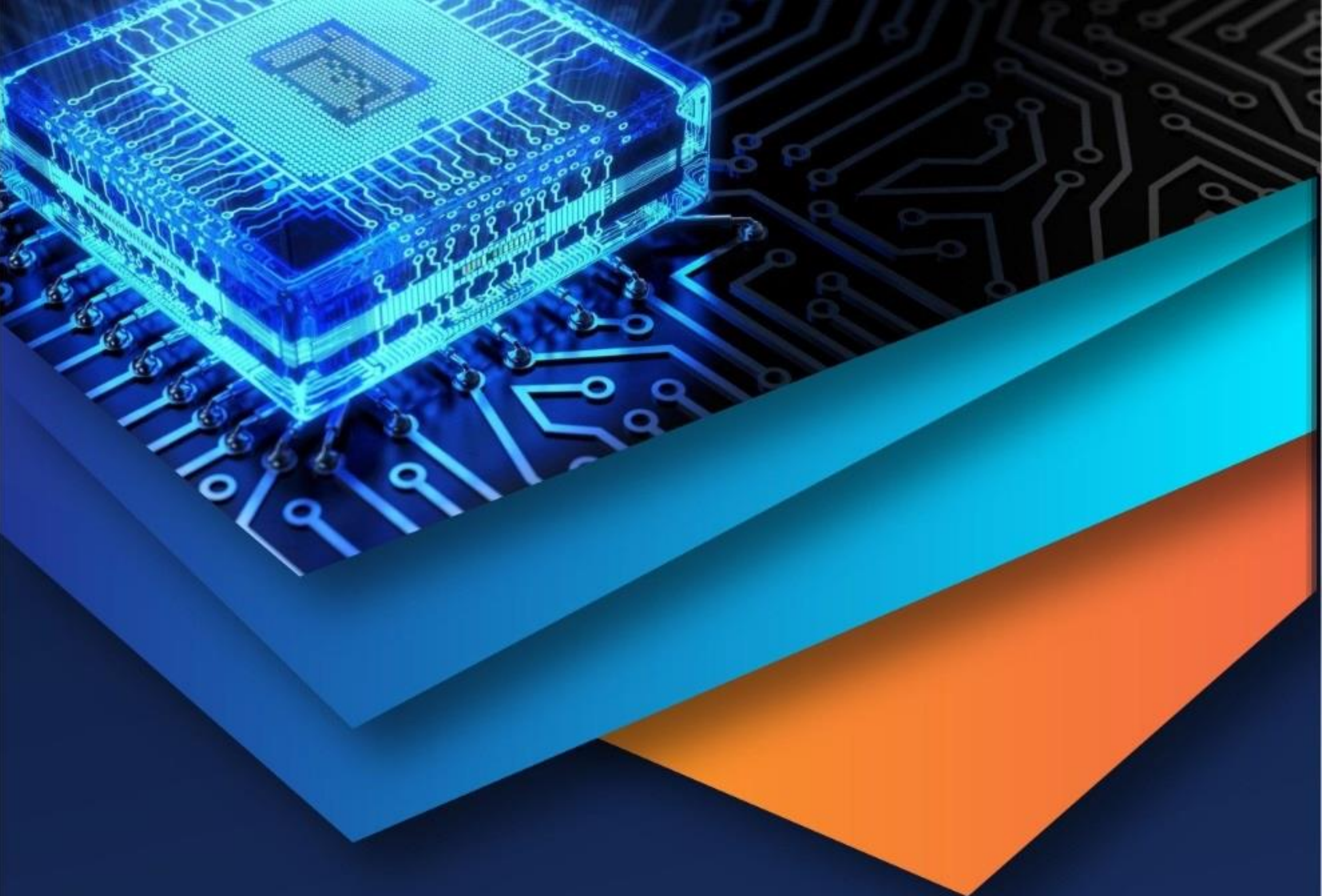
VIII. LIMITATIONS AND FUTURE SCOPE

The present study is subject to several limitations that also define important directions for future research. The dataset is relatively small, comprising only 80 observations across 10 seasons, which limits exposure to extreme climate variability and may reduce model robustness under rare or unusual conditions. In addition, key agricultural management variables such as irrigation scheduling, fertilizer application rates, and pest or disease pressure were not included, thereby restricting the explanatory completeness of the model. The study is further constrained by its single-district focus on Erode, making it necessary to recalibrate the model before applying it to other geographic regions. Moreover, multiple turmeric cultivars were aggregated into a single category, which may reduce prediction accuracy due to cultivar specific yield differences, while soil data derived from soil health card (SHC) grid-level sources may not fully capture fine-scale, plot-level variability to address these limitations, future work should focus on expanding both dataset size and feature richness by increasing farmer participation and adopting structured field data collection system. Integration of remote sensing data, particularly satellite-derived vegetation indices such as NDVI from sentinel-2 imagery and complementary sentine-1 SAR data during cloudy conditions, can significantly improve temporal and spatial crop monitoring. Further improvements can be achieved through advanced machine leaning models such as XG boost, Light GBM and hybrid CNN-LSTM architecture to better capture non liner and temporal relationships. Ultimately, future research should aim to develop a framer-facing decision support system in the form of a mobile application for turmeric growers in Erode, providing yield prediction along with interpretable, SHAP-based insights to support trust, usability, and practical agriculture decision-making.

REFERENCES

[1] Directorate of Economics and Statistics, Tamil Nadu. (2024). Season and Crop Report 2023–2024. Government of Tamil Nadu, Chennai.
[2] Ministry of Agriculture and Farmers Welfare. (2022). Crop Cutting Experiments: Methodology and Limitations. Government of India, New Delhi.
[3] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis D. (2018). Machine learning in agriculture: A review. Sensors,18(8), 2674.
[4] Shook J., Gangopadhyay T., Wu, L., Ganapathy Subramanian B., Sarkar S., & Singh A. K. (2021). Crop yield prediction integrating genotype and weather variables using deep learning PLoS ONE,16(6), e0252402.
[5] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., & Kim, S. H. (2022). Random forests for global and regional crops yield predictions. PLoS ONE, 11(6), e0156571.
[6] Gandhi, N., & Armstrong, L. J. (2020). Applying data mining techniques to predict crop yield in Indian agriculture in proceedings of the IEEE International Conference on Advances in Computer Applications (ICACA) (pp. 95–100).

[7] Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J., & Li, Z. (2021). Identifying contributions of multi-source data for winter wheat yield prediction in China. Remote Sensing,12(5), 750.
[8] [8] Sehgal, V. K., Bhatt, D., & Kumar, S. (2021). Comparative analysis of machine learning methods for crop yield prediction in semi-arid regions of India. Journal of the Indian Society of Remote Sensing, 49(3), 617–629.
[9] Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology, 150(11), 1443–1452.
[10] Verma, A., Sharma, R., & Patidar, N. (2022). Application of Random Forest for turmeric yield prediction: A study from Sangli region. Indian Journal of Agricultural Sciences, 92(4), 487–492.
[11] Padmavathi, K & Prabavathi, M. (2021). Machine learning models for predicting spice crop yield: A review. International Journal of Intelligent Systems and Applications, 13(1), 43–52.
[12] Soil Health Card Scheme. (2024). SHC Portal – Soil Test Data for Erode District. Department of Agriculture and Farmers Welfare, Government of India. soilhealth.dac.gov.in
[13] Department of Economics and Statistics, Tamil Nadu. (2023). District-wise Crop Production Statistics, Erode. Government of Tamil Nadu.
[14] Zhang, S. (2012). Nearest neighbor selection for iteratively KNN imputation. Journal of Systems and Software, 85(11), 2541–2552.
[15] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to Linear Regression Analysis (6th ed.). Wiley.
[16] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
[17] TNAU. (2023). Package of Practices for Spice Crops: Turmeric. Tamil Nadu Agricultural University, Coimbatore
[18] Sentinel Hub. (2024). Sentinel-1 SAR and Sentinel-2 NDVI for Agricultural Monitoring. Copernicus Open Access Hub.
[19] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 30, 4765–4774.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY