ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 15, Issue, 04, pp. 68158-68163, April, 2025
https://doi.org/10.37118/ijdr.29445.04.2025

RESEARCH ARTICLE OPEN ACCESS

PREDICTIVE MODELING OF CROP YIELD USING RANDOM FOREST: A COMPREHENSIVE ANALYSIS OF METEOROLOGICAL AND AGRICULTURAL FACTORS

Sravankumar P[1], Pooja[1] and Ritu R Saxena[2], Roop Shikha Agrawal[3], Shilpi Verma[4] and Ravi R Saxena[1,*]

[1]Department of Agricultural Statistics and Computer Science, IGKV, Raipur (CG); [2]Department of Genetics and Plant Breeding, IGKV, Raipur (CG); [3]Department of Botany, Govt. J. Y. Chhattisgarh College Raipur (CG); [4]Department of Foods and Nutrition, KVK Neemuch

ARTICLE INFO

Article History:
Received 17th January, 2025
Received in revised form
09th February, 2025
Accepted 14th March, 2025
Published online 28th April, 2025

Key Words:

Predicting Crop yield, Machine learning techniques, Random Forest, R-package.

***Correspondingauthor:** **Ravi R Saxena,**

ABSTRACT

Predicting crop yields is a crucial agricultural undertaking that helps with improved resource management and planning. Because machine learning can scan enormous datasets to find hidden patterns and provide reliable forecasts, it has become a useful tool for addressing the issues given by weather variability in agricultural forecasting. In this work, we assess the effectiveness of several machine learning models, including Multiple Linear Regression, K-Nearest Neighbors (KNN), Random Forest Algorithm, and Support Vector Machine (SVM), for predicting rice crop yield in several districts within the Indian state of Chhattisgarh. Key performance metrics were used to evaluate the models. Among the models, Random Forest Algorithm outperformed the others with an R^2 score of 0.9476, MSE of 0.02254, RMSE of 0.15016, and MAE of 0.0036, indicating a high level of accuracy and precision and low error rates in predicting the crop yield, while simpler models such as Multiple Linear Regression showed comparatively lower predictive power. This analysis highlights the importance of selecting appropriate machine learning algorithms for enhancing the precision of crop yieldforecasts. The complete analysis was done by using open-source software namely R-Software.

Citation: Sravankumar P, Pooja and Ritu R Saxena, Roop Shikha Agrawal, Shilpi Verma and Ravi R Saxena, 2025. *"*Predictive modeling of crop yield using random forest: A Comprehensive analysis of meteorological and Agricultural factors*". International Journal of Development Research*, 15, (04), 68158-68163.

INTRODUCTION

Agriculture is inherently sensitive to weather variability, making it one of the most climate-dependent sectors globally. As weather patterns continue to shift due to climate change, the impact on crop yields has become increasingly unpredictableposing significant challenges to food security and agricultural sustainability.An in-depth investigation of the intricate interactions between weather variables including temperature, humidity, wind speed, and rainfall and crop yield is necessary to lessen the negative consequences of climate variability on agriculture. In a climate-sensitive agricultural landscape, accurately Predicting crop yields is a crucial undertaking in agriculture that helps with improved resource management and planning. Because machine learning can scan enormous datasets to find hidden patterns and provide reliable forecasts, it has become a useful tool for addressing the issues given by weather variability in agricultural forecasting. In this work, we assess the effectiveness of several machine learning models, including Multiple Linear Regression, K-Nearest Neighbors (KNN), Random Forest Algorithm, and Support Vector Machine (SVM), for predicting rice crop yield in several districts within the Indian state of Chhattisgarh. Key performance metrics were used to evaluate the models, these linkages can provide farmers, policymakers, and stakeholders with theknowledge they need to make well-informed decisions, maximize yield results, and reduce economic risks. In recent years, machine learning has emerged as a cornerstone for data prediction, playing a crucial role in enhancing performance, productivity, and profitability across a wide range of industries. Its ability to process vast amounts of data, identify patterns, and make accurate predictions has led to widespread adoption in sectors such as agriculture, healthcare, finance and manufacturing. By utilizing advanced algorithms and techniques, machine learning enables organizations to make data-driven decisions, optimize processes and forecast outcomes with greater precision, ultimately driving innovation and growth. Machine learning is a practical and powerful approach for yield prediction, utilizing multiple attributes to generate accurate forecasts. As a branch of Artificial Intelligence (AI), machine learning focuses on the ability to learn from data by identifying patterns and trends without the need for explicit programming. By analysinglarge agricultural datasets that includes factors such as weather conditions, soil properties and crop characteristics, machine learning models significantly enhance prediction accuracy. These models provide critical insights that help optimize crop management practices, ultimately improving productivity. This makes machine learning as a crucial tool in modern agriculture, empowering farmers to make informed, data-driven decisions that improve yield outcomes and enhance overall agricultural efficiency. Crop yield prediction is a vital component of modern agriculture, allowing farmers to enhance productivity due to weather variability and contribute to global food security. With the rapid advancement of data-driven technologies, machine learning has become a powerful tool for improving yield prediction accuracy by processing and analyzing complex datasets that would be difficult to manage using traditional methods. This paper investigates the application of popular supervised machine learning techniques Random Forest, Linear Regression and K-Nearest Neighbours (KNN) for predicting crop yield of different districts in Chhattisgarh.

Related Work: Numerous researchers have made various attempts to estimate and predict crop yield and production at both national and state levels. Among those, Breiman 2001, demonstrated that random forest as a powerful tool for forecasting. However, random forest is resistant to overfitting due to the law of large numbers. They become highly accurate classifiers and regressors when appropriate randomness is introduced. Jamuna *et al.,* 2010 research employs the Cotton seed yield classification based on different stages of growth usingmachine learning techniques which includes Naïve Bayes, Decision Tree and Multilayer Perception were applied to train the model using 90 records. Performance was evaluated through 10-fold cross-validation, with both Decision tree and Multilayer perception yielding the same accuracy in classifying seed cotton yield. Gopal and Bhargavi in 2019 used feature subsets chosen by FFS, CBFS, VIF, and RFVarImp to compare the prediction performance of ANN, SVM, KNN, and RF algorithms. Based on a 745-agriculture dataset, 70% was used for training and 30% for testing. The findings indicated that the best feature selection technique was Forward Feature Selection (FFS). For every feature subset, the RF algorithm consistently produced the best accuracy. Josephine *et al*., 2020 focused on forecasting millet yield using the Random Forest Algorithm. They achieved 99.74% accuracy by utilizing statistical datasets that included various factors such as average humidity, minimum and maximum temperature, moisture, soil properties and millet yield. According to research by Klompenburg *et al*. (2020), models with more features don't always estimate crop yields more accurately. It is crucial to test models with both a bigger and smaller collection of attributes in order to determine the best model. Numerous research has used different methods; therefore, it is impossible to determine which model is the best without further investigation. It is apparent, nonetheless, that some machine learning models are employed more frequently than others. The models that are most frequently employed are gradient boosting tree, random forest, neural networks, and linear regression. To determine which machine model provides the best accurate predictions, the majority of studies use a variety of models. Sunil *et al.,* 2022 studies highlight the potential of integrating remote sensing technology with machine learning for accurate sugarcane yield prediction. Their research assesses random forest and second-degree polynomial regression models using vegetation indices from Sentinel-2 satellite data. Focusing on sugarcane fields in Karnataka, the models demonstrated strong prediction capabilities, with RF showing 90.42% accuracy and polynomial regression achieving 88%. Araujo *et al.,* 2023 studied on Machine Learning applications in Agriculture. They employ the PRISMA methodology to explore the usage of Machine Learning in agriculture. The study explores the key applications of Machine Learning areas such as crop, water, soil and animal management, highlighting its pivotal role in transforming traditional agricultural practices. As expected, Random Forest Algorithm emerged as the most prevalent ML algorithm, constituting 19.2% of overall distribution followed by Support Vector Machine of 15.9% because of their efficacy in both classification and regression tasks.Where Gradient Boosting Tree and Convolutional Neural Network adopt rates of 8.3% and 7.3% respectively, highlighting their significance in agriculture sector. Kavitha *et al*., 2023 estimated the crop yield in Rajasthan, India, on five distinct crops using machine learning approaches such as Random Forest, SVM, Gradient Descent, LSTM, and Lasso Regression. With an R^2 of 0.963, 0.035 RMSE, and 0.0251 MAE, the results show that Random Forest outperformed all other applied techniques. Their study aims to apply crop selection techniques to assist farmers in resolving issues related to crop productivity.

MATERIALS AND METHODS

Study area: Chhattisgarh, located in central India, is known as the “Rice Bowl of India” due to its extensive rice fields and significant contribution to rice production. Agriculture is the backbone of the state’s economy, with over 80% of the population dependent on it. Moreover, this study was conducted in the Raipur District of Chhattisgarh as most of the cropping area is covered in this area.

Table 1. Sample of the dataset

District	Year	Area	Production	RH	Tmax	Tmin	WS	RS	Yield
Raipur	1997	918.60	1016.20	77.18	33.84	19.80	5.76	1059.96	1.106
Raipur	1998	902.00	643.00	76.34	35.12	20.00	6.06	727.74	0.713
Raipur	1999	950.21	1378.88	77.11	34.01	19.81	6.18	954.48	1.415
Raipur	2000	889.78	461.63	71.24	35.41	19.32	5.43	754.10	0.519
Raipur	2001	902.62	1120.82	81.80	33.10	18.95	6.78	966.20	1.242
Raipur	2002	897.84	659.62	71.26	35.48	18.62	5.68	725.40	0.735

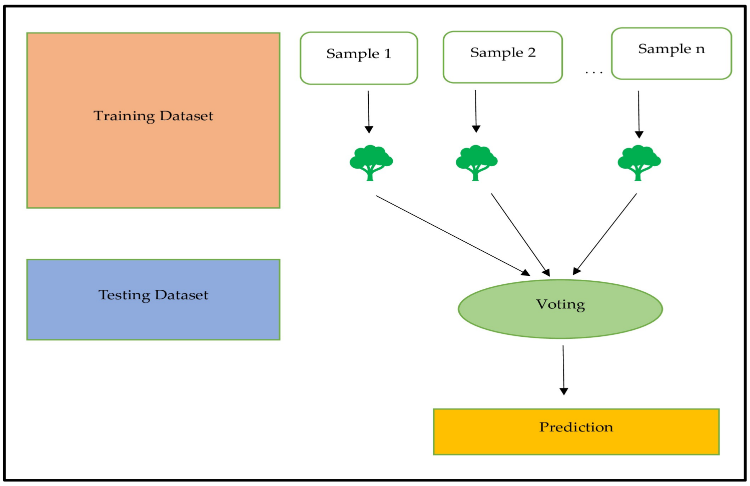

Fig. 1. Random Forest Classifier

Data description: The data covers the state’s most-grown crop concerningarea and production: Rice. Hence, this study uses a random forest algorithm to deal with the prediction of rice crop yield for weather variability in the Raipur District. This study is completely based on secondary data. The dataset was collected for a period of 26 years i.e. from 1997 to 2022 from the source Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, Govt. of Chhattisgarh. The dataset comprises crop parameters of Crop yield (t/ha), Area (1000ha), and Production (1000t) where weather parameters of Relative Humidity (mm), Rainfall (mm), Wind speed (m/s), Maximum and Minimum temperature(C).

Models: In this research, we investigate agricultural yield prediction using the Random Forest model, a machine learning algorithm and compare its effectiveness with other machine learning methods.

Random Forest: Regression and classification problems can be executed with the well-known and potent supervised machine learning method Random Forest. The Random Forest model in regression applies the decision tree algorithm to each of the dataset's many subsets. The outputs of various trees are averaged to provide the final prediction, increasing the predictive accuracy overall. An ensemble of decision tree classifiers is used in the Random Forest approach to improve the performance of the model. It builds decision trees at random from training set instances as a supervised learning method. The outcome is determined by aggregating the predictions made by each decision tree. Random Forest's capacity to manage overfitting—with model accuracy rising as more trees are added—is one of the main factors contributing to its appeal.

Multiple Linear Regression: A statistical technique called multiple linear regression (MLR) is used to model the relationship between a dependent variable (rice crop yield) and several independent factors, including area, production, rainfall, relative humidity, wind speed, and maximum and minimum temperature. This strategy is often used to predict crop yield since it assumes that the dependent and independent variables have a linear relationship.

The following equation represents the MLR model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_{2+\dots+} \beta_n X_n + \epsilon$$

Where:

Y= Crop yield
$X_1, X_2, \dots, X_n$ = Independent variables affecting crop yield
β_0= Intercept
$\beta_1, \beta_{2,}, \dots, \beta_n$= Coefficients that represent the impact of each independent variable on crop yield.
ϵ= Error term

K-Nearest Neighbors (KNN) Algorithm: A straightforward, non-parametric, and intuitive machine learning technique, the KNN algorithm is utilized for both regression and classification applications, including the prediction of agricultural yield. KNN predicts a new observation's result by using the known results of the K closest or most similar data points in the training dataset. Using comparable historical data, KNN can be used to predict crop yields by estimating the yield for a given crop in a given area. A distance metric, such Euclidean distance, is commonly used to quantify the similarity between data points and determine the K closest points. The number of nearest neighbors (K) is a hyperparameter that influences the accuracy of predictions. A small K value may make the model sensitive to noise, while a large K value may smooth out local variations but reduce the model's responsiveness to specific data patterns. A typical range for K is 3 to 10, though this depends on the dataset.

Support Vector Machine Algorithm: A popular family of supervised machine learning techniques for both regression and classification applications are called Support Vector Machine (SVM). SVM can be used in crop yield prediction to simulate intricate correlations between the goal variable, or crop yield, and input data like weather and crop attributes. The goal of the algorithm is to identify the hyperplane that best divides the data points belonging to various classes.

Model performance evaluation and Inter-comparison:

Coefficient of Determination: It is a mathematical metric that indicates that how actual the data is fitted regression line. It is denoted as R^2. Usually, its values varied from 0 to 1 and are indicated as percentage. Performance in relation to the machine learning models used for analysis is called evaluation. Classification report metrics (precision, recall, and F1-score) and Confusion matrix are commonly used to assess the effectiveness of machine learning models. The confusion matrix provides a matrix that displays the frequency of true positives, false positives, true negatives, and false negatives, while the classification report serves as a statistic to assess an algorithm's performance. It provides the model's F1-score, recall, and precision, which may be computed as follows.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ score = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

Where: TP=Total Positives, FP= False Positives, FN=False Negative

RESULTS

This paper reinforces the crop yield prediction with the aid of machine learning techniques.Random Forest model is the technique that results in high accuracy predicted with its yield. The dataset includes the Raipur district of Chhattisgarh plains. The crop data includes District, Year, Area, Production and Yield, and weather parameters like Relative humidity, Rainfall, Windspeed, Maximum and Minimum temperature.Different machine learning models are introduced to predict crop yield. Among those models, Random Forest achieves the highest accuracy of 94% whereasSVM, MLR, and KNN algorithmsattain83%, 68%, and 84% respectively.In terms of MSE, Both Random Forest and Support Vector Machine attainthe lowest 0.002 which makes the smallest average prediction error compared to the other two models where KNN and MLR attain 0.004 and 0.03 respectively. The following plot shows a comparison between Actual Yield and Predicted Yield along with the best-fit line of the Raipur district in the Central Plains of Chhattisgarh visualized with different colours.

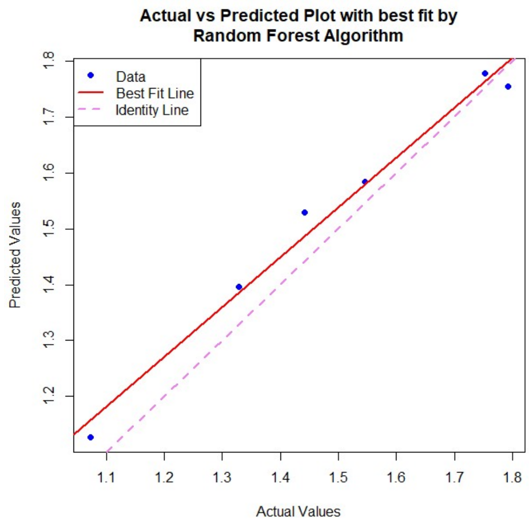

Fig. 2. District-wise plot of Actual and Predicted Yield by Random Forest model

DISCUSSION

From the plotof Fig.2, we can see that the points are clustered along or near the line, indicating that the model's predictors are reasonably accurate for most data points. Points that deviate far from the line suggest where the model overestimated or underestimated the actual yield.Hence, the model provides a good fit for predicting crop yield across the district with most of the predicted values aligning well with the actual values. However, there are no outliers and deviations, particularly in the higher yield range, suggest opportunities for improving the model, especially for extreme values.

Following an analysis of the results for each model individually, the following observations were made:

In the case of the Random Forest algorithm, the R^2 score is the highest, thus it has higher accuracy when compared to other models and is about 94% (figure 2). The RMSE for the selected parameters is about 0.002 and the MAE score is about 0.05. Forthe K-Nearest Neighbors Algorithm commonly referred to as KNN Algorithm, the R^2score is 60%, RMSE is 0.06 and MAE is 0.05.For the SVM algorithm, R^2, RMSE and MAE are 87%, 0.05, and 0.04 respectively. And for the Multiple Linear Regression, the R^2 is 50%, RMSE is 0.36, and MAE is 0.13. The same parameters have been used for all the models to achieve the best possible accuracy.

Table 2. Different Models performance

Models	R^2Score	MSE	RMSE	MAE
RF	0.94	0.002	0.05	0.05
KNN	0.60	0.004	0.06	0.05
SVM	0.87	0.002	0.05	0.04
MLR	0.50	0.03	0.36	0.13

Table 3. Tabular representation of evaluation parameters in the Confusion matrix

Models	Precision	Recall	F1-score
RF	99%	96%	90%
KNN	90%	86%	80%
SVM	75%	84%	85%
MLR	67%	67%	67%

Based on observations, RF is the best-performing model, excelling in both precision and recall, making it highly reliablefollowed by KNN and SVMmodels.Whereas Multiple Linear Regression doesn't. It has poor precision, recall, and balance, making it less effective. Although, the Random Forest model has a higher R^2 value than the other algorithms, the RMSE and MAE are lower than other algorithms. Hence, the Random Forest Algorithm is the best-fitted model among the other models.

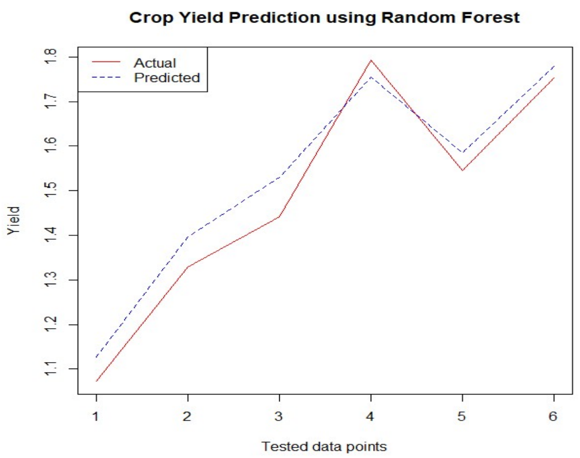

Fig. 3. Model performance by using Random Forest Algorithm

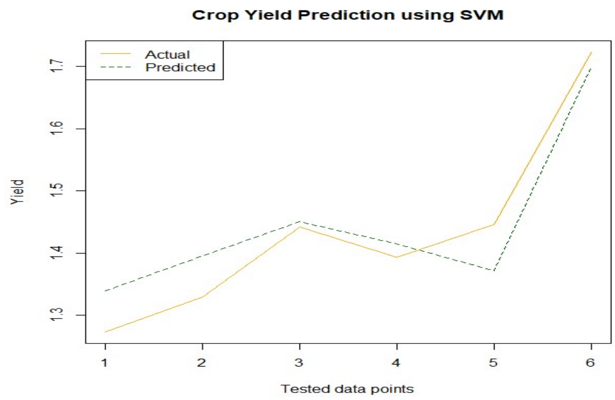

Fig. 4. Model performance by using SVM Algorithm

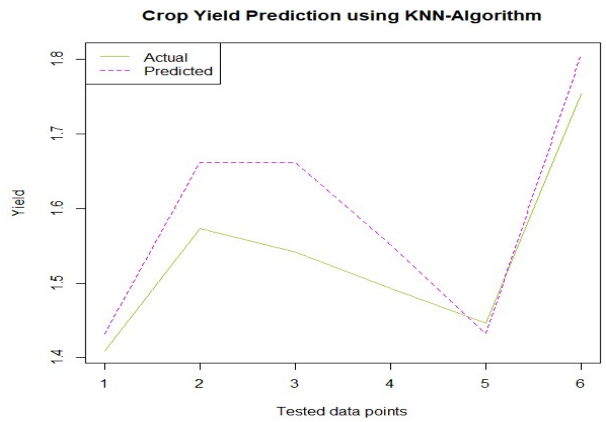

Fig. 5. Model performance by using KNN Algorithm

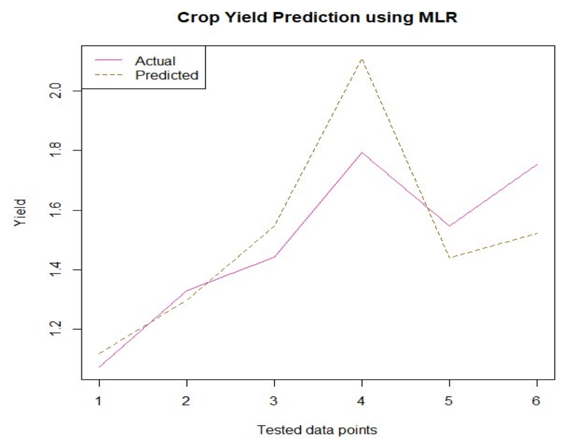

Fig. 6. Model performance by using MLR

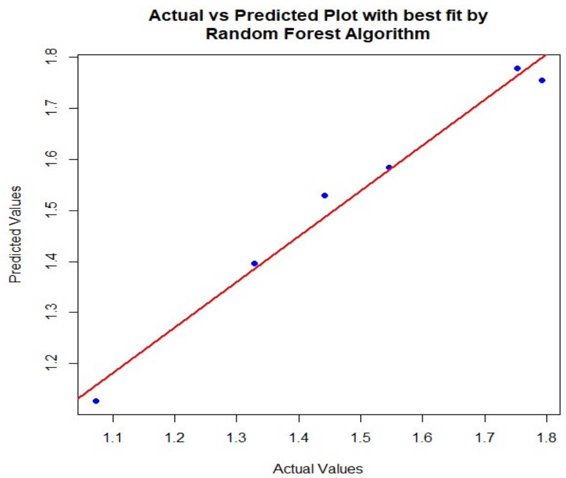

Fig. 7. Best fit by using Random Forest model

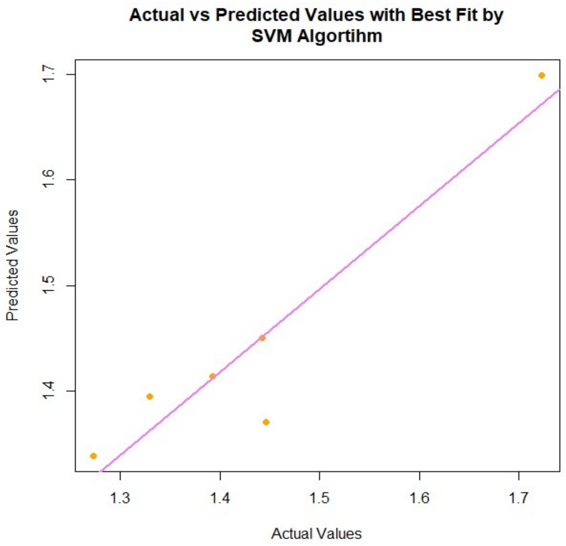

Fig. 8. Best fit by using SVM model

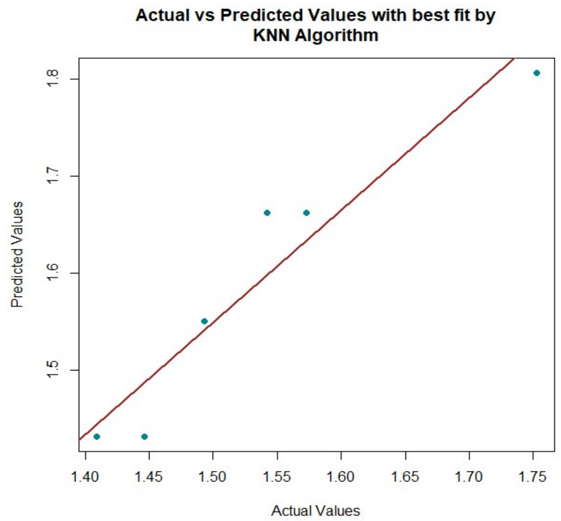

Fig.9 Best fit by using KNN model

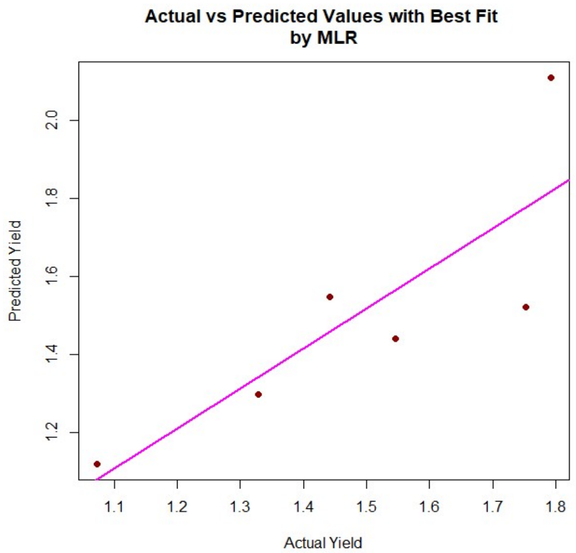

Fig. 10. Best fit by using MLR model

Interpretation: From Figure 7, we can see that the points are close to the red line, indicating the model's ideal fit. However, there is a minimal deviation from the line, indicating very accurate predictions. This graph suggests that the Random Forest model has performed well, with both precision and recall very high. From Figure 8, we can see that the points are more spread out compared to RF and KNN, showing greater deviations from the diagonal line. There is noticeable scatter, particularly for lower and higher actual values, suggesting less accurate predictions. From Figure 9, we observe that data points are slightly more dispersed around the diagonal line compared to Random Forest, but still relatively close. There are small deviations, indicating moderate prediction accuracy. However, the model's performance is still strong but slightly weaker than RF's. From Figure10, we observed many of the points lie relatively farfrom the diagonal line, particularly around the middle range of data, but among all the models, the MLR modelattains the worst accuracy, suggesting that the MLR model evolves weak prediction.

CONCLUSION

In terms of accuracy and precision in crop yield prediction, the Random Forest Algorithm scored better than the other models, with a R^2 score of 94%, MSE of 0.002, RMSE of 0.05, and MAE of 0.05. However, KNN outperformed Random Forest regarding accuracy, with a R^2 of 90% with MSE, RMSE, and MAE values of 0.004, 0.05, and 0.06, respectively. With R^2 scores of 87% and 50%, respectively, and higher MSE, RMSE, and MAE values—especially for SVM and MLR showed less predictive capability. These findings imply that, in comparison to linear regression, ensemble-based approaches, such as machine learning techniques, offer higher accuracy in crop production prediction. This study demonstrates how precise yield forecasting using machine learning techniques can increase agricultural output.

Conflict of interest: The Authors declare no conflict of interest.

REFERENCES

Agrawal, S. and Tarar, S. 2021. A hybrid approach for crop yield using machine learning and deep learning algorithms, *J. Phys.:* Conf. Ser. 17149(1): 24-25.

Agricultural Statistics at a Glance 2015. Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India. June 07: 16-20.

Araújo, S. O., Peres, R. S., Ramalho, J. C., Lidon, F., Barata, J. 2023. Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives. Agronomy. 13(12):29-76.

Aruvansh, N., Saksham, G., Agrawal, A. 2019. Crop Yield Prediction using Machine Learning Algorithms. Fifth International Conference on Image Information Processing. Doi:10.1109/ ICIIP47207.2019.8985951.

Balakrishan, N. and Dr. Muthukumarsamy, R. 2016. Crop production- enable machine learning model for prediction, *International Journal of Computer Science and Software Engineering* (IJCSSE), 5(7): 148-153.

Bharath, S., Yeshwanth, S., Yashas, B. L. and Vidyaranya, R. J 2020. Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction *International Journal of Engineering Research & Technology* (IJERT) NCETESFT. 8 (14):

Breiman, L. 2001. Random forests. Machine learning, 45, 5-32.

Fan, W. C., Chong, G., Xiaoling, Y., Hua and Juyun, W. 2015. Prediction of crop yield using big data, in Proc. 8th Int. Symp. Comput. Intell. Design (ISCID). 255-260, doi: 10.1109/ISCID.2015.191.

Gandhi, N., Armstrong, L.J., Petkar, O. and Tripathy A. K. 2016. Rice crop yield prediction in India using support vector machines, in Proc. 13th *International Joint Conference on Computer Science and Software Engineering* (JCSSE). 1(5)1:3-15, doi: 10.1109/JCSSE.2016.7748856.

Geetha, V., Punitha, A., Abarna, M., Akshaya, M., Illakiya, S. and Janani, A. P. 2020. An effective crop prediction using random forest algorithm. *International Conference on System, Computation, Automation* and 10.1109/ICSCAN49426. 2020.9262311.

Gopal, P. M. and Bhargavi, R. (2019). Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. Applied Artificial Intelligence, 33, 621-642.

Josephine, B. M., Ramya, K. R., Rao, K. R., Kuchibhotla, S., Kishore, P. V. B., & Rahamathulla, S. (2020). Crop Yield Prediction Using Learning. *International Journal of Scientific & Technology Research*, 9(02).

Kavitha, J., and Pratistha, M., Sanchit, J. and Sukriti, N. 2023. Crop yield prediction using Machine Learning and Deep Learning techniques. *International conference on machine learning and data engineering.,* 218: 406-417.

Klompenburg, V.T., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture,* 177, 2102-2106.

Kushwaha, A. K. and Bhattacharya, S. 2015. Crop yield prediction using agro algorithm in Hadoop, *Int.J.Comput.Sci.Inf.Technol*. Secure., 5(2): 271-274.

Maya, P. S., and Bhargavi, R. 2019. Performance Evaluation of best feature subsets for crop yield prediction using Machine Learning Algorithms. *Applied Artificial Intelligence.,* 33(7): 621-642.

Mishra, S., Mishra, D. and Santra, G. H. 2016. Applications of machine learning techniques in agricultural crop production. *Indian Journal of Science and Technology*. 9(38).

Narasimhamurthy, V. 2017. Rice crop yield forecasting using random forest algorithm SML. *Int.J.Res.Appl.Sci. Eng. Technol.* V. 1220–1225. https://doi.org/10.22214/ijraset.2017. 10176

Nishant, P., Venkat, P. S., Avinash, B. L. and Jabber, B. 2020. Crop yield prediction based on Indian agriculture using machine learning. *International Conference for Emerging Technology* (INCET) pp 1-4 doi: 10.1109/INCET49848.2020.9154036.

Paudal, D., Boogard, H., de Wilt, A., Janssen, S., Osinga, S., Pylianidis, C., and Athanasidis, I. N. 2021. Machine learning for large scale crop yield forecasting. *Agricultural Systems*, 187, 103016.

Priya, P., Muthaiah, U., Balamurugan, M. 2018. Predicting yield of the crop using machine learning algorithm. *Int. J. Eng. Sci. Res. Technol.* 7(1): 1–7.

Priya, P., Muthaiah, U., Balamurugan, M.” Predicting Yield of the Crop Using Machine Learning Algorithm”,2015

Ramesh, M., Vijay, S., and Shweta. 2019. Crop Yield Prediction using Machine Learning Techniques, 71(7), 480-91.

Ranjini, B., Kumar, S., and Randhawa, S. 2019. Machine Learning Methodologies for Paddy Yield Estimation in India: A CASE STUDY. International Geoscience and Remote Sensing Symposium. 7254-7257. Doi:10.1109/IGARSS.2019.8900339.

Sangeeta and Shruthi, G. 2020. Design and Implementation of Crop Yield Prediction Model in Agriculture. *International Journal of Scientific and Technology Research.* 8(1):544.

Sunil Kumar Jha, Virupakshagouda C. Patil, Rekha B.U, Shyamal S. Virnodkar, Sergey A. Bartalev, Dmitry Plotnikov, Evgeniya Elkina, Nilanchal Patel, "Sugarcane Yield Prediction Using Vegetation Indices in Northern Karnataka, India," Universal Journal of Agricultural Research, Vol. 10, No. 6, pp. 699 - 721

Suraparaju, V., Mishra, B., and Singh, C.D. 2011. Soybean productivity modelling using decision tree algorithms. *Int. J. Comput. Appl.* 2(7):975–8887.

Suresh, A., Ganesh, P. and Ramalatha, M. 2018. Prediction of major crop yields of Tamil Nadu using K-means and Modified KNN, 3rd International Conference on Communication and Electronics Systems (ICCES). 88-93 doi: 10.1109/CESYS.2018.8723956.

Veenadhari, S., Misra, B, and Singh, C.D. 2014. Machine learning approach for forecasting crop yield based on climate parameters. International Conference on Computer Communication and Informatics.1-5.

****