University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)

ISSN: 2714-3627

A Journal of the Faculty of Computing, University of Ibadan, Ibadan, Nigeria

Volume 15 No. 1, September 2025

journals.ui.edu.ng/uijslictr
http://uijslictr.org.ng/
uijslictr@gmail.com

University of Ibadan
Journal of Science and Logics in ICT Research

Particle Swarm Optimization-Random Forest Weather-Based Crop Yield Prediction Model

[*1] Abisoye O.A., [1]Adedokun J. A., [2]Abisoye B.O., [1]Lawal O.L., [3]Olajire J. A., [3]Kama M.

[1]Department of Computer Science, Federal University of Technology, Minna, Nigeria
[2]Department of Computer Engineering, Federal University of Technology, Minna, Nigeria
[3]Department of Software Engineering, Veritas University, Abuja, Nigeria
***Corresponding Authors Email :** o.abisoye@futminna.edu.ng. opeyemiabisoye1@gmail.com*
**Corresponding Authors ORCID:https:// https://orcid.org/0000-0002-7324-2479*

Abstract
Crop production is a vital source of food for humans, and improving crop yield requires a deep understanding of crop production processes. It has been proven that increasing crop yield reduces poverty, crop failure risk, increases productivity, and optimizes the value of agricultural land. Many factors affect the amount of crop harvested in a specific area and several studies, mainly in the agricultural context, have been conducted to estimate crop yield production with Machine learning (ML) techniques. This study explores five cereal crop yields: rice, maize, wheat, sorghum, and soybeans with Particle Swarm Optimization (PSO) and Random Forest prediction approaches. Performance metrics such as R^2 score, Mean Absolute Error, and Root Mean Squared Error confirm the authenticity of the model. The result of the optimized Crop yield prediction has an R^2 score of 97.13, MAE of 124.75, and RMSE of 1273.73. The model performed better than other existing approaches, such as Random Forest (RF) and Decision Tree (DT). This study will provide farmers with reliable crop yield predictions, enabling better planning based on weather conditions.

Keywords: *Weather-Based, Crop-Yield, Particle Swarm Optimization (PSO), Random Forest, Machine Learning*

1. Introduction

The vast world's population still depends on agriculture for their livelihoods, which has historically been linked to the production of staple crops [3]. But population expansion, resource depletion, droughts, and climate change have all affected agricultural output [7]. These factors have also interrupted food supplies and disproportionately affected the poorest communities during famines. Hunger can spread if population expansion is allowed to outrun food production [11]. These problems are made worse by climate change; low rainfall, extreme weather and other variables, including air pollution, water scarcity, and rising temperatures reduces crop productivity. Farmers can create adaptive strategies to maximize crop development in a more sustainable and effective way by analyzing and forecasting crop production outcomes under various climatic circumstances and management approaches. [2][14].

Crop modelling in agriculture uses quantitative measurements of eco-physiological processes to predict crop growth and development based on environmental factors and management decisions. These models simulate how crops respond to factors such as weather, water, soil, and management inputs throughout the growing season. Crop models mathematically represent different components of the cropping system and are used to simulate growth and development. It was observed that ML models meaningfully increase the accuracy of crop yield predictions in compares with traditional forecasting methods [5][13], but there are certain difficulties to achieving a certain level of granularity and predictive precision especially in a rapidly varying environmental condition. Inspired by the collective behavior of swarms like flocks of birds, schools of fish, Particle Swarm Optimization (PSO) is a vast

Abisoye O. A., Adedokun J. A., Abisoye B. O., Lawal O. L., Olajire J. A. and Kama M. (2025). Particle Swarm Optimization-Random Forest Weather-Based Crop Yield Prediction Model. *University of Ibadan Journal of Science and Logics in ICT Research University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 15 No. 1, pp. 57 - 68.

technique for solving such problems. Because of its simplicity, robustness, low parameter requirements, parallel processing compatibility, easy global optima, rapid convergence, short time complexity, and avoidance of redundant calculations and mutations, it has spanned domains for feature selection and complex optimization problems. Random Forest, a potent tree-based algorithm, generates many decision trees. Many existing models suffer from over fitting, which increases their complexity. By promoting variation among the trees, this intrinsic randomness reduces overfitting and increases the accuracy of the model.

This study aims to address this challenge by implementing a Particle Swarm Optimization (PSO) and Random Forest model to predict crop yield. The study introduction is covered in Section 1 of this document. Literature reviews are covered in Section 2. The project's methodology and algorithms are explained in Section 3. The study Results and Discussions are displayed in Section 4. The Conclusion and future work are discussed in Section 5.

2. Related Works

A model that identifies the best crops for farmers by combining different ML techniques, Artificial Neural Network (ANN), Random Forest, and Decision Tree, is presented by Agarwal and Tarar [1]. Crucial elements such as crop rotation, soil texture, air and surface temperatures, precipitation, and soil nutrients (potassium, phosphorus, and nitrogen) were considered in the model. In Venugopal *et. al.* [18], three ML methods, viz, Random Forest, Naïve Bayes, and Logistic Regression, were used to predict crop yields based on important variables including land area, temperature, and rainfall. Naïve Bayes obtained an accuracy of 91.5%, while Logistic Regression had an accuracy of 87.8%. Nevertheless, Random Forest outperformed other algorithms with the highest accuracy of 92.81%. In the end, the researchers created an application that combines these models, allowing users to enter pertinent information and receive forecasts on the best crop selections and anticipated yields. From these studies climatic variability is a very crucial determinants of crop yield predictions.

Crop yield prediction for maize and Irish potatoes based on weather data from several sources was conducted by [12]. The study identified the ideal meteorological conditions for optimizing yields, with three(3) predictive models: Support Vector Regressor, Random Forest and Polynomial Regressor In Testing Phase, Potatoes has RMSE values of 510.8 for potatoes 129.9 for maize, and corresponding R^2 scores of 87.5 and 81.7 , the Random Forest model perform best with RMSE 740.1, 152.7, R^2 scores 77.3 and 71.6 respectively for potatoes and maize, as well, the Polynomial Regression model came in second. The Support Vector Regressor had the worst performance, generating 971.6, 212.4 RMSE, 56.0, and 54.9 R^2 scores for potatoes and maize, respectively. The resarchers stated that RF model framework can be employed to create more precise crop yield projections while employing IoT technology to track weather and soil moisture

Random Forest (RF) and Multiple Linear Regression (MLR) models were employed to forecast crop yield by Sherif [17] to examine maize, rice, and wheat yields with three (3) combined datasets. Performance criteria like R^2, RMSE, and run time were used to assess the models. Regardless of the feature selection strategy, the results showed that the RF model consistently has better performance than MLR. To counteract this problem, the researcher used cross-validation because of worries about possible overfitting. This modification increased the Random Forest model's accuracy when all variables were used, as well as when stepwise regression and correlation scores were used to pick features and consequently, in these circumstances, the accuracy of the MLR model decreased. Consequently, from these findings, the researchers recommended boosting Random Forest model for crop yield prediction.

The Random Forest technique, which aggregates the majority vote from several decision trees to generate predictions, was used by Doshi *et. al.* [8]. An R^2 score of 72.7% was achieved and the researcher suggested creating a mobile application that would use messaging systems to promptly notify farmers on the best times to plant and harvest. By enabling data-driven decision-making in watering, planting, and harvesting, the study by Elbasi *et. al.* [10] integrated Internet of Things (IoT) sensors with ML algorithms to minimize waste and maximize agricultural productivity. Using Bayes Net and Random Forest for analysis, the study sought to determine how well classification algorithms predicted broad crop groupings. With environmental parameters, temperature, humidity, precipitation, and pH features the Random Forest and Bayes Net models achieved 97.32%, 97.05% respectively. Those findings shows that adding more variables, can increase forecast accuracy and the substantial advantages of using IoT sensors and large amounts of farm data, guarantee farmers of a better decisions and eventually increasing crop yields.

With crop kind and seed variety as input variables, Palanivel and Surianarayanan [16] developed a crop yield prediction model with four (4) machine learning methods; SVM, NB, ANN, and Multiple Linear Regression. The data was preprocessed to remove any noisy or unformatted components. From the study, it was discovered that the ANN and SVM models produced the most precise estimates of crop yield. With Random Forest algorithm, researchers created a crop yield forecast model by Champaneri *et. al.* [4] that integrated important variables like temperature, humidity, rainfall, soil moisture, and weather. Multiple sources of data were gathered, and both descriptive and diagnostic analytics were used for analysis. The model was trained to forecast crop yields on a regional and global scale. The researchers also created an easy-to-use website that lets people create crop yield estimates..

In order to help farmers choose crops depending on soil nutrients and meteorological circumstances, Madhuri *et. al.* [13] sought to identify the best crop prediction model. Researchers compared the performance of a number of popular algorithms, such as RF, DT, and KNN with two metrics: Gini and Entropy. The DT classifier performed moderately, with the Gini criterion (98.86%) surpassing the Entropy criterion, while the KNN algorithm had the lowest accuracy (97.04%). To expand the scope of crop classification, the researchers suggested exploring more ML and DL models. ML techniques were explored by Changulanda and Srikanth [5] to predict crop yields by analyzing historical data involving weather patterns, soil conditions, and other pertinent factors. The study also examines the effects of feature selection, hyperparameter tuning, and model interpretability on prediction accuracy. With feature engineering, to build predictive models, valuable insights from the data were extracted thus, allowing models to effectively capture the crop yield complex features with several approaches, such as DT, RF and ANN. It was observed that ML models meaningfully increase the accuracy of crop yield predictions in compares with traditional forecasting methods, but there are certain difficulties to achieving a certain level of granularity and predictive precision especially in a rapidly varying environmental condition.

From existing studies it was discovered that accurate prediction of yields is crucial for enhancing agricultural productivity and ensuring food security in sustaining the growing population of the world and the old methods which involves farmers using past yield to predict what they may likely have when farming in the current season were very slow, not reliable and sizable quantity of crops are damaged before harvest because of bacterial attacks and lack of adequate information for proper planning. To resolve these challenges, farmers approaches to planting needs to be revolutionized by giving consideration to climatic variability for crop yield prediction, intersecting agriculture and technology with machine learning approaches.

3. Methodology

3.1. Random Forest

Random Forest as a supervised ML algorithm is being used for regression in this study, but it can be used for both classification and regression tasks [15][19]. As an ensemble learning technique, its multiple decision trees improve accuracy of predictions. The basic idea of RF is to create multiple decision trees -based model as shown in Figure 1 to map decisions to their possible outcome and produce a good prediction; generally speaking, more trees are reliable and produces better accuracy. During the training phase, the model leverages on the tree nodes while discarding weaker ones to improve prediction quality.

The architecture of Random Forest is shown in Figure 1. The Random Forest algorithm works in two basic steps: first, it builds a forest by building N decision trees, and then it generates predictions by combining the output from these trees. The mathematical formula for Random Forest is described in equation 1.

$$\bar{Y}\mathrm{i} = \frac{1}{\mathrm{N}}\sum_{i=1}^{N} P_i(x) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (1)$$

where $\bar{Y}$i, $P_i(x)$, N depict predicted output, ith tree prediction, and the tree's total number, respectively.

3.2. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a searching algorithm invented in 1995 [9] which relies on the collective movement of swarms, for example: schools of fish or flocks of birds. PSO is popular because it is easy to use, requires few parameter adjustments, can be processed in parallel, is highly efficient, converges quickly, and has a strong potential to reach global optima. It also avoids redundant computations and mutations, which makes it highly effective in different applications for feature selection. In this study, PSO was used specifically for feature selection. It represents potential solutions as particles that search space to find the optimal results; each particle is made up of a defined position that is evaluated to determine its efficacy, and a velocity that determines the direction of its movement and speed. The mathematical formulas for updating velocity and position are described in equations 2 and 3, respectively.

$$V_i^{t+1} = \mathrm{w}\,.\,v_i^t + c_1\,.\,r_1\,.\,(P_i - X_i^t) + c_2\,.\,r_2\,.\,(g - X_i^t) \qquad (2)$$

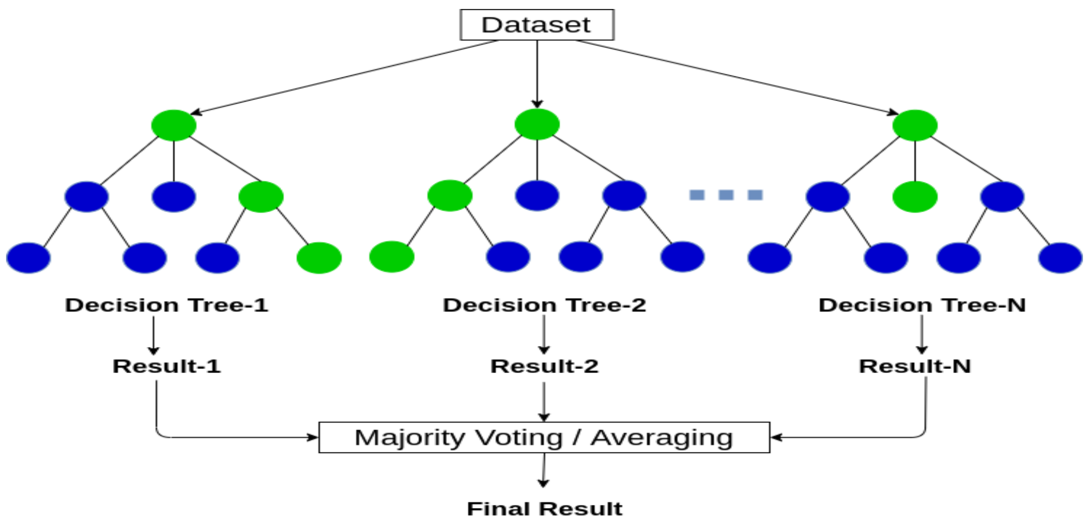
Where V_i^{t+1} represents the velocity update, w is the inertia weight, c_1 and c_2 are cognitive and social coefficients, r_1 and r_2 signify uniform random numbers [0, 1].

$$X_i^{t+1} = x_i^t + V_i^{t+1} \qquad (3)$$

3.3 PSO-RF Crop Yield Proposed Model

3.3.1. Datasets and Attributes

Effective data collection is essential to obtaining accurate results. The eight attributes of the datasets are presented in Table 1 based on five (5) cereal crops totaling 2,595 entries, was collected from Kaggle repository. The Five (5) cereal crop yields —rice, maize, wheat, sorghum, and soybeans were explored in this study and the snippets of the crops distribution is shown in Figure 2. The datasets undergone training and testing in ratio 80:20 respectively and this process allows for the identification of trends from historical data, aiding in the recognition and mitigation of recurrent patterns.



RF Model Architecture (Viswas et al., 2024)

Table 1. Crop Yield Dataset Description

Input	Feature	Crop Demography
	Area	Albania
	Item	True or False
	Year	$1990 \leq Y \leq 2013$
	Average rainfall (mm/year)	$1083.0 \leq R \leq 2875.0$
	Pesticides tonnes (Tonnes)	$1597 \leq P \leq 75000$
	ave_temp (Celsius)	$23.77 \leq T \leq 28.10$
Output	hg/ha_yield (Hectogram per hectare)	$6553 \leq C \leq 385818$

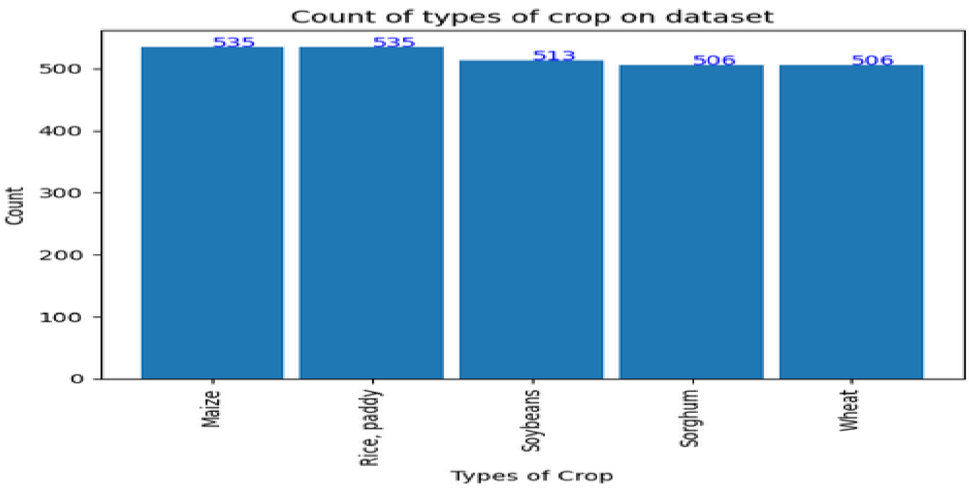


Figure 2: Snippet of the Crops Distribution.

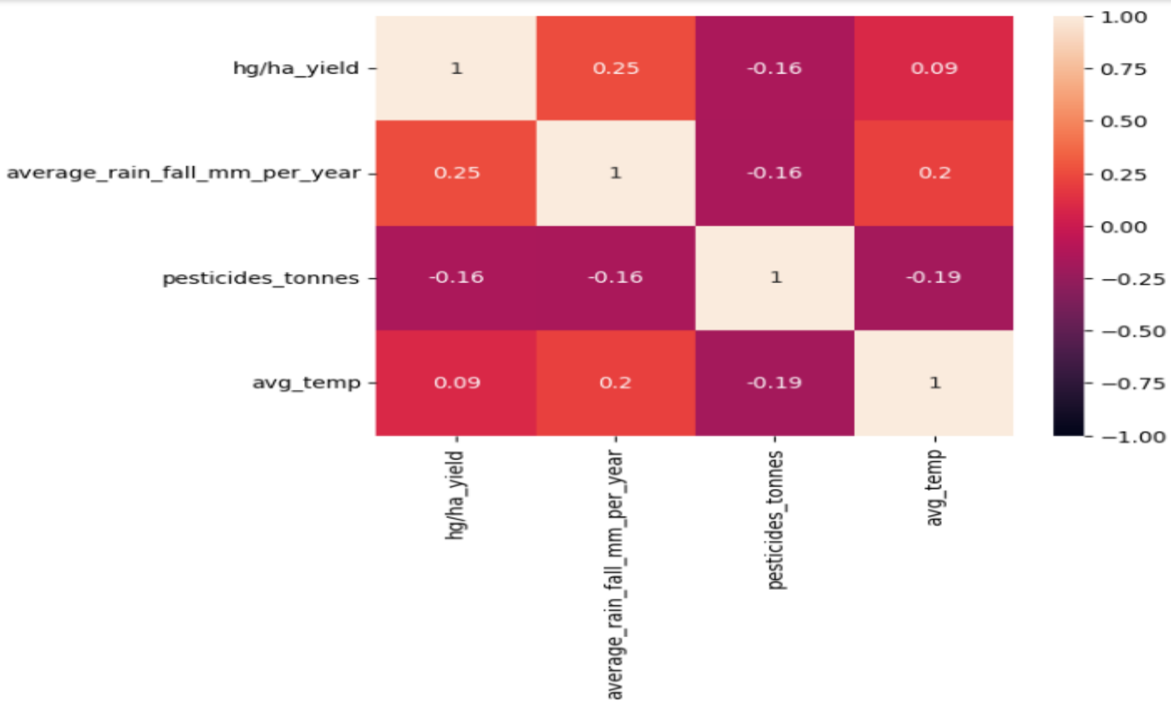


Figure 3: Pairwise correlation between attributes

3.3.2 Pairwise Correlation between Attributes

The Pairwise correlation of the crop yields between attributes and several conclusions regarding the dataset's correlation can be drawn from Figure 3. Pairwise correlation evaluates the linear relationships between two attributes, measuring both the strength and direction of their association. The correlation coefficient, denoted as 'r,' ranges from -1 to +1: An r value of +1 indicates a perfect positive correlation (an increase in one variable result in a proportional increase in the other). An r value of -1 signifies a perfect negative correlation (an increase in one variable leads to a proportional decrease in the other). An r value of 0 indicates no relationship between the variables. The absolute value of the correlation coefficient reflects the relationship's strength: values closer to 1 (whether positive or negative) indicate stronger correlations, while values nearer to 0 suggest weaker correlations. From the analysis, we observe that the correlation between hg/ha_yield and average_rain_fall is 0.25, indicating a weak positive correlation. This means the relationship between these variables is not strong or consistent, but as one variable increases, the other tends to increase as well, albeit weakly. The correlation between hg/ha_yield and pesticides is -0.16, which also suggests a weak negative correlation. This implies that as one variable increases, the other tends to decrease, but again, the relationship is not strong. For hg/ha_yield and avg_temp, the correlation is 0.09, indicating a weak positive correlation similar to that of hg/ha_yield and average_rain_fall as shown in Figure 3. The correlation between average_rain_fall and pesticides is -0.16, indicating another weak negative correlation, suggesting a slight tendency for one variable to decrease as the other increases.

The average_rain_fall and avg_temp correlation of 0.2 indicates a weak positive correlation, meaning that while the relationship exists, it is not particularly strong. Lastly, the correlation between pesticides and avg_temp is -0.19, suggesting a weak negative correlation, where increases in one variable tend to coincide with decreases in the other, but the relationship remains weak.

3.3.3. Data Preprocessing

This stage guarantees that the features of the dataset can be efficiently interpreted by machine learning algorithms. To deal with duplicate entries, outliers, and missing values, extensive pre-processing was done for this investigation. Min-Max Scaler was used for normalization, while Standard Scaler was utilized for scaling. Additionally, categorical variables were converted into a machine learning-compatible format through one-hot encoding.

Algorithm 1: Random Forest model for the regression task.

Input: Crop Yield Dataset
Output: Evaluation Metrics

1. Start
2. Enter the Crop Yield Dataset
3. Normalize the dataset in (2)
4. Perform holdout cross-validation
5. Build decision trees with the chosen instances
6. Determine the number of trees N
7. set i = 1
8. While $T_{i\ <=\ N}$
9. Determine a subclass of the decision tree from the training dataset with replacement
10. Determine a subclass of features Fj from the total feature set F
11. Select the best feature to split based on Mean Squared Error
12. Split the node based on the selected features in (11).
13. Get Prediction $\bar{Y}\text{i} = P_i(x)$
14. Calculate the average of the predictions $\bar{Y}\text{i} = \frac{1}{\text{N}}\sum_{i=1}^{N} P_i(x)$
15. End While
16. Return $\bar{Y}\text{i}$
17. End

3.3.4. Proposed Crop Yield Random Forest Model

To increase prediction accuracy, this model gathers training data from every node in the tree and eliminates weaker nodes. The Random Forest method works in two steps: first, it combines N decision trees to build a random forest, and then it aggregates the outputs from these trees as shown in Algorithm 1 showing regression task to generate crop yield prediction.

3.3.5. Proposed PSO-RF Model

Particles in PSO are candidate solutions that move around the search space in quest of the best results as shown in Algorithm 2. Each particle is identified by its velocity, which establishes the direction and force of its motion, and its position, which gauges its efficacy. PSO selects the best values to optimize RF with the hyperparameters in Table 2.

4. Result and Discussion

A PSO-RF crop yield prediction model was developed for five cereal crops, with considerations to climatic factors as part of input features for the model. PSO helps to select optimal features and provided best hyper-parameter optimization. The RF model achieved an R^2 score of 91.96, MAE of 220.22, and RMSE of 2768.27. In contrast, the PSO-optimized RF (PSO-RF) model significantly improved performance, achieving an R^2 score of 97.13, MAE of 124.75, and RMSE of 1273.73 as shown in Table 3 and Figure 6. Higher R^2 value of the PSO-RF model (97.13%) compared to the RF model (91.96%) depicted in Figure 4, 5 and Figure 8, indicates a stronger alignment between the input features and the target variable thus decreases the variance. The lower MAE of the PSO-RF model (124.75) compared to the RF model (220.22) reflects improved accuracy and precision, with smaller average absolute prediction errors. Additionally, the significantly reduced RMSE of the PSO-RF model (1273.73) as shown in Figure 7 compared to the RF model (2768.27) highlights its effectiveness in minimizing larger prediction errors, which is critical for practical applications.

Table 2: PSO Hyperparameters

S/N	Hyperparameters	Values
1.	n-estimators	200-300
2.	Max_Depth	10-20
3.	Min_Sample Split	2-4
4.	Min_Sample leaf	2-4

Algorithm 2: PSO-RF model for the Regression task.

Input: Crop Yield Dataset
Output: Evaluation Metrics

1. Start
2. Enter the Crop Yield Dataset
3. Normalize the dataset in (2)
4. Initialize particles and subsequent velocities of the swarm
5. Compute particle fitness
6. Update V_i^{t+1} using Equation 2
7. Update X_i^{t+1} using Equation 3
8. Initialize Max_iteration = 5, counter i = 1
9. While $T_i <= \max\_iteration$
10. Choose the required features
11. Repeat 5 to 7
12. End While
13. Output global best, evaluation metric
14. End

PSO-RF model outperformed the RF model as shown from the performance metrics and Distribution plot in Figure 8. PSO-RF captured the underlying data patterns successfully, leading to convergent prediction trends as shown in the Receiver Operating Characteristic (ROC) Curve (Figure 9), Violin Plot (Figure 10) and Box plot (Figure 11). The PSO-RF model prioritized features with the highest impact on performance, resulting in improved accuracy while maintaining the predictive trends observed in the RF model. This shows that PSO's optimization complements RF's inherent strengths, refining its predictions and reducing errors.

Figure 5 depicts the result of the PSO-RF R2 Score result after 5 iterations, and also show the index of the selected features. The selected features are: average_rain_fall, pesticides, Rice, paddy, Sorghum, Soybeans and Wheat.

Table 3: Experiment result

Model	R^2	**MAE**	**RMSE**
RF	91.96	220.22	2768.27
PSO-RF	97.13	124.75	1273.73

```
[35] y_pred=rfr.predict(X_test_scaled)

[36] print("R2 score:","{:.2f}".format(r2_score(y_test, y_pred)*100))
R2 score: 91.96

[37] print("MAE:", "{:.2f}".format(mean_absolute_error(y_test, y_pred)))
MAE: 220.22

rmse = "{:.2f}".format(np.sqrt(mean_squared_error(y_test, y_pred)))
print("RMSE:", rmse)
RMSE: 2768.27
```

Figure 4: Snippet of the Random Forest Results

```
Iteration 1/5, Best R2 Score: 97.09
Iteration 2/5, Best R2 Score: 97.09
Iteration 3/5, Best R2 Score: 97.09
Iteration 4/5, Best R2 Score: 97.09
Iteration 5/5, Best R2 Score: 97.13
Final R2 Score with Selected Features: 97.13
Selected Features: [0 1 5 6 7 8]
```

Figure 5: Snippet of the PSO-RF R2 Score Result

```
Iteration 1/5, Best MAE Score: 124.83
Iteration 2/5, Best MAE Score: 124.83
Iteration 3/5, Best MAE Score: 124.83
Iteration 4/5, Best MAE Score: 124.83
Iteration 5/5, Best MAE Score: 124.75
Final MAE Score with Selected Features: 124.75
Selected Features: [0 1 3 4 6 7 8]
```

Figure 6: Snippet of the PSO-RF MAE Result

Figure 6 depicts the result of the PSO-RF R2 Score result after 5 iterations, and also show the index of the selected features. The selected features are: average_rain_fall, pesticides, India, Maize, Sorghum, Soybeans and Wheat.

Figure 7 shows the result of the PSO-RF R2 Score result after 5 iterations, and also show the index of the selected features. The selected features are: average_rain_fall, pesticides, India, Maize, Sorghum, Soybeans and Wheat.

```
Iteration 1/5, Best RMSE Score: 1275.65
Iteration 2/5, Best RMSE Score: 1275.65
Iteration 3/5, Best RMSE Score: 1275.65
Iteration 4/5, Best RMSE Score: 1275.65
Iteration 5/5, Best RMSE Score: 1273.73
Final RMSE Score with Selected Features: 1273.73
Selected Features: [0 1 3 4 6 7 8]
```

Figure 7: Snippet of the PSO-RF RMSE result

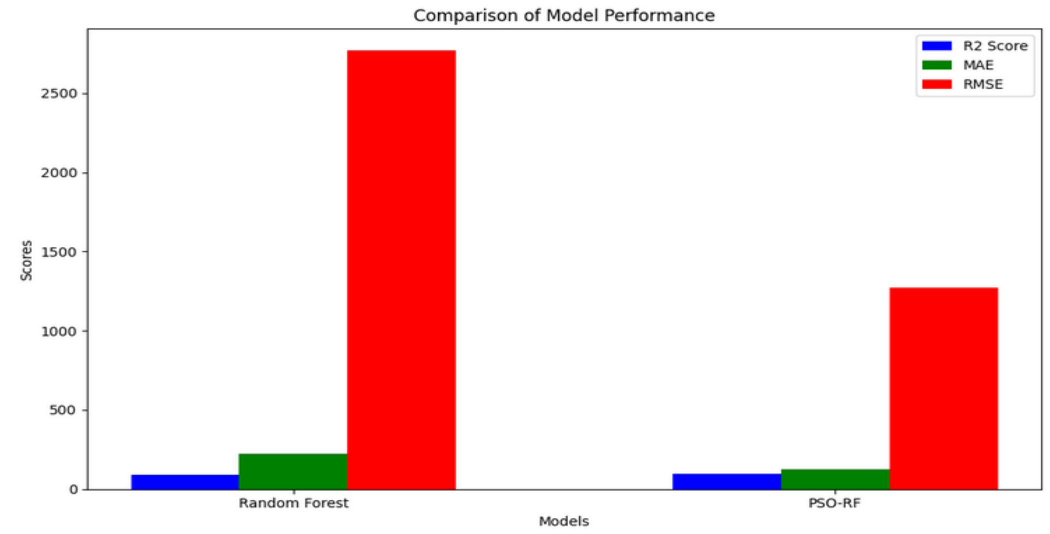
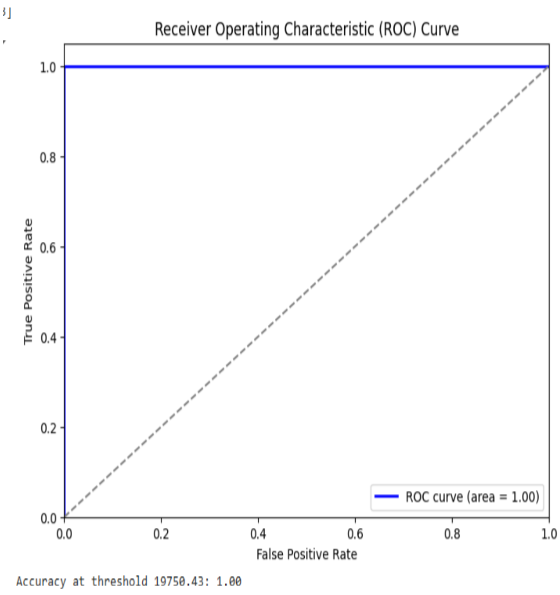


Figure 8. Comparison of Model Performance



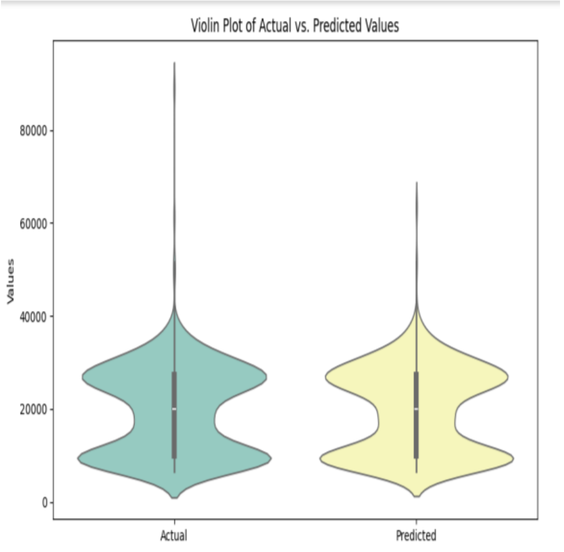


Figure 9. ROC Curve Figure 10: Violin Plot Comparing Actual and Predicted Values

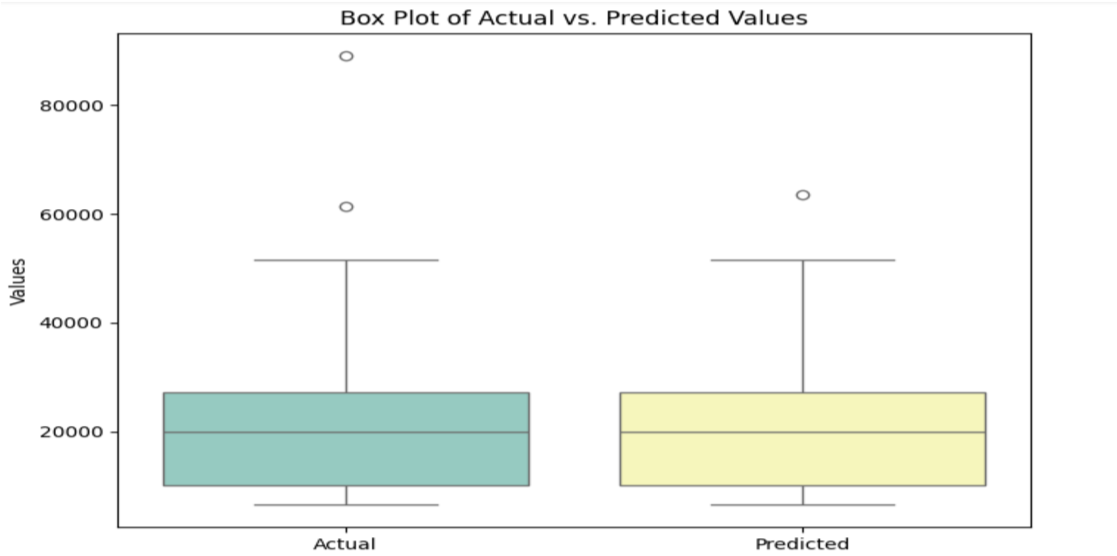


Figure. 11. Box Plot Comparing Actual and Predicted Values

In Table 3 this research obtained accuracy (R2) of 97.13% indicating it ability to measures the proportion of variance that exist in the dependent variable and is predictable from the independent variables. The model obtained a MAE of 124.75 indicating smaller average absolute errors and a RMSE of 1273.73 indicating a minimized larger prediction error. This weather based yield prediction performs better when compared to other existing models. This proposed crop yield prediction model was compared with Baeline literatures for benchmarking as presented in Table 4.

4.2. Performance Comparison between PSO-RF and Baseline work

Table 4: Performance Comparison between PSO-RF and Baseline work

Author/Year	Methodology	Accuracy(R2) %	MAE	RMSE
Venugopal *et al* (2022)	Random Forest, Linear Regressor Naïve Bayes	92.81 87.8 91.50	NA NA NA	NA NA NA
Agarwal and Tarar (2021)	DT& ANN& RF LSTM& RNN&SVM	93 97	NA NA	NA NA
Kuradusenge *et al*. (2023)	RF Regression Polynomial Regression Support Vector Regression	87.5 77.3 56	418.699 563.587 722.779	510.817 740.199 971.633
Viswas *et al*. (2024)	Random Forest	85	NA	16050251.89
Sherif (2022)	ML Regression Random Forest	84.31 88.06 .	NA NA	7059.34 5566.9
Abisoye *et al*., (2025)	PSO-RF	97.13	124.75	1273.73

5. Conclusion and Future Work

A common issue with many existing crop yield prediction models is overfitting, primarily caused by the noisy nature of datasets. This overfitting leads to poor predictive performance and increased model complexity, leading to problems in adapting to changes in data distribution. To address this, the study employed PSO to choose the most optimal parameters. PSO stochastic and global search capabilities, reduces the likelihood of being trapped in local optima, thereby enhancing regularization and improving model performance. Its hyper-parameter tuning ability also helps optimize model parameters, reducing complexity and enhancing adaptability.

After selecting the optimal features through PSO, the PSO-RF model's outcomes were evaluated using metrics such as R2, MAE and RMSE. The optimized PSO-RF model consistently outperformed standard models that do not leverage hyper-parameter optimization. It demonstrated superior accuracy, reduced average prediction errors, and effectively minimized larger errors, making it more robust for crop yield prediction.

This research primarily focused on predicting crop yields based on climatic (weather) conditions using the optimized PSO-RF model.. Emerging technologies such as Internet of Things (IoT) devices can be employed to gather real-time weather data, further improving prediction accuracy. Additionally, a web-based application integrated with an Application Programming Interface (API) could streamline the prediction process, making it more accessible and user-friendly.

References

[1] Agarwal, S., & Tarar, S. (2021). A hybrid approach for crop yield prediction using machine learning and deep learning algorithms. Journal of Physics: Conference Series, 1714.

[2] Archana, S., & Senthil, P. K. (2023). A Survey on Deep Learning Based Crop Yield Prediction. Nature Environment and Pollution Technology, 2, 579 – 592.

[3] Asma, J., Subrahmanyam, D., & Krishnaveni, D. (2023). The global lifeline: A staple crop sustaining two thirds of the world’s population. *Agriculture Archives*.

[4] Champaneri, M., Chandvidkar, C., Chachpara, D., & Rathod, M. (2020). Crop yield prediction using machine learning.

[5] Changulanda, B. B, & Srikanth, V. (2024). Crop yield prediction on agriculture using machine learning. International Journal of Research Publication and Reviews, 5, 165-167.

[6] Chen, X., & Courn`ede, P. (2017). Model-Driven and Data-Driven Approaches for Crop Yield Prediction: Analysis and Comparison. International Scholarly and Scientific Research & Innovation, 7, 334 – 342.

[7] De Wrachien, D., Schultz, B., & Goli, M. B. (2021). Impacts of population growth and climate change on food production and irrigation and drainage needs: A world-wide view. *Irrigation and Drainage*, *70*(5), 981-995

[8] Doshi, A., Chopade, A., & Bodekar, S. Predictive Agriculture Using Data Analysis and Machine Learning. Journal of Emerging Technologies and Innovative Research (JETIR), 8, 15 – 18.

[9] Eberhart, R., Kennedy, J. (1995) A new Optimization using particle swarm theory(C) Micro Machine and Human Science. Proceedings of the Sixth International Symposium on. IEEE, 1995, 39-43.

[10] Elbasi, E., Zaki, C., Topcu, E. A., Abdelbaki, W., Zreikat, I. A., Cina, E., … Saker, L. (2023). Crop prediction model using machine learning algorithms. multidisciplinary digital publishing institute, 1, 1 - 25.

[11] Kent, G. (2015). *Ending hunger worldwide*. Routledge.

[12] Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., … Uwamahoro, A. (2023). Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. Multidisciplinary digital publishing institute,13, 225.

[13] Madhuri, S. R., Singh, A., Subba, N. V. R., & Dinesh, U. A. (2022). Crop prediction using machine learning. Journal of Physics: Conference Series, 2161, 012033.

[14] Mythili. K., & Rangaraj, R. (2021). Deep Learning with Particle Swarm Based Hyper Parameter Tuning Based Crop Recommendation for Better Crop

Yield for Precision Agriculture. Indian Journal of Science and Technology, 14, 1325–1337.

[15] Padma, T., & Sinha, D. (2023). Crop yield prediction using improved random forest. ITM Web of Conferences, 56, 02007.

[16] Palanivel, K., & Surianarayanan, C. (2019). An approach for prediction of crop yield using machine learning and big data techniques. International Journal of Computer Engineering and Technology (IJCET), 10, 110-118

[17] Sherif, H. (2022), Machine Learning in Agriculture: Crop Yield Prediction. RIT Digital Institutional Repository.

[18] Venugopal, A., Aparna, S., Mani, V., Mathew, R., Williams, V. (2021). Crop Yield Prediction using Machine Learning Algorithms. International Journal of Engineering Research & Technology (IJERT), 9, 87 – 9

[19] Viswas, M., Sri, K., Veera, N., Swami, V., Surya, S., Ramu, K., Rama, M., & Reddy, P. (2024). Crop Yield Prediction Using Machine Learning. International Journal of Innovative Science and Research Technology, 9, 1245 – 1248.