Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 12 Issue 1, January - 2025

DETERMINATION OF OPTIMAL YIELD FOR PALM KERNEL OIL EXTRACTION MACHINE USING RANDOM FOREST REGRESSION MODEL

Emmanuel Udama Odeh[1]
Department of Mechanical and Aerospace Engineering
University of Uyo, Akwa Ibom State, Nigeria
emmanuelodeh@uniuyo.edu.ng

Emem Sunday Ezekiel[2]
Department of Mechanical and Aerospace Engineering
University of Uyo, Akwa Ibom State, Nigeria
ememezekiel@gmail.com

OLALEYE, O. Olukayode[3]
Marine Engineering Department,
Maritime Academy, Oron, Akwa Ibom State, Nigeria
kayola_nan@yahoo.com

***Abstract*— In this work, determination of optimal yield for Palm Kernel Oil (PKO) extraction machine using Random Forest Regression (RFR) model is presented. The study utilized 5000 data records of a case study 10-ton PKO extractor machine in Uyo, Akwa Ibom State, Nigeria for the model training and validation. Also, SHAP (SHapley Additive exPlanations) feature importance approach was used to evaluate the importance ranking of each of the three input features to the RFR model. The results show that moisture content with feature importance ranking of 0.16 has the highest impact on the RFR model prediction while the Cone gap with feature importance ranking of 0.137 has the lowest impact. Also, the model perdition had Mean Absolute Error (MAE) of 1.0678 x 10^{-15} and Mean Squared Error (MSE) of 2.14306 x 10^{-30} which are very small (negligible) hence the coefficient of correlation between the actual and the predicted results was 1. Furthermore, the results showed that the highest oil yield of 43.4 % occurred at shaft speed of 18 rpm, cone gap of 1.5 mm and moisture content of 8 %. It means that for maximum PKO, the case study PKO extractor machine should be operated with the input settings as specified in the RFR model optimal solution result.**

***Keywords*— Optimal Yield, Palm Kernel Oil Extraction Machine, SHAP (SHapley Additive exPlanations), Optimization Model, Random Forest Regression Model, Feature Importance**

1. Introduction

In recent years, there has been increasing adoption of Artificial Intelligence (AI) in different sectors [1,2]. The AI approach enables accurate modelling of systems, devices, or events, by relying on historical data pertaining to the case study systems, devices, or event [3,4,5]. The AI model has proven in many cases to be more efficient and accurate than the conventional analytical models [6,7]. As such, researchers are increasingly relying on various types of AI models for characterizing their case study systems, devices, or events [8,9].

Due to the growing adoption of the AI model, the industrial sector is increasingly applying the AI solution to optimize their machines and system, increasing their productivities and cutting down cost [10,11,12]. In any case, the AI models are data intensive, requiring large volume of data records for effective modelling of the case study system [13,14,15]. In view of this requirement, AI solutions have also ignited a new trend whereby different industries keep track of their operations, system configurations, productivity, and maintenance and inventory data for application in data driven model development.

In this study, the application of Random Forest Regression (RFR) model in the determination of optimal palm kernel oil (PKO) yield of a PKO extractor machine is presented [16,17,18]. The study aims at determining the specific input parameters setting that gives the optimal PKO yield for the case study machine. The outcome of such study will enhance productivity, minimize waste and improve on the revenue and profit accruing from the case study machine.

2. Methodology

This work presents detailed theoretical model that underscores the prediction of palm kernel oil (PKO) extractor machine yield using Random Forest Regressor (RFR) model while focusing on feature importance analysis and capturing non-linear interactions. A comprehensive understanding of the underlying relationships between the

input parameters and oil yield, leveraging the strengths of RFR model with architecture shown in Figure 1 is presented.

The aim of the Random Forest Regression (RFR) model as used in this study is to predict the PKO yield denoted as Y based on three parameters: X_1= main shaft speed, X_2= moisture content, and X_3= cone gap. The RFR model prediction of the PKO yield is formulated as follows:

$$\hat{Y} = \frac{1}{T}\sum_{i=1}^{T} f_t(X_1, X_2, X_3) \quad (1)$$

Where, T denotes the number of decision trees in the forest, f_t denotes the prediction of the i^{th} tree, and X_i is the input parameter (main shaft, moisture content, and cone gap).

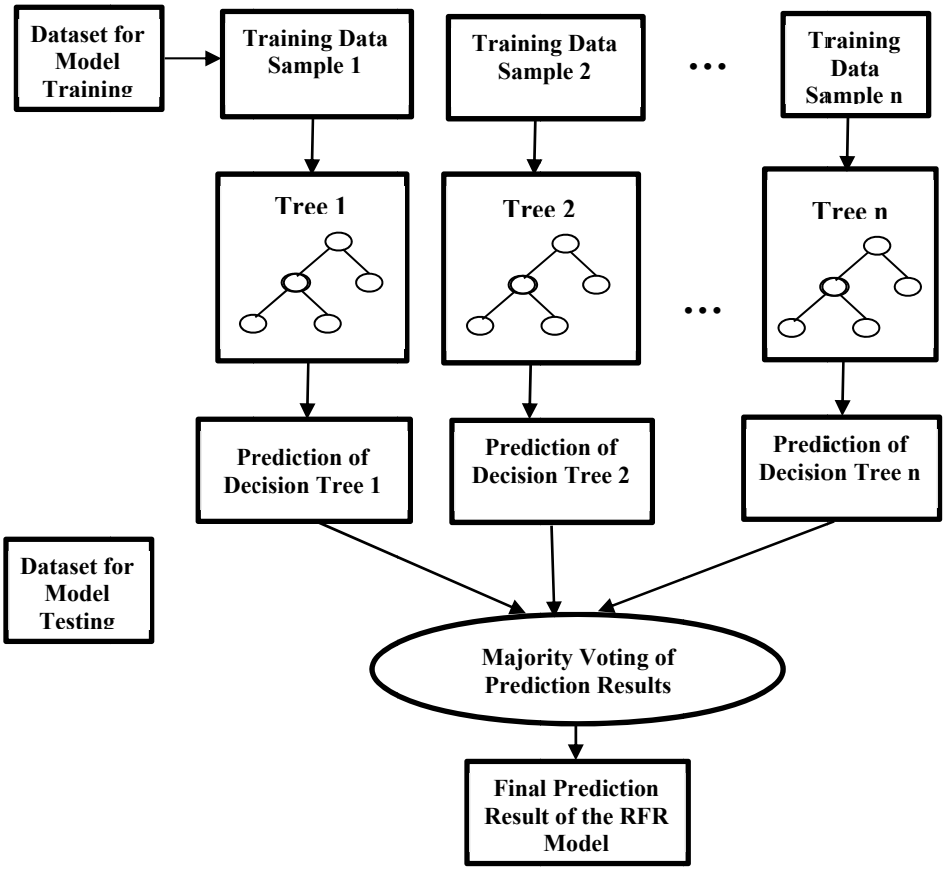

Figure 1 The Random Forest Regression Model Architecture

Each tree in the forest is a regression tree trained on a random subset of data and features. Given that for the i^{th} data input from the PKO extractor dataset, and if the actual oil yield from the extractor machine is denoted as Y_i , then a tree's prediction is calculated as:

$$f_t(X) = \frac{1}{N_l}\sum_{i\in L} Y_i \quad (2)$$

Where, leaf node denoted as L containing X, with N_l number of samples. The feature importance of each parameter X_j is calculated as:

$$I(X_j) = \frac{1}{T}\sum_{i=1}^{T} I_t(X_j) \quad (3)$$

Where, $I(X_j)$ is the importance of feature, in tree t, computing the total decrease in MSE contributed by X_j across all nodes where it was used for splitting. In the feature importance analysis, $I(X_1) > I(X_2) > I(X_3)$ can be used to obtain the feature importance score.. The model performance metrics used include **Mean Squared Error (MSE),** Mean Absolute Error (MAE) and R-Squared (R^2) value given as;

$$MSE = \left(\frac{\sum_{i=1}^{i=n}(x_{Pred(i)} - x_{Act(i)})^2}{n}\right) \quad (4)$$

$$MAE = \frac{\sum_{i=1}^{i=n}|x_{Pred(i)} - x_{Act(i)}|}{n} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{i=n}(x_{Act(i)} - x_{Pred(i)})^2}{\sum_{i=1}^{i=n}(x_{Act(i)} - x_{MeanAct})^2} \quad (6)$$

Where
n is number of data points,
Act indicates the original data
Pred indicates model predicted data
Mean indicates the average of the n data points

The following major steps are taken to implement the RFR model for the case study machine:

i. **Normalization of the Dataset:** To improve model convergence the dataset is normalized with MinMax approach.
ii. **Handling Missing Values:** Imputation with neighboring values approach was used

iii. **Train-Test Split:** 80% by 20% for training and validation respectively was used.

The hyperparameters adopted for the Random Forest performance include:

i. **Number of Trees (n_estimators):** typically 100 to 1000 for stability and reduced variance.

ii. **Maximum Depth (max_depth):** Controls overfitting. Optimal value found via cross-validation.

iii. **Minimum Samples Split (min_samples_split):** 2 or 5 to prevent splits on small sample sizes.

iv. **Minimum Samples Leaf (min_samples_leaf):** 1 to 4 is the best range

v. **Max Features (max_features):** Square root or log2 for reducing correlation between trees.

The summary of the Random Forest Regression model's hyperparameters and their values are presented in Table 1.

Table 1 The Hyperparameter settings for the Random Forest Regression Model

Hyperparameter	Value	Explanation
Number of trees	100	A standard choice; too low may cause underfitting, and too high increases training time.
Maximum depth	Null	Allows trees to expand fully unless constrained by min_samples_split or min_samples_leaf.
Minimum sample split	2	The default; ensures trees can split as long as at least two samples exist in a node.
Minimum sample leaf	1	Ensures leaf nodes can have a single sample; increasing it may reduce model complexity.
Maximum features	auto	-
Learning rate	0.05	-
Random state	42	-

3. Results and discussion

3.1 The Results of the impact of the inputs and the predicted oil yields for the Random Forest Regression (RFR)

The study utilized 5000 data records of a case study 10-ton PKO extractor machine in Uyo, Akwa Ibom State, Nigeria for the model training and validation. Also, SHAP feature importance approach was used to evaluate the importance ranking of each of the three input features to the RFR model. The results of the impact of the inputs parameters on the Random Forest Regression (RFR) model predicted output are shown in Figure 2. It shows that moisture content with feature importance ranking of 0.16 has the highest impact on the RFR model prediction while the Cone gap with feature importance ranking of 0.137 has the lowest impact.

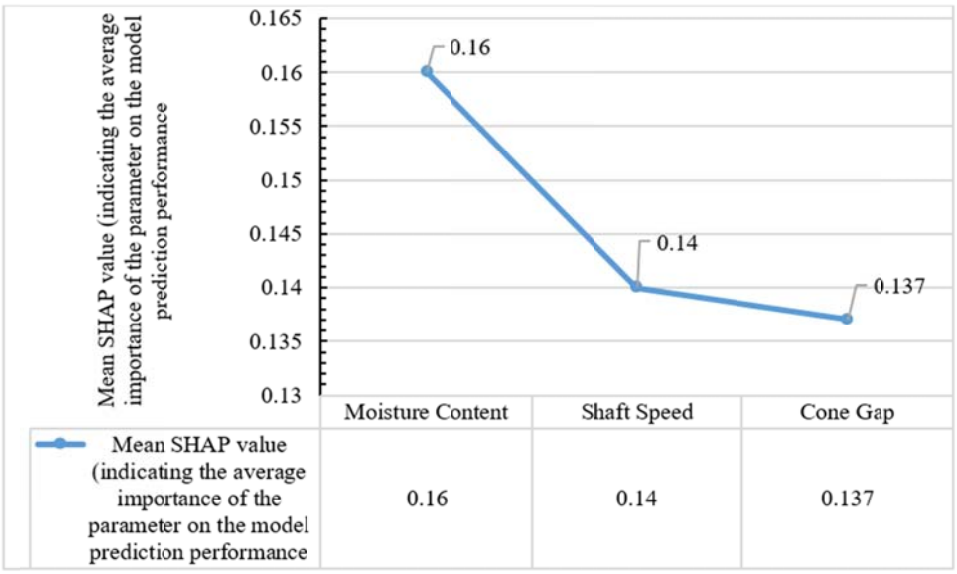

Figure 2: The impact of the inputs parameters on the Random Forest Regression (RFR) model predicted output

Also, the results of the error metrics over epochs for the RFR model are shown in Table 2 and Figure 3. The line chart of the actual versus predicted oil yields for the RFR model is shown in Figure 4. **T**he results show that the MAE =1.0678 x 10^{-15} and MSE = 2.14306 x 10^{-30} are very small (negligible) hence the coefficient of correlation between the actual and the predicted results is 1.

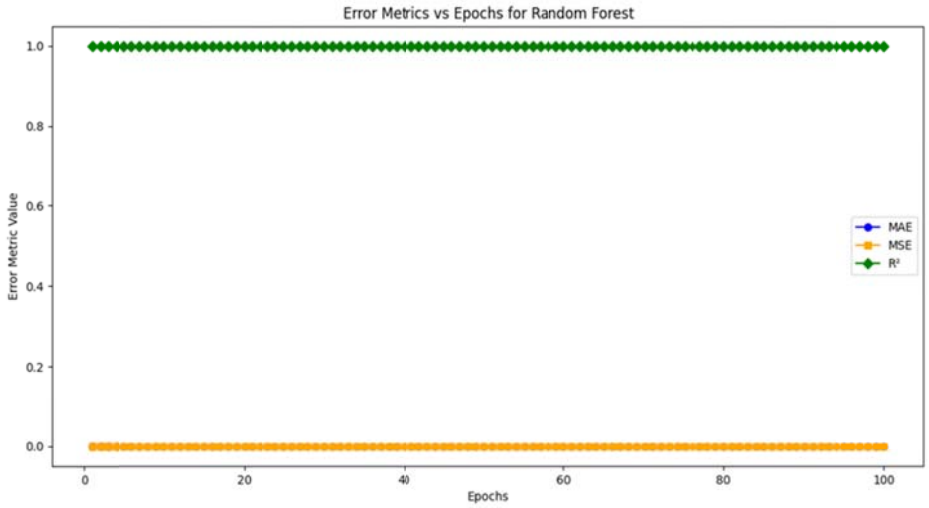

Figure 3 The Plot of the Error Metrics Over Epochs for the Random Forest Regression (RFR) model

Table 2: The Results of the Error Metrics over Epochs for the Random Forest Regression model

Epoch	MAE	MSE	R2
0	1.067178e-15	2.143306e-30	1.0
20	1.067178e-15	2.143306e-30	1.0
40	1.067178e-15	2.143306e-30	1.0
60	1.067178e-15	2.143306e-30	1.0
80	1.067178e-15	2.143306e-30	1.0
100	1.067178e-15	2.143306e-30	1.0

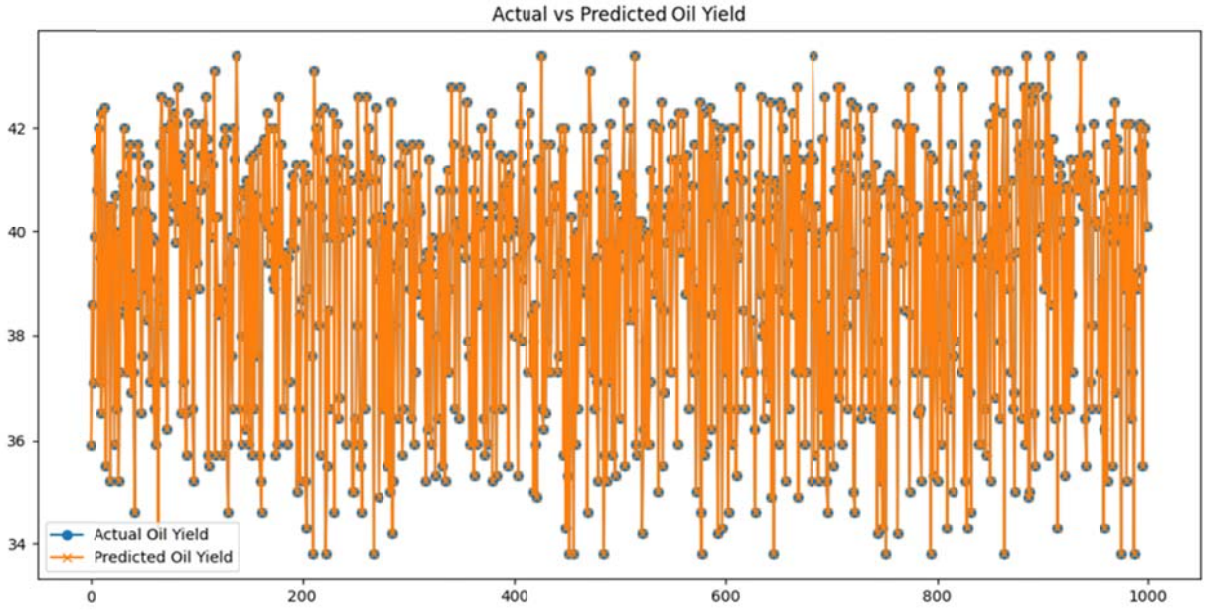

Figure 4: The Line Chart of the Actual Versus Predicted Oil Yields for the RFR Model

3.2 The Results of the Oil Yield for Various Input Variables Configurations for the Random Forest Regression (RFR) Model

The results of the oil yield for various input variables configurations for the RFR model are presented in Figure 5 to Figure 9. The results showed that the highest oil yield of 43.4 % occurred in Figure 7 with shaft speed of 18 rpm, cone gap of 1.5 mm and moisture content of 8 %. It means that for maximum PKO, the case study PKO extractor machine should be operated with the input settings as specified in Figure 7.

Further close examination of the optimal solution using graphical approach applied near the optimal point showed that the exact optimal oil yield based on graphical approach as shown in Figure 10 is PKO yield of 43.44 % at moisture content of 8.4 %. Hence, the RFR results need to be presented with at least two places of decimal to capture the exact optimal point as depicted in Figure 10.

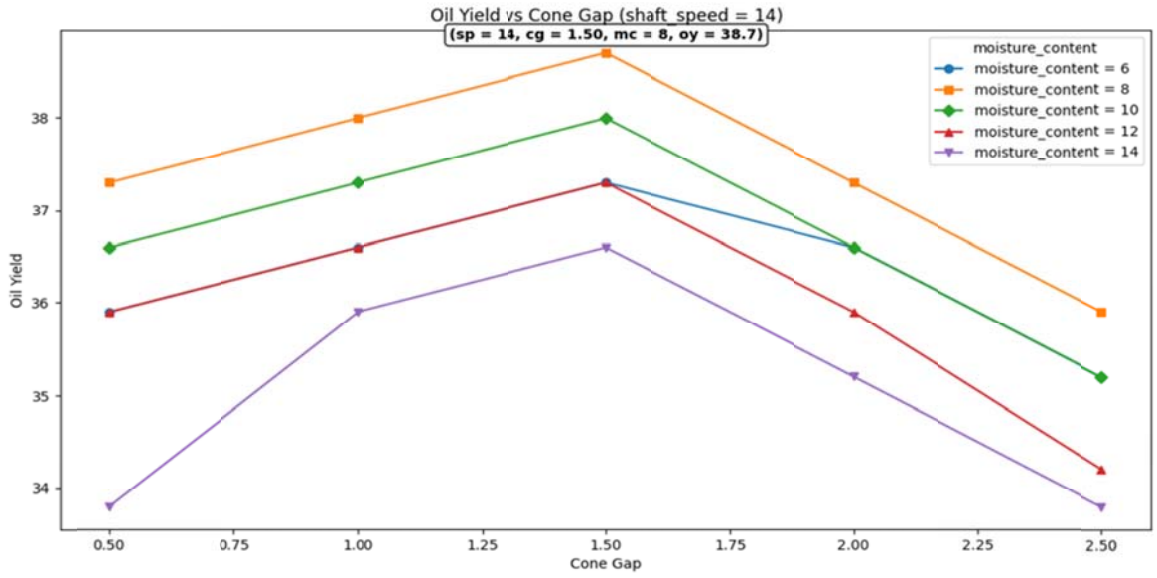

Figure 5: Oil yield versus cone gap at shaft speed = 14rpm and varying moisture content

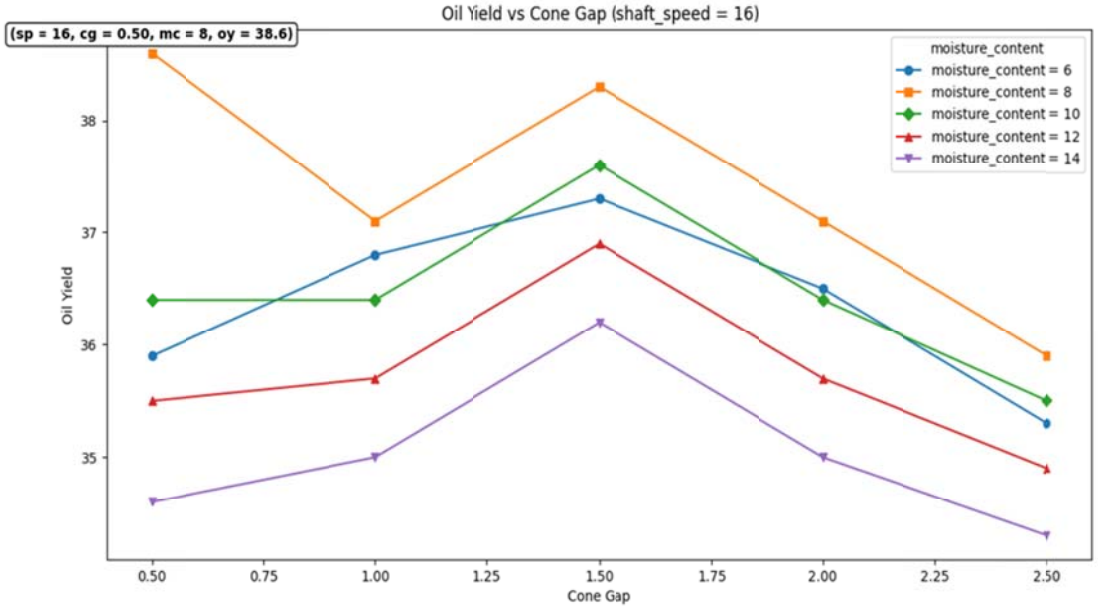

Figure 6: Oil yield versus cone gap at shaft speed = 16rpm and varying moisture content

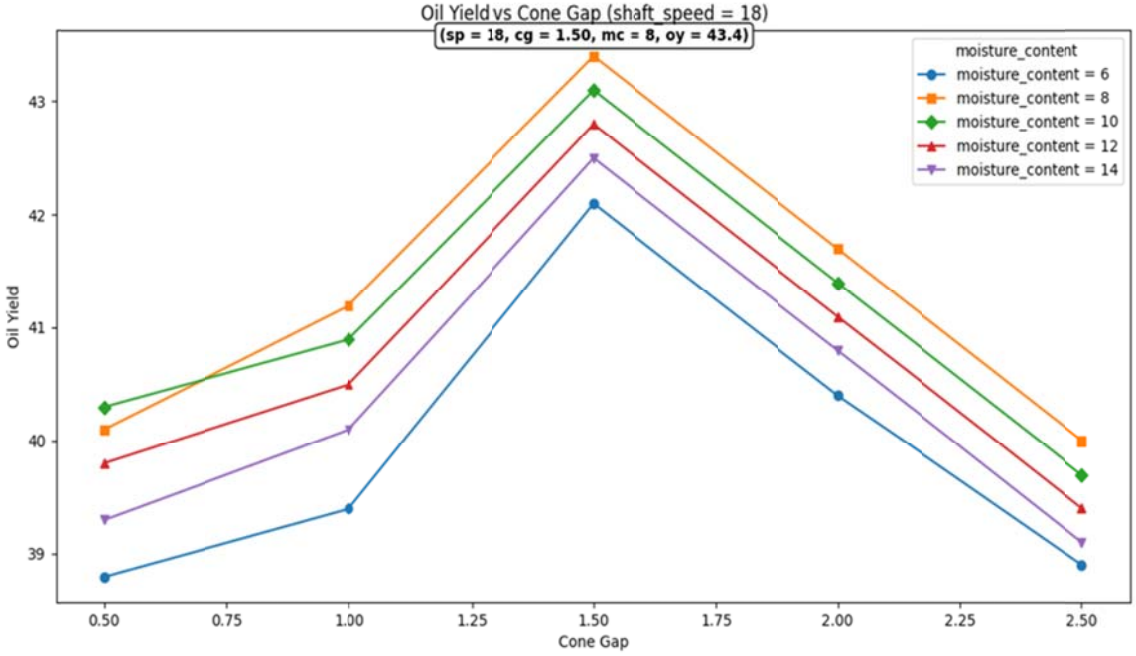

Figure 7: Oil yield versus cone gap at shaft speed = 18rpm and varying moisture content

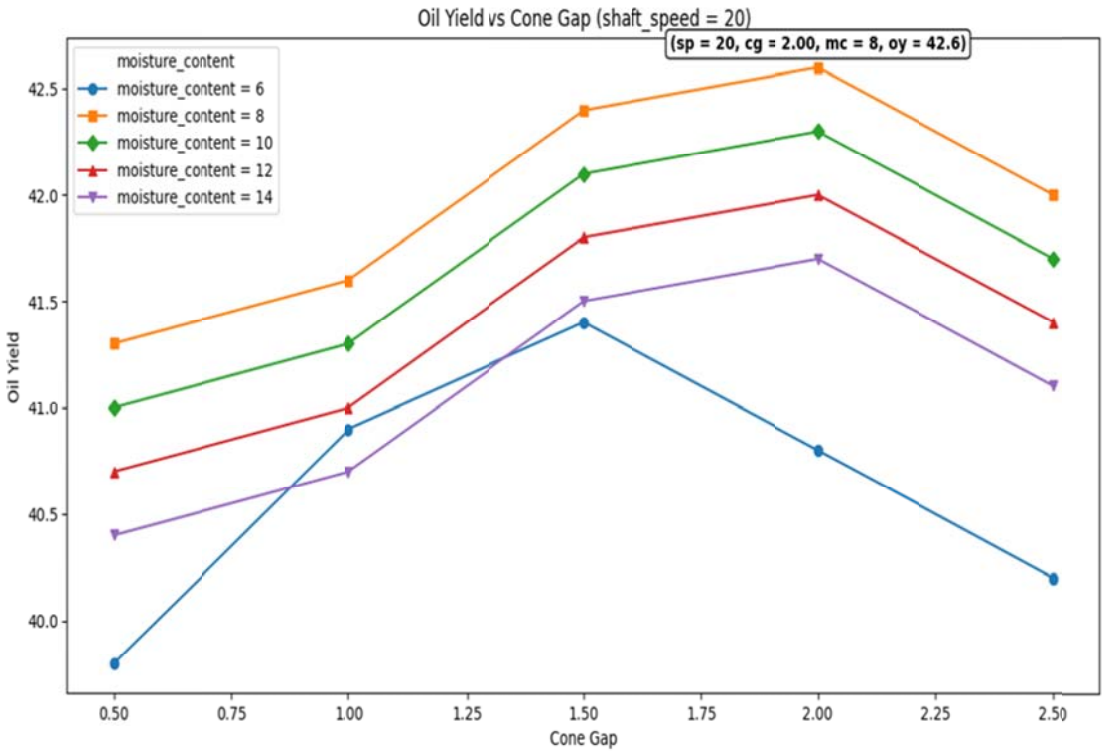

Figure 8: Oil yield versus cone gap at shaft speed = 20rpm and varying moisture content

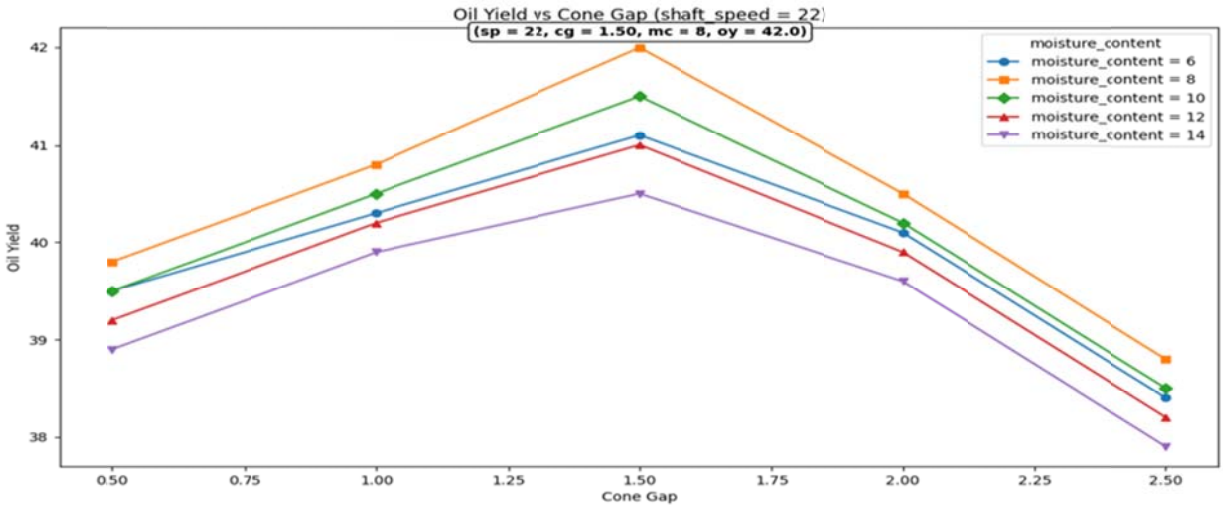

Figure 9: Oil yield versus cone gap at shaft speed = 22rpm and varying moisture content

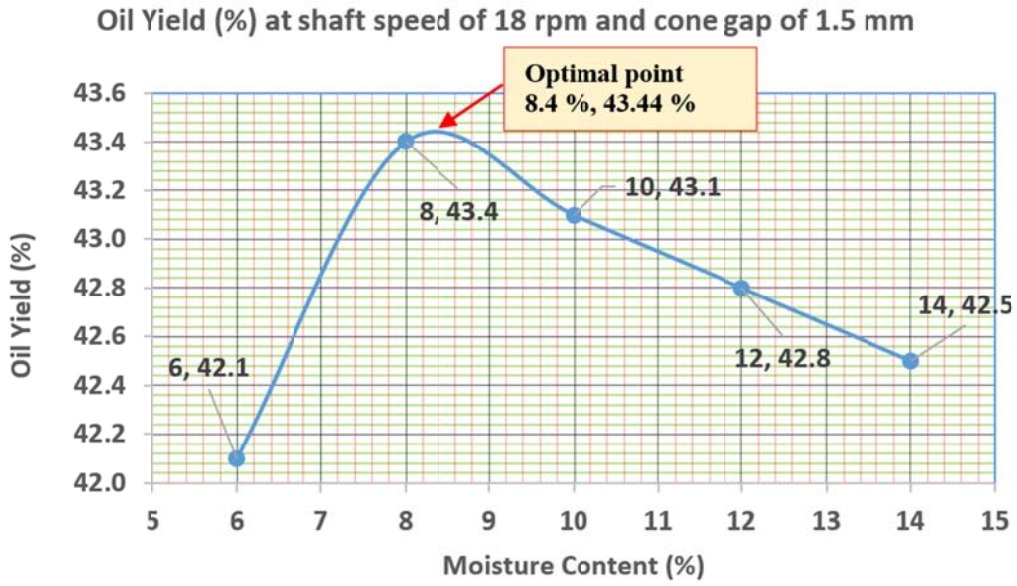

Figure 10 The exact optimal oil yield based on graphical approach

4. Conclusion

The optimal Palm Kernel Oil (PKO) yield for a PKO extractor machine is studied and Random Forest Regression (RFR) model is applied to determine the input configuration that will give the optimal PKO yield. The case study extractor machine in Uyo was considered with data on three input data parameters, namely; the moisture content, the cone gap setting and the machine shaft speed. The feature importance of the parameters was also considered and the moisture content had the highest feature importance score. The RFR model was able to pinpoint the exact optimal settings and the RFR model result was further enhanced using the graphical method.

References

1. Bharadiya, J. P., Thomas, R. K., & Ahmed, F. (2023). Rise of artificial intelligence in business and industry. *Journal of Engineering Research and Reports*, *25*(3), 85-103.
2. Mashood, K., Kayani, H. U. R., Malik, A. A., & Tahir, A. (2023). Artificial intelligence recent trends and applications in industries. *Pakistan Journal of Science*, *75*(02).
3. Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN computer science*, *3*(2), 158.
4. Parihar, V., Malik, A., Bhawna, Bhushan, B., & Chaganti, R. (2023). From smart devices to smarter systems: The evolution of artificial intelligence of things (AIoT) with characteristics, architecture, use cases and challenges. In *AI models for blockchain-based intelligent networks in IoT systems: Concepts, methodologies, tools, and applications* (pp. 1-28). Cham: Springer International Publishing.
5. Elahi, M., Afolaranmi, S. O., Martinez Lastra, J. L., & Perez Garcia, J. A. (2023). A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. *Discover Artificial Intelligence*, *3*(1), 43.
6. Agbehadji, I. E., Awuzie, B. O., Ngowi, A. B., & Millham, R. C. (2020). Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of COVID-19 pandemic cases and contact tracing. *International journal of environmental research and public health*, *17*(15), 5330.
7. Ravichandran, P., Machireddy, J. R., & Rachakatla, S. K. (2023). Data analytics automation with AI: a comparative study of traditional and generative AI approaches. *Journal of Bioinformatics and Artificial Intelligence*, *3*(2), 168-190.
8. Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of management analytics*, *6*(1), 1-29.
9. Baker, N., Alexander, F., Bremer, T., Hagberg, A., Kevrekidis, Y., Najm, H., ... & Lee, S. (2019). *Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence*. USDOE Office of Science (SC), Washington, DC (United States).
10. Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2022). Artificial intelligence applications for industry 4.0: A literature-based study. *Journal of Industrial Integration and Management*, *7*(01), 83-111.
11. Plathottam, S. J., Rzonca, A., Lakhnori, R., & Iloeje, C. O. (2023). A review of artificial

intelligence applications in manufacturing operations. *Journal of Advanced Manufacturing and Processing*, *5*(3), e10159.

12. Mathew, D., Brintha, N. C., & Jappes, J. W. (2023). Artificial intelligence powered automation for industry 4.0. In *New horizons for Industry 4.0 in modern business* (pp. 1-28). Cham: Springer International Publishing.
13. He, M., Li, Z., Liu, C., Shi, D., & Tan, Z. (2020). Deployment of artificial intelligence in real-world practice: opportunity and challenge. *Asia-Pacific Journal of Ophthalmology*, *9*(4), 299-307.
14. Shaw, J., Rudzicz, F., Jamieson, T., & Goldfarb, A. (2019). Artificial intelligence and the implementation challenge. *Journal of medical Internet research*, *21*(7), e13659.
15. Paleyes, A., Urma, R. G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: a survey of case studies. *ACM computing surveys*, *55*(6), 1-29.
16. Khan, N. (2023). *Prediction Of Oil Palm Yield For Smallholders Estates In Tropical Region Using Extra Trees Method* (Doctoral dissertation).
17. SULAIMAN, N. S. B. (2021). DATA-DRIVEN MODELLING AND OPTIMIZATION OF PALM OIL REFINING PROCESS.
18. Tachie, C. Y. E. (2023). *Development and Comparison of Machine Learning Models for Predicting Fatty Acid Classes in Snacks and Authenticating Oils and Margarine* (Master's thesis, Delaware State University).