

Intelligent Maize Yield Prediction Model Based on Plant Attributes and Machine Learning Algorithms

Oyenike Mary Olanrewaju, Eli Adama Jiya, *Faith Oluwatosin Echobu

Faculty of Computing, Federal University Dutsin-Ma, Katsina State, Nigeria.

*Corresponding Author

DOI: <https://doi.org/10.51244/IJRSI.2024.1107087>

Received: 04 July 2024; Accepted: 16 July 2024; Published: 21 August 2024

ABSTRACT

Agriculture is a vital component of the Nigerian economy. The sector is a major source of employment for a large number of Nigerians. Maize is a widely planted crop and consumed in Nigeria, especially in the northern part of the country, with many poor families relying on it as the major source of carbohydrates. Therefore, sufficient provision of the crop is very vital, and prediction of the yield is very essential for proper planning in case of crop failure. This research developed three machine learning models for predicting maize yield using Random Tree, Random Forest and Neural Networks. The work made use of maize yield data from an experimental farm of Federal University Dutsin-ma, Katsina state. From the performance evaluation of the models, the Random Tree model demonstrated better performance than other models. It achieved the lowest MAE, RMSE, RAE, and RRSE values of 0.093, 0.096, 19.7%, and 19.2% respectively. This result indicates a lower error rate and a higher accuracy of almost 80% in predicting the numerical value of the weight of the maize yield. It is recommended that the model here be used to predict future maize yield in the state for proper planning and to ensure food security for the people of the state who are major maize consumers.

INTRODUCTION

Agriculture is a critical component of Nigeria's economy. The sector is a major source of employment for a large number of Nigerians (Cedric et al., 2022; Ebele & Emodi, 2016) and is also considered vital to many world economies (Tandzi & Mutengwa, 2020). While a good number of Nigerians engage in other forms of agriculture like fish farming and animal husbandry, crop production is by far the largest form of agriculture being practiced in the country (National Bureau of Statistics (NBS), 2018); therefore, failures of crop yield affect household incomes and the nation's food security. The recent food inflation which has led to negative responses from the citizens evidenced by protests in various major cities of Nigeria underscores the importance of food security. As successive governments strive to improve internal food production, maize is considered a crop to focus on due to its wide usage for industrial and human consumption (Falade & Labaeka, 2020).

Nigeria is considered the largest producer of maize in Africa, with yields estimated to be 2.2 tons per hectare (Falade & Labaeka, 2020). Maize is a widely planted crop and consumed in Nigeria, especially in the northern part of the country, with many poor families relying on it as the major source of carbohydrates in their diets (Adamgbe & Ujoh, 2013; Wossen et al., 2023). Therefore, sufficient provision of the crop is very vital and prediction of the yield is essential for proper planning in case of crop failure (Akinbile et al., 2020; Edeh & Eboh, 2011; Jiya et al., 2023; Oluwaseyi et al., 2016; Tiamiyu et al., 2015). Maize yield prediction is influenced by several complex and interacting factors that make it a difficult challenge to address (Khaki & Wang, 2019). This challenge has resulted in several approaches by many researchers to

develop actionable maize yield prediction models.

Maize prediction models rely on several methods like subjective judgment by extension workers or counting of grains from samples of the cobs (Tandzi & Mutengwa, 2020), while some methods use process or data-driven methods (Badmus et al., 2011; Chang et al., 2020). Some works (Kumar et al., 2022; Tandzi & Mutengwa, 2020) used formulas to calculate the yield as stated below in equations 1 to 3:

$$\text{Yield} = \text{average number of yield} \times \text{average gain weight per cob} \quad (1)$$

$$\text{Yield (kg/ha)} = \left[\frac{\text{number of kernel rows per ear} \times \text{number of ears per m}^2}{100} \right] \times (\text{weight of 1000-kernel (g)/1000}) \times 10,000 \quad (2)$$

$$\text{Yield (t ha}^{-1}\text{)} = a * \text{NDVI} - b \quad (3)$$

where, NDVI_m = Seasonal maximum NDVI of maize crop, a = 32.37, b = 17.61

These formulas directly measure or estimate maize yield from crop parameters. However, there is no single formula or agreement on the parameters of the independent variables for maize yield estimations. The common features appear to be the estimations of yield directly from plant properties without the consideration of environment variables.

Machine learning models, which move away from direct mathematical formulae are another approach by some authors. These models which rely on experimental or historical data and machine learning algorithms to develop yield prediction are reputed for their high accuracy. The works of Paula et al., (2020) used the random forest algorithm to predict maize crop yield with a rank-based approach. The study used a dataset composed of 33 Vegetation indices extracted from multi-spectral UAV imagery with a result of a Mean Absolute Error of 0.78.

Oiganji et al., (2016) developed a maize yield model using grain yield using the AquaCrop model, the plant biomass and the plant water were used as inputs to the model. The model achieved 86% prediction accuracy. Chang et al., (2020) developed a maize yield prediction model from the daily biomass dataset of maize plants. The data was directly taken from the experimental maize farms. The result of the performance evaluation of the work was 7.16. Khaki and Wang, (2019) used ANN to design a model to predict maize yield from a dataset of 2,267 maize records. The model recorded an accuracy of 12% RMSE and 50% of the standard deviation.

A study of the prediction of multiple crop yield was conducted by (Paudel et al., 2021) using Gradient Boosting, Support Vector Regression (SVR), and k-Nearest Neighbour models. The input features for the prediction models were a combination of weather, remote, and soil data.

With the current challenges and the drive to increase domestic production of maize to meet consumption and industrial usage, it is essential to develop maize yield prediction models based on the plant's features. This allows for predictions based on the plant rather than external environmental variables. Using multiple machine learning models to help stakeholders make proper estimates of farm output (Westerveld et al., 2021). This work primarily used data from farms in Katsina state. The state is located in the northwest region of Nigeria, with a short duration of rainfall and unfavourable climatic conditions (Ebenezer, 2015). Therefore, there is a need for household food security planning that can be driven by accurate models.

METHODOLOGY

This paper adopted a data mining methodology to develop a maize yield prediction model. The steps in the methodology include collection of data from the farm, preprocessing of the collected data, and development of the model using machine learning accuracy checks.

Data Source and Attributes

The study location is Katsina state in northern western Nigeria, and the plant data was collected from the Federal University Dutsin-Ma farm while the environmental data was collected from the Ministry of Agriculture for a single year. The data collected from different plots of the farm in the year 2023 farming season has the following attributes: plant height (cm), plant stem diameter (mm), leaf area (cm²), number of leaves, number of harvested maize, grains weight (kg), average number of grains/cob, number of seed in row, 100 seed weight (g), cob diameter (mm), cob length (cm), dry plant weight (g), and weight of harvested maize (kg).

Features Selection

The key to having an accurate model is the selection of proper attributes that will accurately model the system. To improve the accuracy of our model, this work used all the plant attributes earlier introduced, this was done to improve accuracy beyond the works of (Kumar et al., 2022; Tandzi & Mutengwa, 2020) that used on two or three attributes. Tables 1 and 2 show a sample of the data.

Table 1: sample of the data

| Plant height (cm) | Plant Stem Diameter (mm) | Leaf Area (cm ²) | Number of Leaf | Number of Harvested Maize | Grains Weight (kg) | Aver No of grains/cob |
|-------------------|--------------------------|------------------------------|----------------|---------------------------|--------------------|-----------------------|
| 170.67 | 12.37 | 413.83 | 11 | 32 | 0.71547 | 356.4 |
| 156.3 | 12.6 | 412.47 | 13 | 40 | 0.75675 | 315.2 |
| 180.3 | 12.16 | 616.17 | 12 | 18 | 0.26176 | 288.2 |
| 148 | 16.73 | 367 | 12 | 20 | 0.24281 | 210 |
| 161.3 | 10.7 | 416 | 13 | 14 | 0.22533 | 263.8 |

Table 2: sample of the data

| Number of seed in row | 100 seed weight (g) | Cob Diameter (mm) | Cob Length (cm) | Dry Plant Weight (g) | Weight of Harvested Maize (kg) |
|-----------------------|---------------------|-------------------|-----------------|----------------------|--------------------------------|
| 13.2 | 13.72 | 36.4 | 11.32 | 200.76 | 0.94745 |
| 13.6 | 13.94 | 37.42 | 9.7 | 160.37 | 1.02021 |
| 13.6 | 12.2 | 33.64 | 10.92 | 208.07 | 0.37219 |
| 13 | 11.83 | 32.82 | 8.78 | 167 | 0.35293 |
| 13.4 | 9.6 | 30.02 | 9.84 | 137.27 | 0.31143 |

Data Preprocessing

The initial data had the weight of the harvested maize measured in grams; therefore, further processing was performed to convert the data to kilograms to aid model development. The data was divided into 75% for training and 25% for testing.

Algorithms

The research utilised three machine learning algorithms for the yield prediction. These included Random Forest (RF), Random Trees (RT), and Artificial Neural Network (ANN).

RT is a variant of tree-based algorithms and data structure. the RT formation began by constructing several subtrees. The subtrees are formed by randomly sampling the input data and using it as the features or nodes (Gupta et al., 2016). The general theory is that each tree has a probability of being sampled.

The Random Forest algorithm is a tree-based technique that generates prediction trees using the attributes of the systems being modelled (Guo et al., 2016). This technique uses random sampling of the entire dataset to generate numerous subtrees which are further merged as the final solution.

ANN is a computing algorithm which excels in processing large volumes of data and providing graphic-like connections between the various components of a system. This element of Neural networks was a concept directly modelled from the pattern that the human brain uses to process data (Alanazi et al., 2021; He et al., 2018; Okewu et al., 2019).

ANN consists of nodes which receive input and a processing layer that sums up the input to produce the output. All the nodes are connected using weighted edges. Though several topologies of ANN exist, however, all the topologies have 3 basic layers which map input to output to find patterns in the training data.

Metrics for Performance Evaluation

For each of the model developed in this paper, their accuracy or performance evaluation was done using Root Mean Square Error (RMSE), Root-Relative Square Error (RRSE), Root Absolute Error (RAE) and Mean Absolute Error (MAE) to assess their efficacy. Some of the errors can be mathematically expressed as follows:

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{2}$$

RESULT AND DISCUSSION

This section provides the output from various maize yield prediction models developed. Figures 1 and 2 below are the graphical representations of two of the predictive maize yield models.

Figure 1: Random Tree Maize Yield model

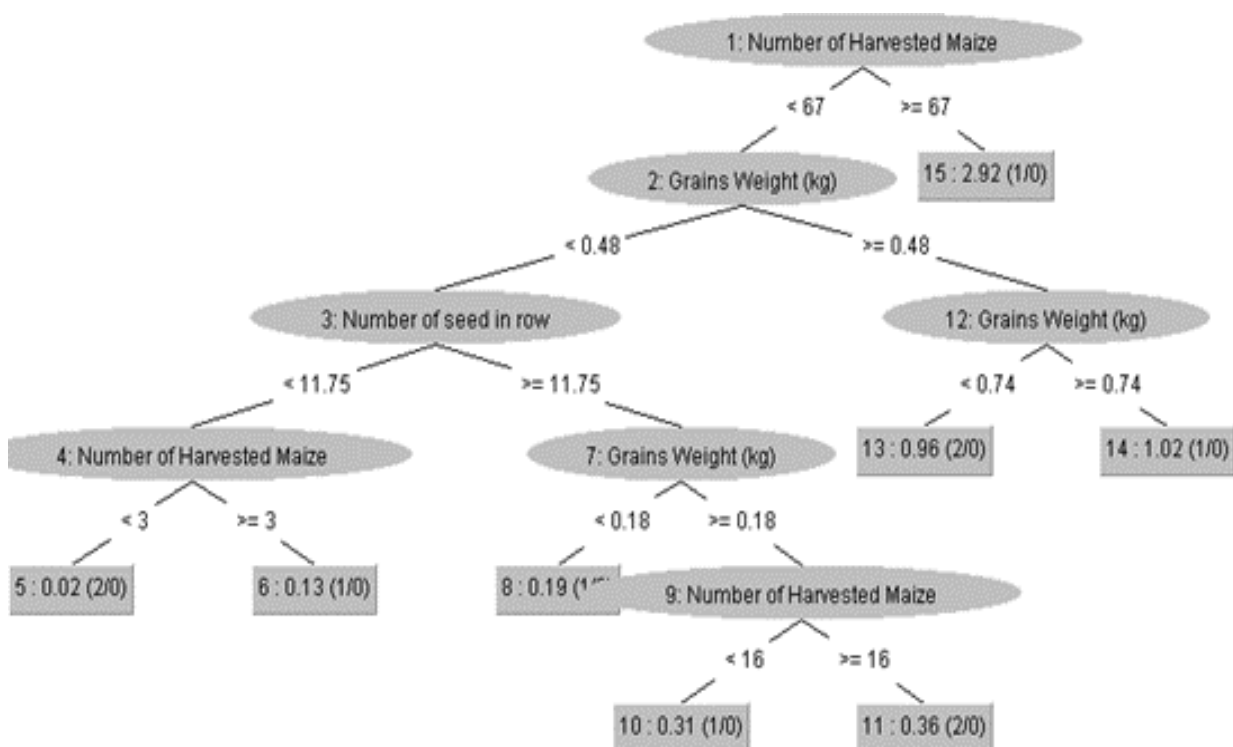


Figure 2: ANN Maize Yield Model

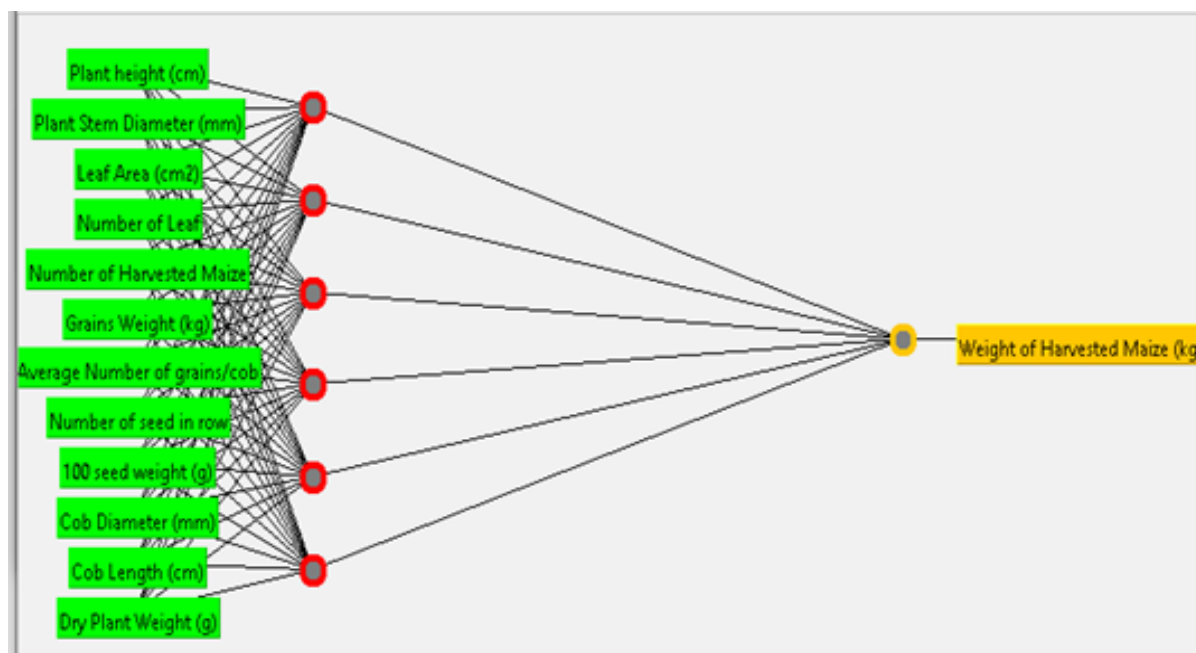


Figure 1 represents the maize yield forecasted using RT algorithm. The model uses various nodes for stem height, leaf size, number of leaves, number of harvested maize, weight of the grains and other essential parameters from the plant to predict the weight of harvested maize. The parameters are used as the leaves of the tree. Also, the ANN model in Figure 2 uses various nodes of the stem height, leaf size, number of leaves, number of harvested maize, weight of the grains and other essential parameters from the plant to predict the weight of harvested maize. In the case of ANN, the parameters form the input nodes of the ANN

Model Accuracy Measure

Three models were developed in this research, Table 3 provides a summary of all the models with their various accuracy measures.

Table 3: Error measure of the models

| | MAE | RMSE | RAE | RRSE |
|------------|-------|-------|--------|--------|
| ANN | 0.3 | 0.33 | 63.75% | 67.29% |
| RF | 0.15 | 0.17 | 33.05% | 35.3% |
| RT | 0.093 | 0.096 | 19.7% | 19.2% |

The accuracy metrics MAE, RMSE, RAE, and RRSE were used to evaluate the prediction models performances in predicting the values of yield output of maize.

From Table 1, ANN shows model error values in terms of MAE, RMSE, RAE and RRSE values as 0.3, 0.33, 63.75%, and 67.29% respectively. RF shows model error values in terms of MAE, RMSE, RAE and RRSE values as 0.15, 0.17, 33.05%, and 35.3% respectively. The last model RT shows values of MAE, RMSE, RAE and RRSE as 0.093, 0.096, 19.7%, and 19.2%.

Discussion of Result

This research investigated the capacities of three machine learning algorithms to predict maize yield based on plant features. The developed models were each evaluated based on performance metrics of MAE,

RMSE, RAE, and RRSE. From the analysis of the metrics, RT demonstrated better performance than the other model results. It achieved the lowest MAE, RMSE, RAE, and RRSE values of 0.093, 0.096, 19.7%, and 19.2% respectively. This result indicates a lower error rate and a higher accuracy of almost 80% in predicting the numerical value of the weight of the maize yield. The RF model follows closely by having performance results with MAE, RMSE, RAE, and RRSE values of RF 0.15, 0.17, 33.05%, and 35.3% respectively. Though the error rate is higher than that of RF, the result of 35.3% RRSE shows that the model is still reliable. ANN performed with, MAE, RMSE, RAE, and RRSE values of RF 0.3, 0.33, 63.75%, 67.29% respectively. This result shows that ANN has a higher error rate in RRSE of 67.29%, showing a result of poor performance in predicting maize yield using parameters from the plant.

The general interpretation of the performance evaluation of the models suggests RT and RF are able to predict maize yield with more than 70% accuracy and therefore are better models than for predicting maize yield. The evaluation of the ANN model suggests that neural network models are too weak to be used for the prediction of maize yield in this case.

In comparison with the work of Paula et al., (2020) which developed a model for maize yield prediction using RF, the model in this research performed better with lower error, MAE of 0.15 as against 0.78.

CONCLUSION

This paper developed three machine learning models for predicting maize yield using Random Tree, Random Forest and Neural Networks. From the performance evaluation of the models, the Random Tree model demonstrated better performance than other models. It achieved the lowest MAE, RMSE, RAE, and RRSE values of 0.093, 0.096, 19.7%, and 19.2% respectively. This result indicates a lower error rate but a higher accuracy of almost 80% in predicting the numerical value of the weight of the maize yield. It is recommended that the model be used to predict future maize yield in the state for proper planning and to ensure food security for the people of the state who are major maize consumers.

ACKNOWLEDGEMENT

This research was funded by the Tertiary Education Trust Fund (TETFUND). We extend our heartfelt appreciation to TETFUND for their commitment to advancing educational research and development.

REFERENCES

1. Adamgbe, E. M., & Ujoh, F. (2013). Effect of Variability in Rainfall Characteristics on Maize Yield in Gboko , Nigeria. *Journal of Environmental Protection*, 2013(September), 881–887.
2. Akinbile, C. O., Ogunmola, O. O., Abolude, A. T., & Akande, S. O. (2020). Trends and spatial analysis of temperature and rainfall patterns on rice yields in Nigeria. *Atmospheric Science Letters*, 21 (3), 1–13. <https://doi.org/10.1002/asl.944>
3. Alanazi, S. A., Kamruzzaman, M. M., Sarker, N. I., Alruwaili, M., Alhwaiti, Y., Alshammari, N., & Siddiqi, M. H. (2021). Boosting Breast Cancer Detection Using Convolutional Neural Network. *Journal of Healthcare Engineering*, 2.
4. Badmus, M. A., Ariyo, O. S., & Ishin, J. I. (2011). Forecasting Cultivated Areas and Production of Maize in Nigerian using ARIMA Model. *Asian Journal of Agricultural Sciences*, 3(3), 171–176.
5. Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., & Kimpolo, C. L. M. (2022). Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*, 2(December 2021), 100049. <https://doi.org/10.1016/j.atech.2022.100049>
6. Chang, Y., Latham, J., Licht, M., & Wang, L. (2020). A data-driven crop model for maize yield prediction. *COMMUNICATIONS BIOLOGY*, 2023. <https://doi.org/10.1038/s42003-023-04833-y>

7. Ebele, N., & Emodi, N. (2016). Climate Change and Its Impact in Nigerian Economy. *Journal of Scientific Research and Reports*, 10(6), 1–13. <https://doi.org/10.9734/jsrr/2016/25162>
8. Ebenezer, T. (2015). Drought , desertification and the Nigerian environment : A review. *Journal of Ecology and the Natural Environment Review*, 7(7), 197–209. <https://doi.org/10.5897/JENE2015>.
9. Edeh, H. O., & Eboh, E. C. (2011). *Analysis of Environmental Risk Factors Affecting Rice Farming in Ebonyi State , Southeastern Nigeria*. 7(1), 100–103.
10. Falade, A. A., & Labaeka, A. (2020). A review of production constraints confronting maize production. *African Journal of Sustainable Agricultural Development*, 1(4), 11–20. <https://doi.org/10.46654/2714>
11. Guo, F., Wang, G., Su, Z., Liang, H., Wang, W., Lin, F., & Liu, A. (2016). What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests. *International Journal of Wildland Fire*, 25(5), 505–519. <https://doi.org/10.1071/WF15121>
12. Gupta, S., Abraham, S. K., Sugumaran, V., & Amarnath, M. (2016). Fault Diagnostics of a Gearbox via Acoustic Signal using Wavelet Features, J48 Decision Tree and Random Tree Classifier. *Indian Journal of Science and Technology*, 9(33), 1–8. <https://doi.org/10.17485/ijst/2016/v9i33/101328>
13. He, L., Li, H., Holland, S. K., Yuan, W., Altaye, M., & Parikh, N. A. (2018). NeuroImage : Clinical Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework. *NeuroImage: Clinical*, 18(October 2017), 290–297. <https://doi.org/10.1016/j.nicl.2018.01.032>
14. Jiya, E. A., Illiyasu, U., & Mudashiru, A. (2023). *Rice Yield Forecasting : A Comparative Analysis of Multiple Machine Learning Algorithms*. 5(2), 785–799. <https://doi.org/10.51519/journalisi.v5i2.506>
15. Khaki, S., & Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks. *Plant Sci.*, 10 (May), 1–10. <https://doi.org/10.3389/fpls.2019.00621>
16. Kumar, D. A., Neelima, T. L., Srikanth, P., Devi, M. U. M. A., Suresh, K., & Murthy, C. S. (2022). Maize yield prediction using NDVI derived from Sentinel 2 data in Siddipet district of Telangana state. *Journal of Agrometeorology*, 1665(24), 165–168.
17. National Bureau of Statistics (NBS). (2018). *Demographic Statistics Bulletin*. National Bureau of Statistics.
18. Oiganji, E., Igbadun, H. E., Mudiare, O. J., & Oyebode, M. A. (2016). *Calibrating and validating AquaCrop model for maize crop in Northern zone of Nigeria*. 18(3), 1–13.
19. Okewu, E., Misra, S., Sanz, L. F., Ayeni, F., Mbarika, V., & Damaševičius, R. (2019). Deep Neural Networks for Curbing Climate Change-Induced Farmers-Herdsmen Clashes in a Sustainable Social Inclusion Initiative. *Problems of Sustainable Development*, 14(2), 143–155.
20. Oluwaseyi, A., Nehemiah, D., & Zuluqurineen, S. (2016). Genetic Improvement of Rice in Nigeria for Enhanced Yield and Grain Quality – A Review. *Asian Research Journal of Agriculture*, 1(3), 1–18. <https://doi.org/10.9734/arja/2016/28675>
21. Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine Learning for large-scale crop yield forecasting. *Agricultural Systems*, 187. <https://doi.org/https://doi.org/10.1016/j.agsy.2020.103016>
22. Paula, A., Ramos, M., Prado, L., Elis, D., Furuya, G., Nunes, W., Cordeiro, D., Pereira, L., Teodoro, R., Antonio, C., Capristo-silva, G. F., Li, J., Henrique, F., Baio, R., Marcato, J., Eduardo, P., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, 178(July), 105791. <https://doi.org/10.1016/j.compag.2020.105791>
23. Rugimbana, C. (2019). *Predicting Maize (Zea Mays) Yields in Eastern Province of Rwanda Using AquaCrop Model*. University of Nairobi.
24. Tandzi, L. N., & Mutengwa, C. S. (2020). Estimation of Maize (Zea mays L .) Yield Per Harvest Area : Appropriate Methods. *Agronomy*, 10(19), 1–18.
25. Tiamiyu, S. A., Eze, J. N., Yusuf, T. M., Maji, A. T., & Bakare, S. O. (2015). Rainfall Variability and its Effect on Yield of Rice in Nigeria. *International Letters of Natural Sciences*, 49(November),

- 63–68. <https://doi.org/10.18052/www.scipress.com/ilns.49.63>
26. Westerveld, J. J. L., van den Homberg, M. J. C., Nobre, G. G., van den Berg, D. L. J., Teklesadik, A. D., & Stuit, S. M. (2021). Forecasting transitions in the state of food security with machine learning using transferable features. *Science of the Total Environment*, 786(May), 147366. <https://doi.org/10.1016/j.scitotenv.2021.147366>
27. Wossen, T., Menkir, A., Alene, A., Abdoulaye, T., Ajala, S., Badu-apraku, B., Gedil, M., Mengesha, W., & Meseka, S. (2023). Drivers of transformation of the maize sector in Nigeria. *Global Food Security*, 38(January), 100713. <https://doi.org/10.1016/j.gfs.2023.100713>