International Conference on Sciences, Engineering & Environmental Technology (ICONSEET), 10(15): 94-100, 2023

ISSN 0749-9650

www.repcomseet.org

Crop Yield Prediction Using Machine Learning: A Comparative Analysis and Review of Classification Algorithms

Ramoni Tirimisiyu Amosa*, Fabiyi Aderanti Alifat, Etudaiye Abdulmumuni Isah, Lawal Moshood Olatunji, Olorunlomerue Adam Biodun, Oni Esther Kemi, Adejola Aanu Adeyinka.

Department of Computer Science, Federal Polytechnic, Ede, Osun State, Nigeria.
Correspondence Author: * ramoni.tirimisiyu@federalpolyede.edu.ng;

Abstract- *Global civilization began with agricultural practises. Although human lifestyles have evolved over time, agriculture continues to play a significant role in providing for fundamental requirements and basic needs. Agriculture contributes significantly to nation-building beyond just feeding people. One such contribution is the growth of foreign reserves via the sale of agricultural goods and products. Predicting and estimating agricultural production is necessary because of the rapidly changing climate in our surroundings, which negatively affects most crops at harvest time. The forecast assists farmers in selecting the best crop to plant in order to maximise yield by examining variables such as soil quality, temperature, humidity, rainfall, and field area. Crop production forecasts make it easier for farmers to plan all aspects of sale and storage. This paper reviews common, state of earth, effective and recently used prediction and classification algorithm for crop yield.*

Keywords*: Agriculture, Algorithm, Crop, Farmer, Forecast, Prediction.*

1. Introduction

Crop yield is the amount of crop harvested per unit area of land. Crop factors, soil, weather, and environment all play a role in predicting crop production (Pant et al., 2021; Oluwole et al., 2022) Among the most significant and ancient industries in the world is agriculture. A few of the many issues facing the agriculture industry are population expansion, climate change, and the need to increase the productivity of food production. The forecasting of crop yields is an approach or methodology used to identify the many types of crops that must be grown in order to maximise productivity, maximise profit, and reduce soil pollution (Gupta et al., 2023). But with the advent of machine learning (ML) technology, farming is becoming more productive, efficient, and sustainable, completely changing the way farmers work Elijah et al. (2022). Agriculture's digital revolution has led to the evolution of management practises into artificial intelligence systems. This is done in an effort to extract value from the vast amounts of data coming from a variety of sources. Machine learning, a subset of artificial intelligence, holds great promise for addressing many obstacles in the development of knowledge-based farming systems (Benos et al., 2021). Since the dawn of agriculture, people have wished for perfect knowledge of yield prior to harvest. This is because seasonal crop yield forecasts are crucial for decision-making for a variety of stakeholders, including farmers, policymakers, governments, and commodities traders (Basso and Liu, 2019). Algorithms for machine learning often aim to maximise task performance by utilising examples or prior knowledge. ML is particularly useful for reconstructing a knowledge architecture and generating effective associations with respect to data inputs. Precision agriculture's (PA) main objective is to increase agricultural yield production and quality while lowering operational expenses and pollution to the environment. Numerous factors that affect output, including terrain, weather, soil characteristics, irrigation, and fertiliser management, affect potential growth and yield. The rising use of remote and proximal sensing technologies is a result of the requirement for timely and precise sensing of these inputs for large agricultural fields (Chlingaryan, 2018). Predicting crop yields is crucial for the production of food. To improve national food security, policymakers must make timely decisions about imports and exports based on reliable forecasts. Previously, crop and field expertise of the farmer was taken into account when predicting yield (Pant et al., 2021).

2. Applications of Machine Learning in Agriculture

One of the forefront of usage of machine learning in agricultural sector is the forecasting and prediction of crop end result or output. Decisions about which crop species to plant and what tasks to complete during the growing season are made more easily thanks to technology (Nigam et al., 2019). Future yields are the result of a thick layer of pre-harvest operations called crop management. That being said, this is one of the hardest phases of the agricultural life cycle. Crop resistance may be impacted by rising temperatures, erratic wetting and drying cycles, and an increase in the frequency of drought. As a result, many use machine learning developments to accelerate this stage. Reddy and Kumar (2021) opined that Spraying is essential to crop health because it keeps pests and pathogens out. This is also addressed by machine learning initiatives in the agricultural sector. In the field of agriculture, precision or targeted spraying represents the pinnacle of intelligent software and computer vision technologies. Consequently, the system gathers target data, like the plant's dimensions and morphology, and then uses herbicides as necessary. Soil sensor calibration also makes use of machine learning. Together, these factors then aid in the prediction of water stress and nutrient deficits. Soil moisture affects several important agricultural tasks, such as choosing the right crops and when to till and harvest them. Typically, factors from soil and crops as well as weather data are used to forecast moisture. Next, machine learning, regression, and empirical methods are used to reinforce the forecast. This programme makes it feasible to plan water resources more data-drivenly, with better returns and at reduced prices. (Muruganantham, 2022). Pant et al. (2021) submitted that Yield mapping is essentially a component of precision farming. It aids in highlighting the variations in the soil found in different farm regions. In addition to providing information on moisture content, mapping helps farmers handle a number of associated problems on their property. Using data from previous years and machine learning is a popular method for yield mapping. Farmers are able to determine which areas will be best and worst for crops in this way. To get more information, this application may additionally use a range of sensors, such as ones that measure header position and grain flow. Another supplementary tool for optimising fertiliser use is yield mapping. According to Morales and Villalobos, 2023) Among the important applications and methods of machine learning in agriculture are those related to animal welfare and cattle productivity. There are numerous applications for the technology. These include evaluating the welfare of animals, forecasting animal output through predictive modelling, and calculating the environmental effect of livestock operations. Thus, by keeping an eye on vital signs, daily activity levels, and food consumption, farmers can gain a better understanding of cattle well-being. Chewing signals can also be used in animal behaviour classification to determine whether dietary adjustments are necessary. Categorising the movement patterns of animals allows for the tracking of their stress levels. This use of machine learning in agriculture opens up completely new possibilities for improving livestock management and agricultural profitability. Modern techniques for obtaining a far more accurate crop price prediction have been made possible by the development of machine learning. These projections have the potential to be very useful tools for improving financial choices. Commodities used in agriculture have unstable prices. Many factors, including as demand, governmental policy, and the climate, might affect them. Organisations in applied economics and agriculture can benefit from machine learning by using it to better identify pricing variations and provide risk management strategies. The government may lend money to the industry, for instance. Additionally, farmers can prepare by varying the amount of crops they produce. Selecting the right crops to plant can also be aided by price prediction. From a technical perspective, machine learning in predictive analytics uses agricultural datasets (Salpekar, 2019). Gupta (2023) submitted that Robots are able to identify particular fruits and vegetables more accurately because to machine learning technology, which enable them recognise shapes, sizes, and colours. By reducing the amount of time needed for harvesting, automated robots help farmers make more money.

3. Research Method

The research works published within the last six years were to be downloaded because the search year range was set to 2018–2023. Approximately forty-three research publications on crop yield prediction in this range were obtained and examined; only twenty-three were deemed primarily relevant and are listed below. Several algorithms were used by the researcher ranging from tree-based, lazy classifiers, function and rule-based classifiers.

Table 1: Classifiers Acronyms and Meaning

S/n	Acronyms	Meaning
1	DT	Decision Tree
2	RF	Random Forest
3	SVM	Support Vector Machine
4	LR	Linear Regression
5	ANN	Artificial Neural Network

6	DL	Deep Learning
7	MLR	Multiple Linear Regression
8	PR	Polynomial Regression
9	LSTM	Long-Short Term Memory
10	RNN	Recurrent Neural Network

Table 2: Summary of the review and comparison of the crop yield prediction

S/N	Authors(Year)	Algorithms	Dataset	Machine learning tool	Result
1	Agarwal & Tarar (2021)	SVM, LSTM, RNN	Public dataset (kaggle)	Rapid Miner	LSTM outperform other algorithms with accuracy of 95%
2	Aravind (2021)	LR, SVM	Public dataset	Weka	LR and SVM has the same accuracy
3	Kumar et al.(2021)	LR, RF, Gradient Boost	Real time dataset	Knime software	Gradient boos gives more accurate result
4	Bondre et al. (2019)	LR, RF	websites.data.gov.in	Tensorflow	RF outperform LR
5	Iniyan et al.(2023)	RF, KNN,	UCI	Weka	RF outperform KNN
6	Kuradusenge et al., (2023)	RF, PR, SVM	historical weather data and yield statistics from 2005 to 2021. (Irish Maize & Potato)	Keras	Random forest: R2 of 86.7%, PR: 77.3% and SVR: 56%
7	Olofintuyi et al., (2023)	An ensemble deep learning approach was proposed (CNN-RNN with LSTM)	Historical Yield data(1988-2017) and Weather data (Cocoa)	Tensorflow	CNN-RNN with LSTM had the least MAE and RMSE of 26.11 and 28.41 respectively when compared with individual CNN, RNN and LSTM.
8	Cedric et al., (2022)	Decision Tree(DT), Multivariate Regression & K-Nearest Neighbor(KNN)	Annual Rainfall data, Climate data, Yield data & chemical data (Maize, Rice, Cassava, Seed cotton, Yam & Beans)	Phyton	DT: 95.3%, KNN: R^2 = 93.15% and Multivariate Regressor : 89.78%
9	Srikantaiah and Deeksha (2021)	MLR, Ridge Regression & Random Forest Regression	Historical Leaf Yield data and Soil data (pH, EC, OC, N & P) **(Mulberry leaf)**	Tensorflor	Random Forest:94.6%, Ridge : 79.4% and MLR :74.32%
10	Banua and Geethab (2021)	Random Forest and Deep Neural Network.	Yield Data (Season, Year, Production, Humidity, Temperature, Rainfall & Fertilizer) from 2009-2018. **(Rice)**	Weka	RF-DNN: accuracy of 92.1% Random forest : 86.9% Deep neural network 75.25% accuracy.
11	Pant et al., (2021)	Decision Tree Regressor, Random Forest (RF), SVM, and Gradient Boosting Regressor (GB)	Yield data, Average Rainfall per year, Pesticide Tonnes and Average Temperature. (Potatoes, Rice, Wheat & Maize) Country: India	Phyton	Decision Tree Regressor :92% $\mathbf{R^2}$ score RF: 68% GB:89% SVR: 20%

12	Simon et al., (2020)	Xgboost was used for prediction while SVM was used for fertilizer recommendation. Performance Comparison : Random Forest, KNN and RF	Climate data, Crop properties, Area cultivated, and soil features (maize, sorghum, millet, yam, beans, pumpkin, spinach, rice, tomato, groundnut, and cassava)	Weka	Xgboost had the highest R2 score 95.28% for yield prediction and SVM had 97.86% as compared with other models.
13	Abbas et al., (2020)	Linear Regression, Entropy, K-Nearest Neighbor, and Support Vector Regression (SVR).	Yield samples for each growing season and data about the field dimensions. (Potato Tubers)	Tensorflow	The SVR model outperformed all other models for NB-2017, NB-2018, PE-2017, and PE-2018 dataset with RMSE of 5.97, 4.62, 6.60, and 6.17 t/ha, respectively.
14	Nayak (2020)	LSTM, SVM	State, District, Crop Year, Season, Crop, Area and Production (2000-2015). (Coconut)	Rapidminer	GBR had a considerable level of accuracy of 81%.
15	Shah et al., (2018)	SVM (regression), Random Forest, and Multivariate Polynomial Regression were the algorithms used.	Historical yield data and weather data. (Corn)	Tensorflow	The three regression-based models were evaluated and SVR had the best possible results for predicting the crop yield.
16	Gupta et al. (2023)	RF, ANN	Kaggle.com	Weka	RF has the accuracy of 92% while ANN has the accuracy of 87%.

Table 3: Summary of the paper reviewed by year published

S/n	Year	No of Paper
1	2018	1
2	2019	1
3	2020	3
4	2021	6
5	2022	1
6	2023	4

4.0 **Results and Discussion**

This section shows various output of our finding after extensive review of literatures. The review in section 3.0 was categorized into year basis which spread across 2018 to 2023. Table 3 shows the statistics of the paper reviewed by year. Figure 1 shows the graphical representation of the reviewed. Here it is observed six journals of year 2021 were reviewed which is the highest followed by year 2023 where four papers were reviewed.

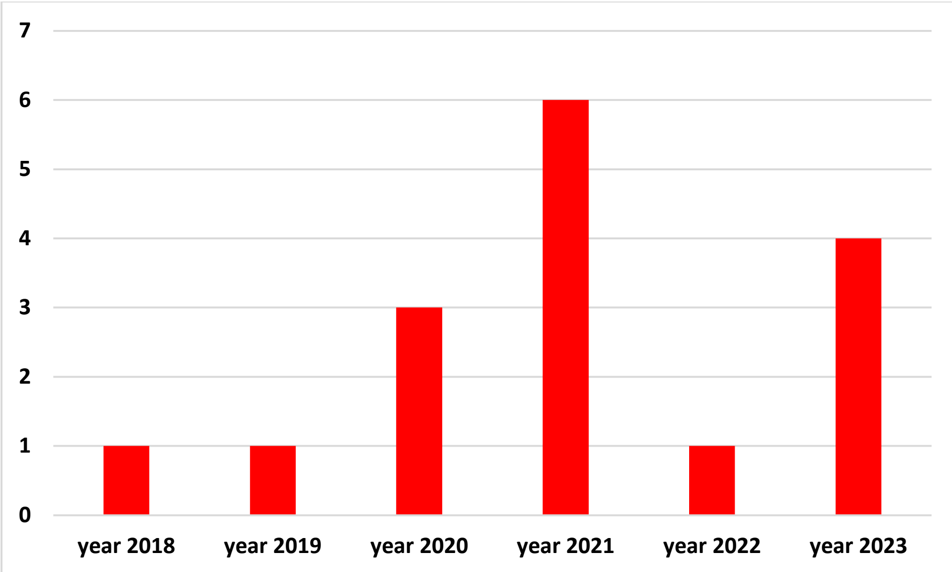

Figure 1: No of paper reviewed by Year

Altogether, ten common algorithms/classifiers are being used for crop prediction and it was observed in the reviewed that total occurrence for all the algorithm is thirty two (32). The result shows that RF is most commonly used and proved to outperform all other algorithms. Figure 2 show the list of the algorithm and their occurrence.

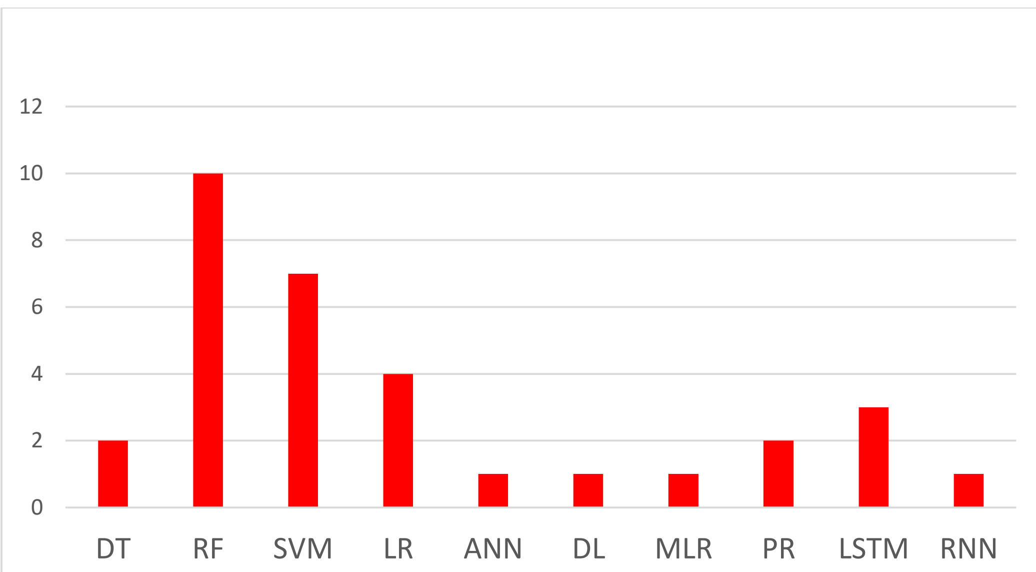

Figure 2: List of algorithms and the Occurrence

Although, when combined with ensemble, SVM proved to had the highest accuracy. Ensemble is the combination of two or more algorithms but Random forest proved to be common reliable and more accurate and outperform other algorithms without ensemble learning.

Conclusion and Recommendation

The bulk of people in our nation get their living from agriculture, which is the foundation of our nation. Farmers will need to adopt smart farming practises to address contemporary concerns like weather variability, supply exceeding demand, and water limitations. Because of these factors, poor irrigation systems, diminished soil fertility, ageing farming practises, and changing climatic conditions, it is anticipated that crop yields will decline. In agriculture, there are a number of methods for predicting crop yield, including machine learning. It therefore becomes pertinent for farmers to predict the particular type of crop suitable for a land or region and how well such a crop will optimally yield. Hence the implementation of machine learning method for crop yield prediction. Our finding revealed that random forest is the most suitable and reliable algorithm for crop yield prediction.

References

Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). Machine learning in Agriculture: A Review. *Agronomy*, *10*(7), 1046.

Agarwal, S., & Tarar, S. (2021). A hybrid approach for crop yield prediction using machine learning and deep learning algorithms. In *Journal of Physics: Conference Series* (Vol. 1714, No. 1, p. 012012). IOP Publishing.

Aravind, T. (2021, December). Review of Machine learning models for Crop Yield Prediction. In *Proceedings of the First International Conference on Combinatorial and Optimization, ICCAP 2021, December 7-8 2021, Chennai, India*.

Banua, E., & Geethab, A. (2021) Rice crop yield prediction using random forest and deep neural network-An integrated approach.

Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. *advances in agronomy*, *154*, 201-255.

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758.

Bondre, D. A., & Mahagaonkar, S. (2019). Prediction of crop yield and fertilizer recommendation using machine learning algorithms. *International Journal of Engineering Applied Sciences and Technology*, *4*(5), 371-376.

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and electronics in agriculture*, *151*, 61-69.

Elijah, O., Rahim, S. K., Abioye, E. A., Salihu, Y. O., & Oremeyi, A. A. (2022, April). Decision Support Platform for Production of Chili using IoT, Cloud Computing, and Machine Learning Approach. In *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)* (pp. 1-5). IEEE.

Gupta, T., Maggu, S., & Kapoor, B. (2023). Crop Prediction using Machine Learning.

Kumar, T., Jyoti, K., & Singla, S. K. (2021). Design and Development of Machine Learning Model for Crop Yield Prediction.

Morales, A., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past or the future. *Frontiers in Plant Science*, *14*, 1128388.

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, *14*(9), 1990.

Nayak, S. K. (2020). A Comparison between major artificial intelligence models for crop yield prediction: case study of Midwestern United State. *International Journal of Modern Agriculture*, *9*(4), 359-369.

Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019, November). Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE.

Oluwole, O. E., Osaghae, E. O., & Basaky, F. D. (2022). Machine learning Solution for Prediction of Soil Nutrients for Crop Yield: A Survey. *Machine learning*, *9*(9).

Pant, J., Pant, R. P., Singh, M. K., Singh, D. P., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*, *46*, 10922-10926.

Ramu, K., & Priyadarsini, K. (2021, October). A review on crop yield prediction using machine learning methods. In *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1239-1245). IEEE.

Reddy, D. J., & Kumar, M. R. (2021, May). Crop yield prediction using machine learning algorithm. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1466-1470). IEEE.

Salpekar, H. (2019, October). Design and implementation of mobile application for crop yield prediction using machine learning. In *2019 global conference for advancement in technology (GCAT)* (pp. 1-6). IEEE.

Shah, A., Dubey, A., Hemnani, V., Gala, D., & Kalbande, D. R. (2018). Estimating and understanding crop yield with deep learning in the Indian Wheat belt

Simon, Y. T., Ojochogwu E. C., Adeiza, R. M., Onyekwere, O., Abiodun, A. A., and Abdulkadri, I. (2020). IoT- based crop yield prediction system. *International Journal of Data Science and Analysis*, *6*(3), 72. https://doi.org/10.11648/j.ijdsa.20200603.11