Application of Machine Learning Techniques in Crop Yield Prediction for Improved Farming Recommendations: A Review

Yahaya Musa Kayode
E-Extension Department
NAERLS, ABU
Zaria, Nigeria
ymkayode@gmail.com

Mohammed Aminu Umar
E-Extension Department
NAERLS, ABU
Zaria, Nigeria
mohammedaminuumar@gmail.com

Muhammad Lawal Tijani
E-Extension Department
NAERLS, ABU
Zaria, Nigeria
mltbnahmad@gmail.com

Aliyu Fatima Hassan
E-Extension Department
NAERLS, ABU
Zaria, Nigeria
aliyufatima1991@gmail.com

Abstract **- In modern agriculture, the utilization of cutting-edge technologies has become essential for optimizing crop production and ensuring food security. One such technology that has gained significant attention is Machine Learning (ML), which has proven to be a powerful tool for enhancing crop yield prediction and guiding precision farming practices. This paper presents a review of recent advancements in the application of machine learning techniques to crop yield prediction and its subsequent impact on providing informed farming recommendations. The paper emphasized the critical role that machine learning algorithms play in accurately forecasting crop yields. It highlights the dominance of algorithms like Random Forest and Support Vector Machines in the literature, while also acknowledging the rising influence of Deep Learning techniques such as Deep Neural Networks and Convolutional Neural Networks. This trend suggests a paradigm shift towards more sophisticated and automated predictive models that leverage the power of complex neural networks for improved accuracy. Furthermore, the study also brings to the fore key features that contribute to effective crop yield prediction, variables such as temperature, humidity, rainfall, and soil type as pivotal factors for accurate forecasts. The paper also captured popular metrics used to evaluate ML model performance.**

Keywords - Machine learning, crop yield prediction, precision farming, deep learning, and algorithms.

I. INTRODUCTION

Agriculture has been the backbone of human civilization for millennia, providing sustenance and livelihood to billions worldwide. As the global population continues to grow, the pressure on the agricultural sector intensifies, demanding increased crop productivity and sustainable farming practices. In this pursuit, even before the advent of the computing age, forward-thinking farmers engaged in what is now recognized as precision agriculture. By meticulously observing their fields throughout growing seasons and during harvest operations, these farmers were able to identify sources of field variability and determine the best course of action to achieve improved crop yields. Relying on their accumulated knowledge and experience, they deduced the optimal strategies for the upcoming year [1]. In recent years, the application of machine learning techniques in agriculture has revolutionized farming practices by providing data-driven insights and predictions. One of the most promising applications is in crop yield prediction, where machine learning models analyze historical data and environmental factors to forecast crop performance accurately. Crop yield prediction poses a significant challenge within the realm of precision agriculture, leading to the development and validation of numerous models to address this issue. Addressing this problem necessitates the utilization of multiple datasets, as crop yield is influenced by a diverse array of factors encompassing climate, weather, soil composition, fertilizer usage, and the choice of seed variety [2]. This implies that crop yield prediction is far from a straightforward task; rather, it involves several intricate steps that must be carefully executed [3]. These predictions play a crucial role in shaping farming recommendations, helping farmers optimize their agricultural practices, reduce resource wastage, and maximize yields.

The continuous rise in data-producing devices and sensors has become a prevalent trend in agriculture, empowering farmers to embrace data-driven decision-making. This practice is commonly referred to as smart farming. Consequently, as technology advances and datasets grow, we can expect even more sophisticated machine-learning solutions that will enable farmers to meet the challenges of feeding a growing world population while fostering environmental stewardship. This research explores the application of machine learning techniques in crop yield prediction, shedding light on its potential to drive improved farming practices.

II. RELATED WORK

Crop yield prediction plays a critical role in facilitating rapid decision-making for policymakers at international, national, and regional levels. Its significance lies in empowering farmers to make informed choices regarding crop selection and optimal timing for cultivation. Various methodologies exist for crop yield prediction, and this review article thoroughly explores the application of machine learning techniques in this domain. By delving into the existing literature, the review provides valuable insights into the advancements and progress made in utilizing machine learning for accurate crop yield prediction.

[4] In this paper, a hybrid machine learning (ML) algorithm is proposed to predict and improve the accuracy of crop yield forecasting using hybridized ML algorithms, that combines stacked generalization, gradient boosting, random forest, and least absolute shrinkage and selection operator (LASSO) regression. Results from the proposed hybridized ML algorithm outperform the others with 88.89% accuracy. Thus, they conclude the proposed approach is effective for predicting crop yield.

[5] The authors proposed to predict crop yield quantity based on factors like temperature, humidity, soil moisture level, and field area. which is displayed as an output to the user that aids the farmer to harvest the crop. Based on experimentation of different ML algorithms the authors report that the Random Forest algorithm is effective for crop yield prediction.

[6] This paper discusses the most recent research on machine learning-based strategies for efficient crop yield prediction and proposes a machine learning technique that will provide high precision in crop yield Prediction. They opinionated that Machine learning models are effective for crop yield prediction and that Artificial Neural Networks (ANN) can produce more precise results.

[7] The paper presents an investigation of crop yield prediction using machine learning algorithms such as Support Vector Machine, Random Forest, and ID3. The results of the investigation show that the machine learning-based framework can predict crop yield with high accuracy. The paper reports that the Random Forest algorithm outperformed the other two algorithms in terms of accuracy. However, the paper also notes that the performance of the machine learning algorithms depends on the quality and quantity of the data used for training and testing. Therefore, the accuracy of the framework may vary depending on the data used.

[8] In this paper, the authors designed a system with a machine learning algorithm that recommends the best crop based on the nitrogen, phosphorus, and potassium content of the soil, which can also provide the answers to the farmer's query and subsequently helps to increase profit margin. Decision tree gives higher accuracy in results.

[9] In this paper, a graph-based recurrent neural network (RNN) was proposed for crop yield prediction, which incorporates both geographical and temporal knowledge in the model and further boosts predictive power. The proposed model outperforms existing state-of-the-art methods on various metrics.

[10] In this paper, a hybrid approach for crop yield prediction is proposed, utilizing both machine learning (ML) and deep learning algorithms. The study focuses on the agricultural sector, particularly in India, where farming significantly contributes to the economy. The model incorporates parameters such as temperature, rainfall, pH value, relative humidity, and area to predict the best crop for cultivation. The hybridized approach employs a support vector machine (SVM) as a machine learning algorithm, while long-short term memory (LSTM) and recurrent neural network (RNN) are employed as deep learning algorithms. The proposed model demonstrates an improved accuracy of 97% compared to existing methods, providing more accurate crop yield predictions.

The authors in [11] present a novel algorithm for early yield prediction in different grapevine varieties using computer vision and machine learning. The authors utilized a mobile platform with an RGB camera to acquire images of grapevines at the pea-size berry stage. A SegNet architecture was employed for the semantic segmentation of clusters, berries, and canopy elements. The canopy features obtained from the segmented images were used to train support vector regression (SVR) models for predicting the number of actual berries and yield. The method achieved promising results in detecting visible berries and canopy features, with good accuracy and correlation between actual and estimated values. The algorithm allows predicting yield up to 60 days before harvest across various grapevine varieties.

[12] In the study, the authors address the issue of declining agricultural productivity in Japan. They propose a data-driven approach to predict paddy rice yields by integrating weather information, cultivation data, and location details. The research employs the

XGBoost machine learning algorithm to establish a predictive model. Key stages of the rice cultivation process, including nursery, tillering, panicle development, and ripening, are highlighted. Weather variables such as daily maximum and minimum temperatures and sunshine time are identified as crucial predictors. The study emphasizes the importance of these factors in accurately forecasting rice yields. The integration of XGBoost demonstrates promising results in predicting paddy rice yields, contributing to the field of Agri-informatics, and providing insights into improving agricultural planning and management.

[13] In this paper, the authors propose a data-driven approach for crop yield prediction, with a focus on the agricultural sector in India. The study aims to tackle the challenging task of estimating agricultural production by taking into account various influential factors including temperature, precipitation, humidity, and soil type. The authors underscore the critical importance of precise crop yield predictions in enabling effective agricultural planning and informed decision-making. Through the utilization of the Random Forest algorithm, the authors provide a robust framework for accurate crop yield forecasting. The proposed methodology showcases impressive accuracy, reaching 98%, which signifies the potential of data mining techniques to revolutionize crop yield prediction and enhance agricultural productivity.

[14] This study presents a novel approach to predicting crop yield using Convolutional Neural Networks (CNNs) based on data acquired from Unmanned Aerial Vehicles (UAVs). The goal is to predict crop yield at an intra-field scale using RGB and NDVI data. The study explores various aspects of CNN training, such as training algorithms, network depth, regularization strategies, and hyperparameters. The CNN model achieved its best performance using early growth RGB data, with a mean absolute error (MAE) of 484.3 kg/ha and a mean absolute percentage error (MAPE) of 8.8%. The study demonstrates that the CNN architecture performed better with RGB data compared to NDVI data. The research emphasizes the potential of CNNs for accurate crop yield prediction and their applicability in precision agriculture.

III. METHODOLOGY

A. Research Methodology (Review protocol)

The primary objective of this research is to investigate various machine learning methods applied in crop yield prediction; the features used in crop yield prediction and assess the effects of these models. To achieve this, an analysis of secondary quantitative data was conducted.

Before commencing the research review, a review protocol is established. The review process adheres to the well-recognized guidelines outlined by [15]. The initial step involves defining the research questions. Once the research questions are formulated, various pertinent scholarly articles sourced from reliable and credible online repositories such as Google Scholar, Springer, and Science Direct among others were consulted. The numerical data presented in these papers were extracted and subsequently analyzed to address the research questions.

B Research Questions

This research aims to delve into the existing literature in the realm of machine learning and its application to crop yield prediction. Various studies have been scrutinized from diverse angles to enhance comprehension. The study is guided by the following set of three research questions (RQs).

- **RQ1:** What machine learning algorithms have been employed in the literature for predicting crop yield?
- **RQ2:** What features have been utilized in the literature for predicting crop yield using machine learning?
- **RQ3:** What evaluation metrics have been applied in the literature for assessing models' performance for crop yield prediction?

IV. RESULT AND DISCUSSION

RQ1 - Algorithms used: As depicted in Fig. 1, Random Forest is the most frequently employed algorithm in the literature followed by Support Vector Machine. However, it's crucial to understand that while Random Forest and Support Vector Machine are being frequently used do not necessarily correlate with superior performance. Interestingly, there has been a surge of interest in applying Deep Learning (DL), specifically, Deep Neural Networks (DNN) and Convolutional Neural Networks (CNNs) which are subsets of Machine Learning, to crop yield prediction. Recent studies have demonstrated the potential of DL methods in this domain. Notably, DL methods offer advantages such as automated feature extraction and impressive performance. The emergence of DL in crop yield prediction holds promise for future advancements in this field. It's worth noting that among the selected publications, both classifiers, regression, and clustering algorithms have been utilized. However, some publications employ clustering algorithms in connection with machine vision using image data rather than traditional numerical datasets as shown in Table 1. While crop yield is a continuous value, it is observed that

classification algorithms and metrics were mostly used in literature.

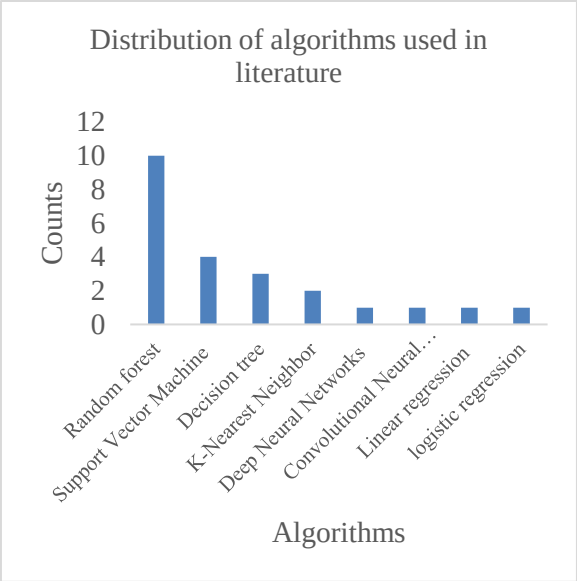

Fig. 1. Distribution of algorithm used in the literature

RQ2 - Features used: The commonly utilized features as revealed by the literature encompass temperature, humidity, rainfall, field data, and soil type, as indicated in Fig. 2. In addition to these prevalent attributes, certain studies incorporate specific features unique to their contexts. These distinctive elements include pH value, NDVI, RGB, season, and perhaps crop-related information as shown in Table 1.

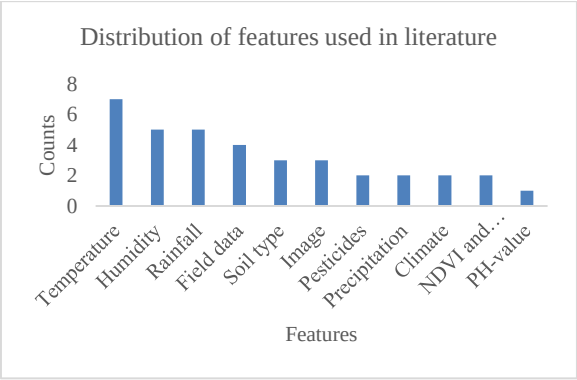

Fig. 2. Distribution of features used in the literature

RQ3 - Metrics used: The selected papers exhibit a limited range of evaluation parameters. The predominant metrics employed across these studies to assess model performance are Accuracy and Root Mean Squared Error (RMSE). Some alternative evaluation metrics include Mean Absolute Error (MAE), R-Squared (R2), and Mean Absolute Percentage Error (MAPE). However, specific studies introduced additional metrics, and many of them appear to combine the aforementioned metrics as depicted in Table 1. Remarkably, the majority of models achieved commendable accuracy levels according to these evaluation metrics, signifying accurate predictive capabilities.

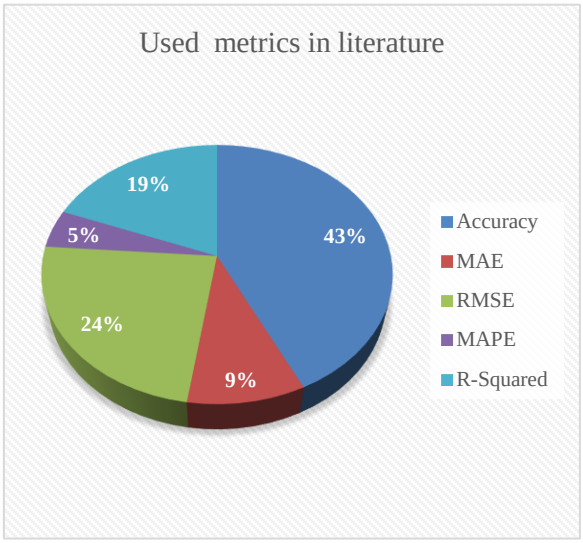

Fig. 3. Distribution of metrics used in the literature

TABLE 1. RELATED WORKS SYNTHESIS

Reference	Feature Used	ML Algorithm Used	Metric Used	Result
[16]	Soil type, climate statistics (e.g., rainfall), yields from previous years	Random Forest Classifier, K-Nearest Neighbor, Support Vector Machine	Accuracy	75%
[17]	State, crop year, season, crop, area, and production	Random forest	Accuracy	90%
[18]	Genotype, Environment Data (Weather, Soil type)	Deep Neural Networks (DNN)	RMSE	8.9%
[19]	Image	CNN-LSTM	RMSE R-Squared and	RMSE: 353.74 R2: 0.74
[20]	Remote sensing satellite imagery	Ensemble model based on Long Short-Term Memory (LSTM) and Convolutional LSTM	R-Squared and RMSE	Soybean: 0.73 366.14. Maize: 0.49 413.22
[21]	Hg/ha_yield, Average_rain_fall_mm_per_year, Pesticides_tones, Avg_temp	Gradient Boosting Regressor, Random Forest Regressor, SVM, Decision Tree Regressor	Accuracy	96%
[22]	Color images	Convolutional Neural Networks (Faster R-CNN, SSD, SSDLITE)	MPE	8.84%
[23]	Multispectral and LIDAR data, meteorological data	CNN-attention-LSTM model	R-Squared and RMSE	0.78 and 8.27%
[24]	Agricultural data, annual rainfall data, climate data, weather data, and chemical	Decision tree, Multivariate logistic regression, K-nearest neighbor models	R-Squared and MAE	95.3%, 89.78%, and 93.15%,
[25]	Weather parameters (rainfall, temperature, slope, humidity, soil moisture)	Pattern matching technique	Accuracy	98%
[26]	High-resolution satellite data, soil data, field data	Linear regression (LR), Decision tree (DT), and Random Forest (RF) regression models	R-Squared and RMSE	LR: R2: 0.85-0.95, RMSE: 2.06-2.66; DT: R2: 0.95-0.97, RMSE: 1.54-2.06; RF: R2: 0.97-0.98, RMSE: 1.40-1.85

V. CONCLUSION

This paper brings to light the transformative potential of machine learning in modern agriculture. Through an in-depth analysis of recent advancements and research in the field of crop yield prediction, the paper underscores the profound impact that data-driven insights and predictions can have on optimizing agricultural practices and ensuring food security. The paper emphasizes the critical role that machine learning algorithms play in accurately forecasting crop yields. It highlights the dominance of algorithms like Random Forest and Support Vector Machines in the literature, while also acknowledging the rising influence of Deep Learning techniques such as Deep Neural Networks and Convolutional Neural Networks. This trend suggests a paradigm shift towards more sophisticated and automated predictive models that leverage the power of complex neural networks for improved accuracy. Furthermore, the study dives into the features that contribute to effective crop yield prediction. It identifies key variables such as temperature, humidity, rainfall, and soil type as pivotal factors for accurate forecasts. Also, the paper made mention of popular metrics used in literature.

REFERENCES

[1] Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.J., "Big Data in Smart Farming – A Review. Agric. Syst." 153, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023. 2017.

[2] Alexandros Oikonomidis, Cagatay Catal & Ayalew Kassahun, " Hybrid Deep Learning-based Models for Crop Yield Prediction", Applied Artificial Intelligence, 36:1, DOI: 10.1080/08839514.2022.2031823. 2022.

[3] Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., & Zhu, M. "Design of an Integrated Climatic Assessment Indicator (ICAI) for Wheat Production : A Case Study in Jiangsu Province, China. Ecological Indicator, 943–953. https://doi.org/10.1016/j.ecolind.2019.01.059. 2019.

[4] Anbananthen KSM, Subbiah S, Chelliah D., "An Intelligent Decision Support System for Crop Yield Prediction Using Hybrid Machine Learning Algorithms". https://doi.org/10.12688/f1000research.73009.1 .2021.

[5] C., K. and Srinivas. "Crop Yield Prediction Using Machine Learning. International Journal for Research in Applied Science and Engineering Technology", 9:553-555. doi: 10.22214/IJRASET.2021.37413. 2021.

[6] S., Sairamkumar. "Design of ANN Based Machine Learning Method for Crop Prediction. 3(3):223-239. doi: 10.36548/JIIP.2021.3.005. 2021.

[7] GS. Sajja., SS. Jha., H. Mhamdi., M. Naved., S. Ray., K. Phasinam. "An Investigation on Crop Yield Prediction Using Machine Learning". 5-11. doi: 10.1109/ICIRCA51532.2021.9544815. 2021.

[8] V. Vanarase, V. Mane, A. Harsha, Bhute, A. Tate., S. Dhar. "Crop Prediction Using Data Mining and Machine Learning Techniques". doi: 10.1109/ICIRCA51532.2021.9544724. 2021.

[9] J, Fan, J. Bai, Z. Li, A. Ortiz-Bobea, C. P, Gomes. "A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction. arXiv: Learning. 6-9. https://doi.org/10.1609/aaai.v36i11.21444. 2021.

[10] S. Agarwal and S. Tarar. J. Phys.: Conf. Ser. 1714 012012 DOI 10.1088/1742-6596/1714/1/012012. 2021.

[11] F. Palacios, M. Diago, P. Melo-Pinto, J. Tardaguila. "Early yield prediction in different grapevine varieties using computer vision and machine learning". Precision Agriculture. https://doi.org/10.1007/s11119-022-09950-y . 2022.

[12] Y. Maeda, T. Goyodani, S. Nishiuchi, E. Kita, "Yield Prediction of Paddy Rice with Machine Learning" Int'l Conf. Par. and Dist. Proc. Tech. and Appl. | PDPTA'18. 360-361. 2022.

[13] P. Kamath, P. Patil, S. Shrilatha, Sushma, S. Sowmya, "Crop yield forecasting using data mining, Global Transitions Proceedings", Volume 2, Issue 2, 402-407, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.08.008. 2021.

[14] P. Nevavuori, N. Narra, T. Lipping. "Crop Yield Prediction with Deep Convolutional Neural Networks", Computers and Electronics in Agriculture. Volume 163, https://doi.org/10.1016/j.compag.2019.104859. 2019.

[15] Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., Visaggio, G., "Guidelines for performing Systematic Literature Reviews in Software Engineering". Retrieved from https://userpages.uni-koblenz.de/~laemmel/ esecourse/slides/slr.pdf. 2007.

[16] K. Rao, K. Yashwanth, M. Goud "Crop Yield Prediction by using Machine Learning Techniques." Annals of the Romanian Society for Cell Biology, 25(6), 2551–2557. Retrieved from http://annalsofrscb.ro/index.php/journal/article/view/5869. 2021.

[17] P. Bhasha, J. Babu, M. Vadlamudi, K. Abraham, S. Sarangi, "Automated Crop Yield Prediction System Using Machine Learning Algorithm". Journal Of Algebraic Statistics: 2512-2522. 2022.

[18] S. Khaki and L. Wang, "Crop Yield Prediction Using Deep Neural Networks". Frontiers in Plant Science, Volume 10, 2019. https://doi.org/10.3389/fpls.2019.00621. 2019.

[19] Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. "County-level soybean yield prediction using deep CNN-LSTM model". Sensors, 19(20), 4363. https://doi.org/10.3390/s19204363. 2019.

[20] M. Divakar, M. Elayidom, R. Rajesh, "Forecasting crop yield with deep learning-based ensemble model". Materials Today: Proceedings, Volume 58, Part 1, Pages 256-259. https://doi.org/10.1016/j.matpr.2022.02.109. 2022.

[21] J. Pant, R. Pant, M. Singh, D. Singh, H. Pant, "Analysis of agricultural crop yield prediction using statistical techniques of machine learning". Materials Today: Proceedings Volume 46. 2021.

[22] D. Tedesco-Oliveira, R. Pereira da Silva, W. Maldonado, C. Zerbato, "Convolutional neural networks in predicting cotton yield from images of commercial fields", Computers and Electronics in Agriculture, Volume 171, 2020, 105307, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2020.105307. 2020.

[23] W. Zhou, C. Song, C. Liu, O. Fu, T. An, Y. Wang, et al, "A Prediction Model of Maize Field Yield Based on the Fusion of Multitemporal and Multimodal UAV Data: A Case Study in Northeast China". Remote Sens. 2023, 15(14), 3483; https://doi.org/10.3390/rs15143483. 2023.

[24] L. Cedric, W. Adoni, R. Aworka, J. Zoueu, J. Mutombo, M. Krichen, C. Kimpolo, "Crops yield prediction based on machine learning models: Case of West African countries". Smart Agricultural Technology, Volume 2, 100049, ISSN 2772-3755, https://doi.org/10.1016/j.atech.2022.100049. 2022.

[25] Anjana, A. Kedlaya, A. Sana, B. Bhat, S. Kumar, N. Bhat,

"An efficient algorithm for predicting crop using historical data and pattern matching technique", Global Transitions Proceedings, Volume 2, Issue 2, Pages 294-298, ISSN 2666-28. 2021.

[26] M. Boori, K. Choudhary, R. Paringer, A. Kupriyanov, "Machine learning for yield prediction in Fergana Valley, Central Asia", Journal of the Saudi Society of Agricultural Sciences, Volume 22, Issue 2, 107-120, ISSN 1658-077X, https://doi.org/10.1016/j.jssas.2022.07.006. 2023.