Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Predictive Modeling of Agricultural Water Footprints in Iraq Using Machine Learning

[1]Ali Hasan Taresh, [2]Huda Mowafek Kadhim
[1]*University of Information and Communications Technology*
Email: alihtaresh@uoitc.edu.iq ORCID: 0000-0002-8776-0322
Mobile: +9647704842545
[2]*University of Information and Communications Technology*
Email: ms202320754@iips.edu.iq Mobile: +9647807535204

ARTICLE INFO

Received: 15 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

The research used machine learning and data mining to improve the use of water in agriculture in Iraq, assessing the water footprints of crops. Datasets from 4TU.ResearchData and Google Earth Engine were merged. The datasets were integrated into classification or regression algorithms to predict either crops locally planted or imported. The algorithms that used SVM and Random Forest achieved high accuracy successfully, which can help with the management of sustainable water and the selection of crops in regions with limited water resources.

Keywords: Machine learning, Data mining, Crop prediction, Water footprint, Agriculture, Iraq

1. INTRODUCTION

Jabaar et al. 2023 show that there is an important tool known as the water footprint that evaluates the consumption and use of green, gray, and blue water in agriculture, which is used in Iraq because it is considered one of the countries that suffer from water shortages.[1]. Alzubaidi et al. 2024 explain that because of dam construction projects from neighboring countries and climate change, Iraq faces water problems, which greatly affect agricultural crops[2]. Wedaa et al. 2023 indicate that from the differences in climate and irrigation, regional variations in water footprints arise from differences in climate and irrigation methods, so they imply the trade of virtual water, which involves importing crops that consume a lot of water; it could improve water efficiency and sustainability [3].

Ansorge 2020 shows that the water footprint can be used in life cycle assessment (LCA) and volumetric approaches, which assess environmental impact and analyses water consumption.[4]. Kong et al. 2021 suggested a method to integrate data with statistical evaluation called Water Footprint-Based Principal Component Analysis (WFPCA). This method attempts to optimize the use of water and deal with scarcity while balancing the needs of economic and environmental [5].

Singh et al. 2020 show in their study that deep learning models such as convolutional neural networks (CNN) were shown to detect water shortage in crops by analyzing texture and leaf color, thus indicating that the use of machine learning and artificial intelligence plays an important role in water management.[6]. Campi et al. 2024 explain that the machine learning techniques used for real-time monitoring use data from the satellite, which includes Random Forest and Support Vector Machine (SVM).[7]. Mehla 2022 shows that it is possible to have resilience to climate change and food security from using technologies based on artificial intelligence, which improve irrigation accuracy, enhance water sustainability, and reduce the waste of water[8].

2. RELATED WORK

In 2023 Padriya and Patel. presented a system for predicting crop yields and water requirements based on the Internet of Things (IoT) and machine learning, using a random forest algorithm that achieved 97% accuracy [9].

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

In 2023 Joshila Grace et. al. Have developed a model using random forests to predict crop yields and prices, which helped farmers choose crops achieving high accuracy [10].

In 2023 Rajdeep Chatterjee et. al. Proposed a system smart farming that combines logistic regression and the Internet of Things and achieves in real-time crop prediction over 98% accuracy [11].

In 2019 Suhas L. et al. presented a prediction of agricultural production based on environmental factors such as rainfall and temperature. Multiple linear regression was used which achieved the MS is 0.02 and CoD: 0.66 and decision tree regression, MSE is 0.18 and CoD: 2.72[12].

In 2024, Ashrakat A. Lotfy et al. developed machine learning models to predict wheat's blue and green water footprints in Egypt's Nile Delta. Hybrid algorithms like XGB-LASSO achieved perfect accuracy (R^2 = 1.0) under optimal scenarios [13].

In 2022, Kuradusenge et al. researchers used Random Forest, Polynomial Regression, and Support Vector Regressor to predict Irish potato and maize yields, with Random Forest achieving the highest accuracy (R^2 = 0.875 for potato, 0.817 for maize) [14].

In 2021, A. S. Sreerama et al. the study applied Decision Tree, Gradient Boosting, and Random Forest Regressor to predict global crop yields, with Decision Tree and Gradient Boosting showing the best R^2 scores [15].

In 2023, Ms.V.Mahalakshmi et al. location-specific crop yield prediction was enhanced using SVM, Random Forest, Neural Networks, and hybrid models, with hybrid models like SVM + RNN-LSTM reaching up to 97% accuracy, despite processing limitations [16].

3. PROBLEM STATEMENT

Iraq is facing water shortages due to climate change and neighboring countries' river water policies. Rising temperatures and reduced precipitation, coupled with damming activities in Turkey and Iran, threaten the country's water security, agriculture, and overall stability [17]. Machine learning is applied to forecast the solution of this data to achieve precision based on machine learning predictions [18]. In Previous studies have examined the water footprint issue using various methods, such as Random Forest, Logistic Regression, Multiple Linear Regression, and Decision Tree Regression, among others. However, these studies have not adequately addressed the specific challenges of the Iraqi environment, prompting the need for further investigation. To address the scarcity of data in this area, A hybrid model was proposed to process the data in this field and was applied, which led to accurate results with an accuracy rate of 98%. This model's performance was compared with five algorithms (Logistic Regression, SVM, Random Forest, Ridge Regression, and Polynomial Regression) which indicate from the results the model was found to be better in terms of accuracy. These results proven the efficiency of the proposal to solve the water footprint problem in Iraq effectively.

4. METHODOLOGY

4.1 Data Collection

The data in the research came from two sources and were combined into one dataset, and this data came from one and two repositories. The first repository provided the first part of the dataset that contains crop production (tons/hectare), harvested area (hectare), water footprint (green and blue), production (tons), percentage of irrigated areas, water consumption of crops (mm/year), etc., covering 66 crops from Iraq during the period from 2007 to 2019. The second repository, through which data were obtained that include temperature (°C), evapotranspiration (ET, mm), humidity (%), rainfall amounts (mm), etc. for the same research area [19][20]. As shown in Figure.1

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

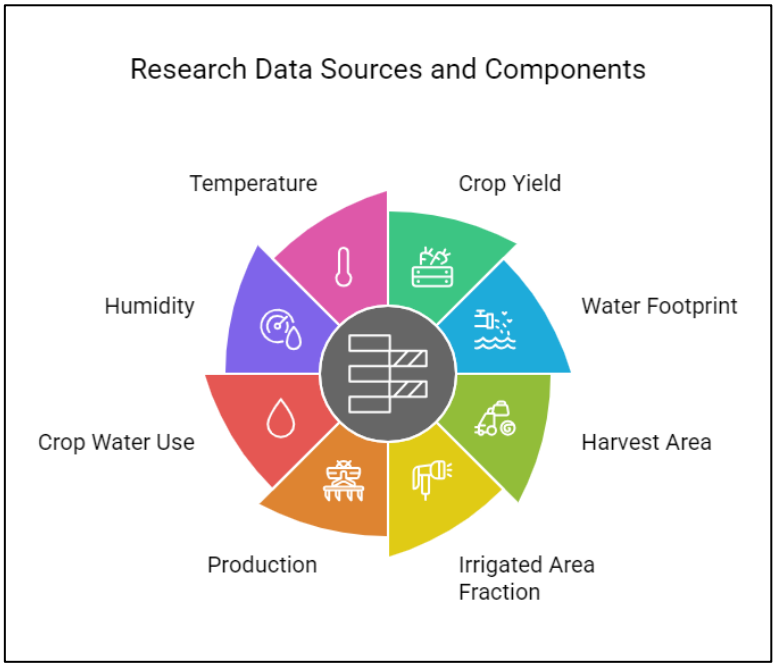

Figure.1 Data Collection

4.2 Data Pre-processing

In this research, several data pre-processing operations were used. This is an important step in analyzing the data and making it suitable for the prediction model. It helps improve the quality of the data. These operations are used before building the model, which include clean data, normalize data and reduce noise data, as showing in figure.2.

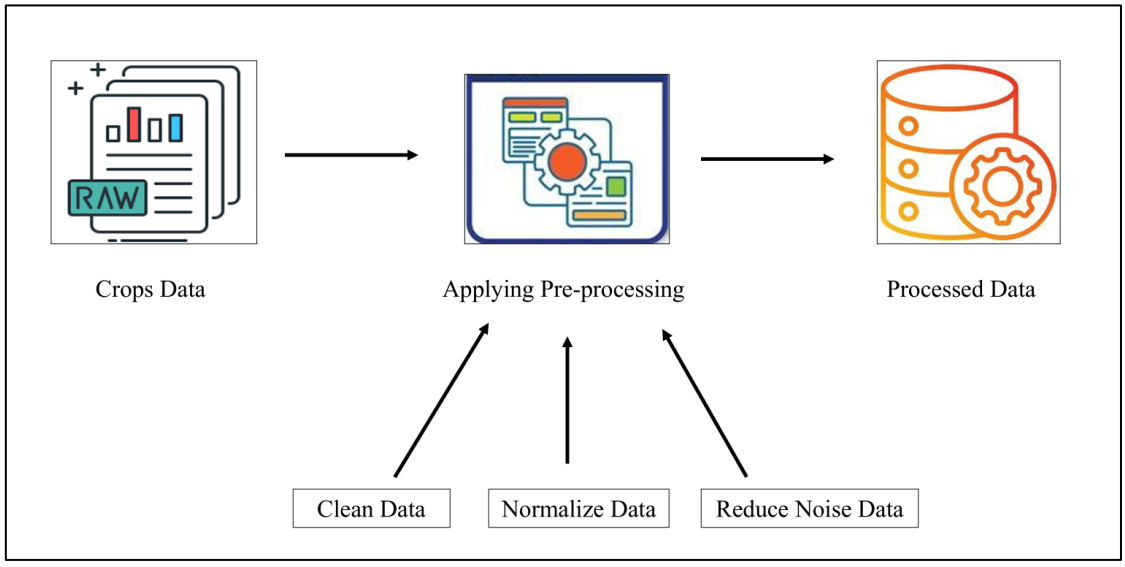

Figure.2 Data Pre-processing

4.3 Model Selection

In this research, regression and classification algorithms as shown in figure.2, were used to solve the water footprint problem and determine the most appropriate methods for water management by determining whether crops are better grown locally or imported from other countries, using a threshold for the total water footprint of crops. If the total water footprint is less than 1534 m^3/ton, it is considered suitable for local agriculture, and if it exceeds that, these crops are preferable to be imported.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

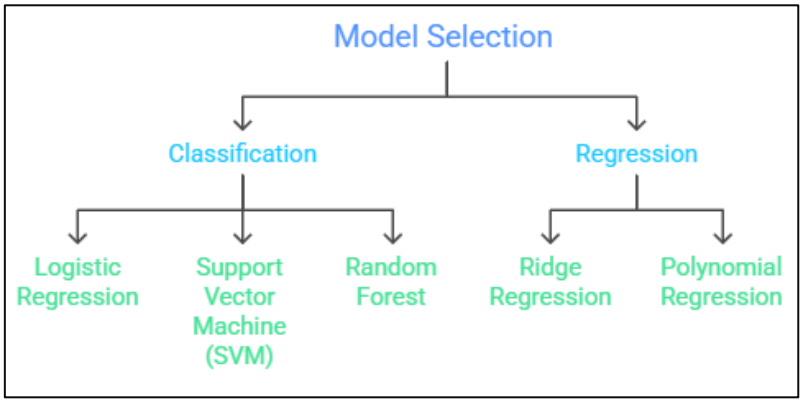

Figure.3 Model Selection

4.4 Workflow

Hybrid approach as shown in figure.4 start with preprocessing operations used on Iraq crops data such as clean and normalize then use future selection to extract important features like water footprint, harvest area, crop yield, climate conditions...etc. The dataset is split into training (80%) and testing (20%) subsets. A Random Forest Classifier is trained on the training data to predict whether crops should be planted locally or imported, based on water footprint analysis. The model is evaluated on the testing dataset, and the system outputs a decision recommendation to optimize water resource usage.

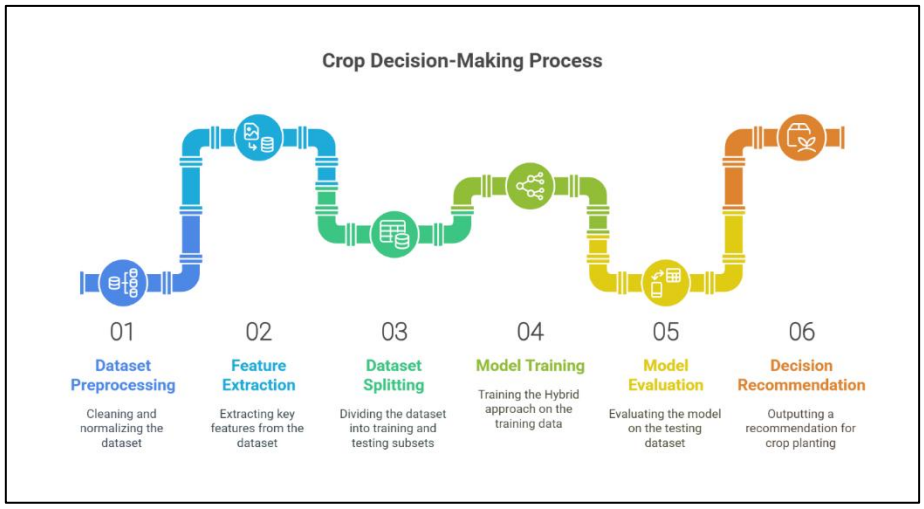

Figure.4 Model Workflow

4.5 Model Design

The hybrid model integrates Logistic Regression, SVM, Random Forest, Ridge Regression, and Polynomial Regression to achieve high accuracy in analyzing water fingerprints in Iraq. Logistic Regression handles binary classification, such as sustainable water usage detection, while SVM classifies complex patterns in water data. Random Forest improves accuracy by combining decision trees and reducing overfitting. Ridge Regression provides stable predictions for water demand, and Polynomial Regression captures non-linear relationships, such as environmental impacts. Through this, in order to improve the use of water resources, reduce costs, and benefit from data mining and artificial intelligence to manage water effectively, these algorithms work together, as shown in figure.5.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

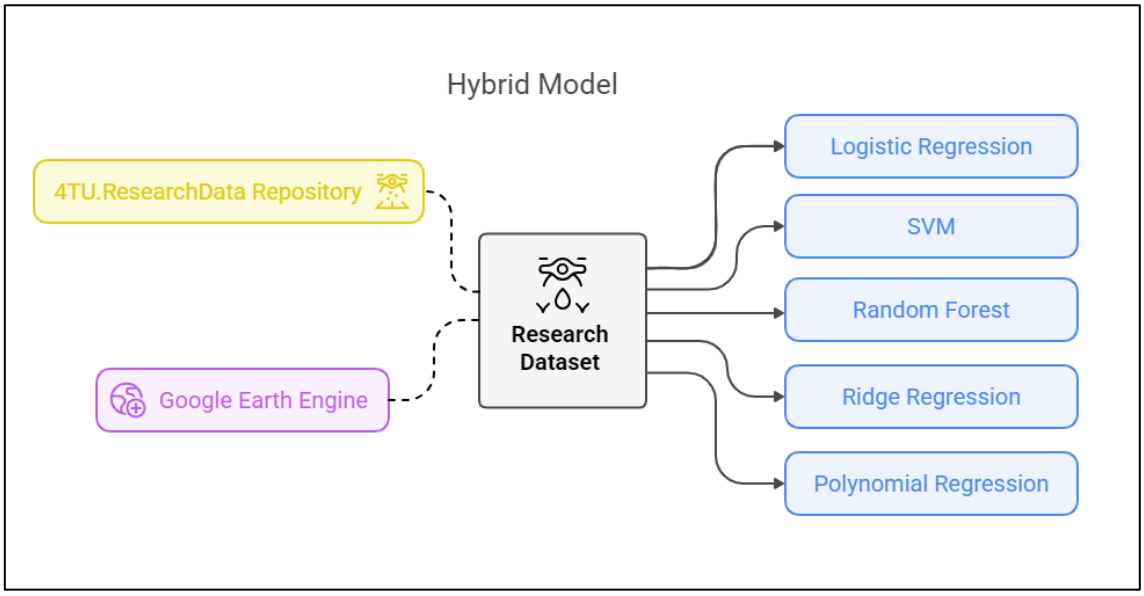

Figure.5 Hybrid Model

4.6 Training and Evaluation

To ensure that the models are evaluated fairly, the data is split into 80% for training and 20% for testing. These algorithms (classification and regression) are trained first and then tested to evaluate the performance of the model.

Training Process:

- Classification algorithms, Logistic Regression, Support Vector Machine (SVM), and Random Forest, were trained to classify crops as either planted locally or imported based on a water footprint threshold of 1534 m^3/ton.
- Regression algorithms, Ridge Regression and Polynomial Regression, were trained to estimate whether the crop is locally grown or should be imported.

Evaluation Metrics:

To evaluate the performance of the predictive model in classification, metrics such as precision, recall, F1 score, and AUC-ROC were used, meanwhile metrics such as R2 score and mean square error (MSE) were used to evaluate the regression prediction models.

5. RESULTS

5.1 Model Results

The Random Forest algorithm was used to classify Iraq's crops into either locally grown or imported from other countries. The results showed that crops suitable for local cultivation such as beans, spinach, tomatoes, watermelons, etc. due to their low water footprint for irrigation. However, crops such as almonds, apples, rice, soybeans, and wheat had been identified as fit for import according to their high irrigation water footprints. The Random Forest algorithm showed robust efficacy in executing these classifications, establishing an accurate foundation for decision-making in water resource management. From 2025 to 2027, the algorithm gave reliable predictions for optimizing water use in agriculture. Figure.5 shows how many crops were grown locally versus imported from 2025 to 2027. It helps see if more crops were being produced locally or brought in from other places during those years, giving a clear picture of farming and import trends.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

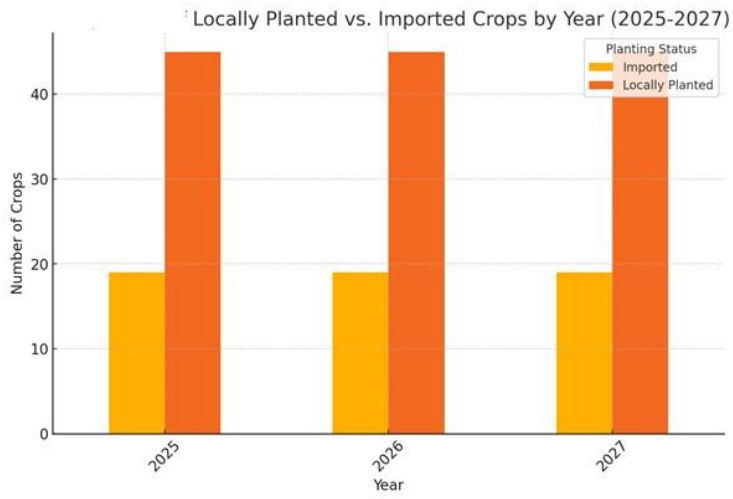

Figure.5 Result chart

5.2 Performance Evaluation

The performance of classification and regression predictive was evaluated using the testing dataset. Table 1 summarizes the performance metrics for all models:

Table 1: Algorithm Performance Metrics

	Classification			Regression	
Algorithm / Evaluation Metrics	Logistic Regression	SVM	Random Forest	Ridge Regression	Polynomial Regression
Accuracy	74%	98%	98%	---	---
Recall	78%	100%	100%	---	---
Precision	50%	93%	95%	---	---
MSE	---	---	---	64.6	36.8
R^2 Score	---	---	---	99%	99%

Discussion

The results and accuracy that were processed by the proposed model can be observed, as we notice that the ratio has outperformed other models in our proposed model with high accuracy, as you can see in the tables. Random Forest and SVM outperformed Logistic Regression in classification tasks, achieving 98% accuracy and 100% recall, excelling in handling complex, non-linear data for tasks like crop decisions. Logistic Regression, while simpler, showed lower accuracy (74%) and precision (50%), struggling with intricate datasets. In regression, polynomial regression (MSE = 36.8) outperformed ridge regression (MSE = 64.6) in capturing non-linear relationships, such as environmental impacts on water footprints. Key variables like rainfall, temperature, and crop yield were critical, though missing data limited the dataset.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6. CONCLUSION

This research used machine learning to determine whether crops in Iraq should be grown locally or imported based on their water footprints. Crops suitable for local agriculture such as beans, watermelon and tomatoes due to their low water footprint, while rice, almonds and wheat are preferred for import due to their high-water consumption, so their water footprint is high. Random forest algorithms and support vector machine showed high accuracy of 98%, outperforming logistic regression which showed an R2 of 99%, but the mean square error appeared high. The research presents an idea to improve water use in Iraqi agriculture to contribute to achieving food security and improving water use efficiency. For the future work establishing an archive is recommended to update and store the latest data to makes sure that model stay accurate and flexible for changing conditions.

REFERENCES

[1] M. H. Jabaar and S. A. Abed, "Estimating the Water Footprint of the Four Important Cereal Crops in the Euphrates River Basin, Iraq," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1215, no. 1, 2023, doi: 10.1088/1755-1315/1215/1/012053.

[2] Eman F.M . Alzubaidi Waleed Ibrahim sultan Karar mahdi, "An Economic and Analytical Study to Estimate the Water Footprint and Virtual Water Trade for Grain Crops in Iraq," vol. 4, pp. 191–201, 2024, doi: : https://doi.org/10.56286/ntujavs.v2i2.

[3] Z. Wisam Wedaa, S. A. Abed, and S. H. Ewaid, "The Agricultural Water Footprint of Al-Qadisiyah Governorate, Southern Iraq," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1029, no. 1, 2022, doi: 10.1088/1755-1315/1029/1/012025.

[4] L. Ansorge, "Water Footprint: Two Different Methodologies," *Tecnura*, vol. 24, no. 66, pp. 119–121, 2020, doi: 10.14483/22487638.15903.

[5] X. Kong, D. Wang, J. Zhao, C. Wang, and R. Cheng, "Evaluation of Carrying Capacity on Water Resources Based on Improved Water Footprint Method Evaluation of Carrying Capacity on Water Resources Based on Improved Water Footprint Method", doi: 10.1088/1755-1315/793/1/012033.

[6] N. Singh, C. Subir, K. Chakraborty, Y. Anand, and R. Kumkum, "Identifying crop water stress using deep learning models," *Neural Comput. Appl.*, vol. 4, 2020, doi: 10.1007/s00521-020-05325-4.

[7] P. Campi, A. F. Modugno, G. De Carolis, F. Pedrero Salcedo, B. Lorente, and S. Pietro Garofalo, "A Machine Learning Approach to Monitor the Physiological and Water Status of an Irrigated Peach Orchard under Semi-Arid Conditions by Using Multispectral Satellite Data," *Water (Switzerland)*, vol. 16, no. 16, 2024, doi: 10.3390/w16162224.

[8] M. K. Mehla, "Regional water footprint assessment for a semi-arid basin in India," vol. 2031, 2022, doi: 10.7717/peerj.14207.

[9] N. Padriya and N. Patel, "Predicting yield of crop type and water requirement for a given plot of land using machine learning techniques," *Int. J. Reconfigurable Embed. Syst.*, vol. 12, no. 3, pp. 503–508, 2023, doi: 10.11591/ijres.v12.i3pp503-508.

[10] L. K. Joshila Grace, M. Paul Selvan, and A. Professor, "Price Prediction and Crop Yield using Machine Learning Algorithm," 2023.

[11] R. Chatterjee, O. Das, R. Kundu, and S. Podder, "Machine Learning Inspired Smart Agriculture System with Crop Prediction," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 1, pp. 1511–1517, Jan. 2023, doi: 10.22214/ijraset.2023.48841.

[12] S. L., Sangamesh, P. Kumar, and Supriya B. N., "Rice crop yield prediction using machine learning techniques," *Int. J. Adv. Res. Ideas Innov. Technol.*, vol. 5, no. 3, pp. 1037–1039, 2019.

[13] A. A. Lotfy *et al.*, "Forecasting Blue and Green Water Footprint of Wheat Based on Single, Hybrid, and Stacking Ensemble Machine Learning Algorithms Under Diverse Agro-Climatic Conditions in Nile Delta, Egypt," *Remote Sens.*, vol. 16, no. 22, 2024, doi: 10.3390/rs16224224.

[14] M. Kuradusenge *et al.*, "Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize," *Agric.*, vol. 13, no. 1, pp. 1–19, 2023, doi: 10.3390/agriculture13010225.

[15] A. S. Sreerama and B. M. Sagar, "A Machine Learning Approach to Crop Yield Prediction," no. May, pp. 6616–6619, 2020.

Research Article

[16] D. M. M. Ms.V.Mahalakshmi, "APPLICATION OF MACHINE LEARNING ALGORITHMS IN LOCATION SPECIFIC CROP YIELD PREDICTION," *Int. J. Res. Comput. Appl. Robot.*, vol. 11, no. 12, pp. 1–15, 2023.

[17] D. S. Kashmer and S. A. Abed, "the Agricultural Water Footprint of Al-Najaf Governorate, Iraq," *J. Agric. Biol. Sci.*, vol. 2, no. 6, pp. 1–20, 2024, [Online]. Available: https://webofjournals.com/index.php/8/article/view/1612

[18] O. A. Wani *et al.*, "Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-77687-x.

[19] O. Mialyk, J. F. Schyns, M. J. Booij, H. Su, R. J. Hogeboom, and M. Berger, "Water footprints and crop water use of 175 individual crops for 1990–2019 simulated with a global crop model," *Sci. Data*, vol. 11, no. 1, pp. 1–16, 2024, doi: 10.1038/s41597-024-03051-3.

[20] Google Earth Engine, "Google Earth Engine." Accessed: Oct. 15, 2024. [Online]. Available: https://earthengine.google.com/