DOI: http://dx.doi.org/10.26483/ijarcs.v16i3.7250

ISSN No. 0976-5697

Volume 16, No. 3, May-June 2025

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AN INTELLIGENT APPROACH FOR CROP WATER FOOTPRINT PREDICTION

Preethi Harris, Priyaharshini E and Varshika S
Sri Ramakrishna Engineering College,
Coimbatore, India

Abstract: Agriculture is the largest consumer of water; enhancement of the water level irrigation is essential for sustainability. This project employs the Random Forest Regressor for crop specifications as per hectare with considering the features such as crop type, seasonal data, location and meteorological data. To improve the robustness of the model performance, data preprocessing, Feature Engineering and Exploratory Data Analysis are used. The trained model is incorporated with a Flask Based web application, enabling the user, farmer, researchers and policymakers to custom their inputs and obtain their regional and crop specific predictions of water footprint. An in-built water calculator helps in manual estimations of predicting the water level required by specific crops along with yield area in cubic meters. By the combination of Machine Learning with user interface, it helps in the prediction of water footprint by considering the different features and improving the water conservation.

Keywords*:* Water Footprint, Irrigation, Crop Yield Prediction, Random Forest Regressor, Machine Learning, Flask Web Application

1. INTRODUCTION

Agriculture accounts for 70% of global issues of withdrawing the freshwater, making the enhancement of water irrigation practices provides a crucial stage in achieving the agricultural productivity. Other Traditional methods of irrigation are defective for predicting the water footprint with the diverse climatic conditions and specific crops across specified regions. In response the machine learning (ML) method provides an active tool for the optimization of decision making in precision agriculture.

The basis of this project is steams from Kulkarni et al. [1], who validated the techniques of Machine Learning in agriculture by prognosticating the crop yields using climatic and pesticide data. Expanding on such intelligent modelling, researchers have developed resilient Machine Learning techniques for predicting the yield and classification of crop aimed at sustainability [2]. Furthermore, extensive evaluations of ML approaches for smart agriculture system [3] and that was targeted on yield prediction studies specifically on cotton [4] which visualizes the versatility of these models. Recent critiques on developing the human-assist and digital technologies also emphasize the increasing role of intelligent systems in sustainable farming. Human involvement remains central to the chosen and success of sustainable agricultural technologies. Emerging tools aim to complement, rather than replace human decision-making. This not only give about the prediction of water footprint but also helps the farmers, researchers and policymakers to empower with each other along with the agricultural needs [5].

Kashyap et al. [6] presented an IoT-based intelligent irrigation system using deep learning method called Deep Neural Network (DNN) to maximize water use efficiency and automate the scheduling of irrigation. Concurrently, Jain et al. [7] discussed how agro-inundation practices, with integrated optimized irrigation design and crop selection approaches, can remarkably raise yield and water efficiency. The importance of mathematical and scientific modelling in determining the boundaries of irrigation can also be identified through the research work of Turaev et al. [8], which outlines algorithms to determine the limits of operation in irrigation networks. Machine learning methods have emerged to the forefront to maximize water allocation in agriculture. Kanmani et al. [9] applied data-driven methods to distribute water resources more effectively in agricultural environments to facilitate real-time decision- making. Similarly, the Hydro-Sense system by Thigale et al. [10] is an ideal example of employing sensor networks and data analytics to monitor soil moisture levels and maximize irrigation. Regional crop water footprint is provided by Rodríguez et al. [11] who estimated Argentinean rainfed crops' green water footprint with climate variability. Monchusi et al. [12] also further explained the combined application of IoT and AI to monitor and manage water, especially in small farms, for improved precision agriculture practice. The sensitivity of crop water needs to climate change is illustrated by Suwannakhot et al. [13], who looked into glutinous rice farming in Thailand and proposed adaptation strategies to mitigate expected threats. Furthermore, decision support models for considering crop seasonality and climatic conditions for sustainable irrigation planning are gaining prominence, as shown in a study by Díaz et al. [14].

Advanced machine learning and control systems such as those proposed by Agyeman et al. [15] unlock new opportunities for intelligent irrigation scheduling. On the micro- level, Geerthana et al. [16] reported an AI-supported model for determination of the crop water footprint with a digital solution for resource planning. Finally, the case study by Saicharan and Shwetha [17] quantifies rice's virtual water content in Mysore district, giving one an idea about region-specific consumption patterns of water.

When compared to the earlier research that generally concentrated on yield prediction, irrigation control or water footprint estimation specific to regions, this work proposes a inclusive method of crop water footprint prediction using a variety of features in the datasets like rainfall, irrigation sources, pesticides used and crop production specific to India. Although previous studies have already shown good efficiency of machine learning in agriculture, fewer studies combine more than one parameter of agriculture across seasons and regions for good estimation of water footprint accuracy.

In this we have used Decision Tree and Random Forest algorithm integrated with K-Fold Cross Validation is used to prevent overfitting and to ensure robust performance. In fact, the results obtained were 1.00 for Decision Tree and 0.99 for Random Forest, respectively. In addition, another innovation of this work is in the creation of a user- interactive web- based platform through which farmers, researchers and policy-makers can feed crop and regional information to get precise water footprint forecasts. This project address about the voids between perception and practical execution by providing the exceptional, scalable and accessible solution for supporting imperishable irrigation planning and water resource management.

2. PROJECT DESCRIPTION

The Project, An Intelligent Approach for Crop Water Footprint Prediction Leverages Machine Learning and web- based technologies to calculate the irrigation requirements of various (water footprint types such as blue and green water footprint) specified crops based on specific geographical and climatic conditions. The idea was to identify and develop a predictive model that use Random Forest Regressor to analyze the agricultural inputs such as state, district, crop type and season and outputs the calculated water usage per hectare. The system aims to support environmental-safe irrigation planning by providing the farmers and policymakers a fast tool access water need. The design begins with the collection and preprocessing the datasets with the cleaning techniques in accordance with the varieties of features and characteristics of crop and environment. Using Pandas and Numpy, the data gets cleaned and the features are getting encoded with the numeric values to endure the adaptability and compatibility with the Machine Learning model.

2.1 Block Diagram

Figure 1 depicts the system architecture for predicting the Water footprint of crops using the Machine Learning tool. The process begins with user input, where vital agricultural parameters that are provided through a web interface. These inputs are then preprocessed and cleaned to endure the continuous state in formatting and to handle any missing or noisy values. Preprocessing may include normalization, encoding of categorical variables into numbers and scaling of numerical data to prepare it for model consumption.

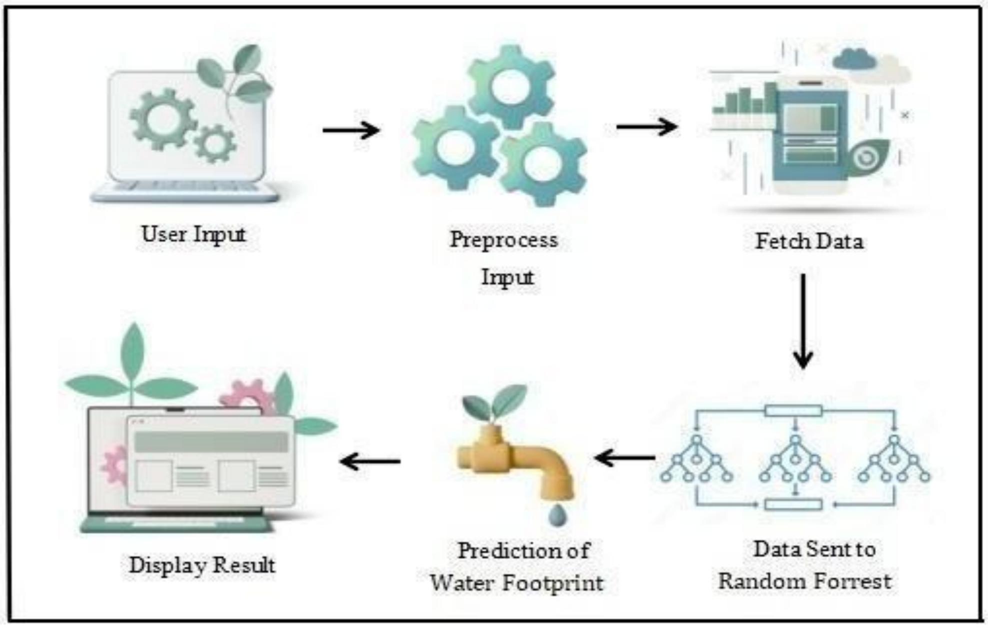

Figure 1 System Architecture for Crop Water Footprint Prediction

2.2 Working Principle

Input Data Collection
The process begins with user input, where users provide vital agricultural data with varieties of features such as state, district, crop type and cultivation area. These serve as a basic parameter for predicting the water requirements and are entered via a user-friendly web interface.

Data Preprocessing
Once the data collected, the next phase is data

preprocessing, that checks on the input values that it is cleaned, standardized and compatible.

The steps involve:

1.Encoding variables on different categories into numerical values. 2.Normalizing numeric features to remove the redundancy.

3. Handling the missing values.

4. Ensuring the formatting of different data.

Preprocessing is vital for optimizing the accurate prediction level and stability of the model and check on the obtaining results.

Data Fetching

After preprocessing, the historical data are needed to be fetched along with the environmental data relevant to the given inputs by the user.

Historical irrigation requirements. Crop yield data. Meteorological and geographical conditions.

These steps are required for the enrichment of the relevant background information that enables the high accuracy level.

Model Predictions Using Random Forest

After pre-processing, the dataset is pre-processed and enriched in terms of useful features and fed to a trained Random Forest regressor model. Random Forest algorithm is selected for its better accuracy and noise-robustness and the ability to minimize overfitting based on ensemble learning. This model appears to be able to comprehensively represent complex, non-linear relationships of a large number of input features by building several decision trees with their outputs averaged together. To improve its predictive performance, the model is fine-tuned using methodologies such as k-fold cross-validation and tested based on the key metrics including Root Mean Square Error (RMSE) and the coefficient of determination (R^2 score), thus inferring reliability and generalizability.

Processing and Interface

The whole prediction workflow is easily embedded into a simple to use web application built using the Flask framework. This interface eases the flow of interaction between the user and the backend model. Essential technologies in the areas of Pandas and NumPy are utilized for effective data manipulations, computation of numericals, while Scikitlearn is used for model training and evaluation. Furthermore, Matplotlib also provides the possibility to generate useful visualizations to help interpret the predictions. This architecture provides ease of use as well as effective deployment of the machine learning pipeline.

3. RESULTS

In Figure 2, one can observe the correlation heatmap for numeric features used in the analysis. From this visualization, there are linear relationships between key variables such as Area, Production, Rainfall (mm), Estimated Irrigation Requirement (cu.m), Crop Water Productivity (CWP), Rainfall Contribution and Blue Water Footprint (BWF). Interestingly, Area has a high positive correlation with Estimated Irrigation Requirement (0.98) and BWF (0.98) implying that high cultivation areas are closely linked to high water needs and high use of blue water. Similarly Estimated Irrigation Requirement is very much correlated with BWF (1.00) indicating their direct dependency. Rainfall (mm) has weak to moderate positive correlations with Rainfall Contribution (0.26) and CWP (0.09), indicating that natural precipitation as a contribution contributes partially in increasing crop water efficiency as well as overall water contribution. Surprisingly, CWP is associated with few or no variables such that water productivity is regulated by complex or crop specific factors. These correlations reveal useful information about interdependency between features and are essential in selecting predictors for machine learning models for water footprint estimation.

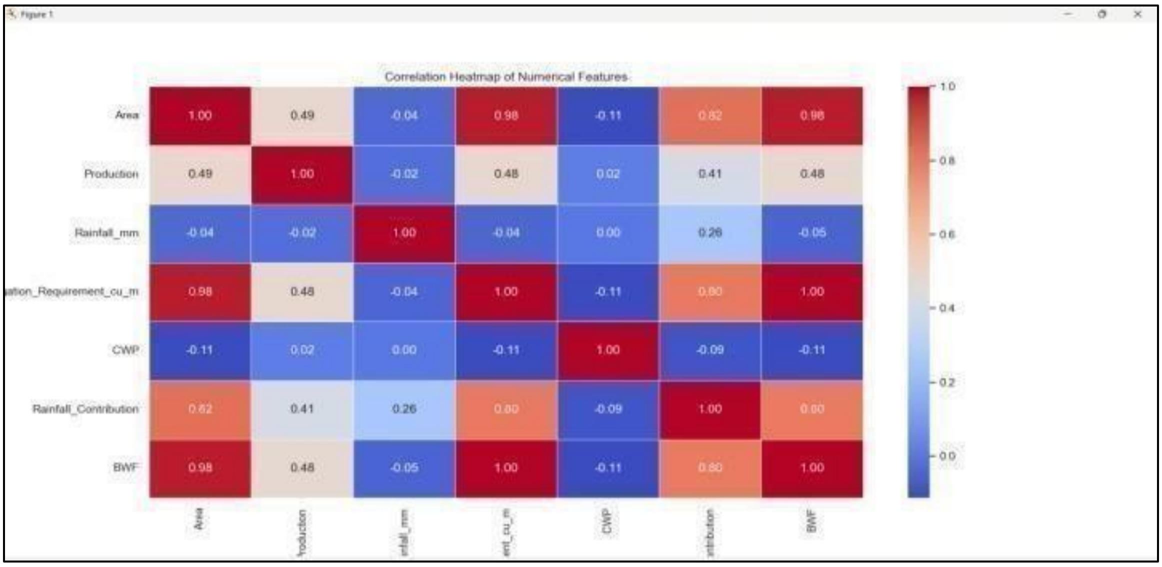

Figure 2 Feature Correlation Heatmap

Figure 3 presents the distribution plots for various features used in the water footprint estimation model. The features visualized include Area, Production, Rainfall (mm), Estimated Irrigation Requirement (cu.m), Rainfall

Contribution, CWP (Crop Water Productivity) and BWF (Blue Water Footprint). Most distributions, such as those for Area, Production, Estimated Irrigation Requirement, Rainfall Contribution, BWF and CWP, exhibit strong right-skewness, indicating that a majority of the values are concentrated near the lower end, with a few extreme values extending toward the higher end. This suggests the presence of outliers or uneven data spread, which is common in agricultural datasets due to regional or crop-specific variations. In contrast, the distribution of Rainfall (mm) appears relatively more uniform, with multiple peaks, suggesting seasonal or regional fluctuations in rainfall patterns. These plots provide critical insights into data behavior and helped guide further preprocessing steps, such as outlier handling and normalization, before model training. Understanding these distributions is essential for accurate prediction and interpretation of water footprint metrics in agriculture.

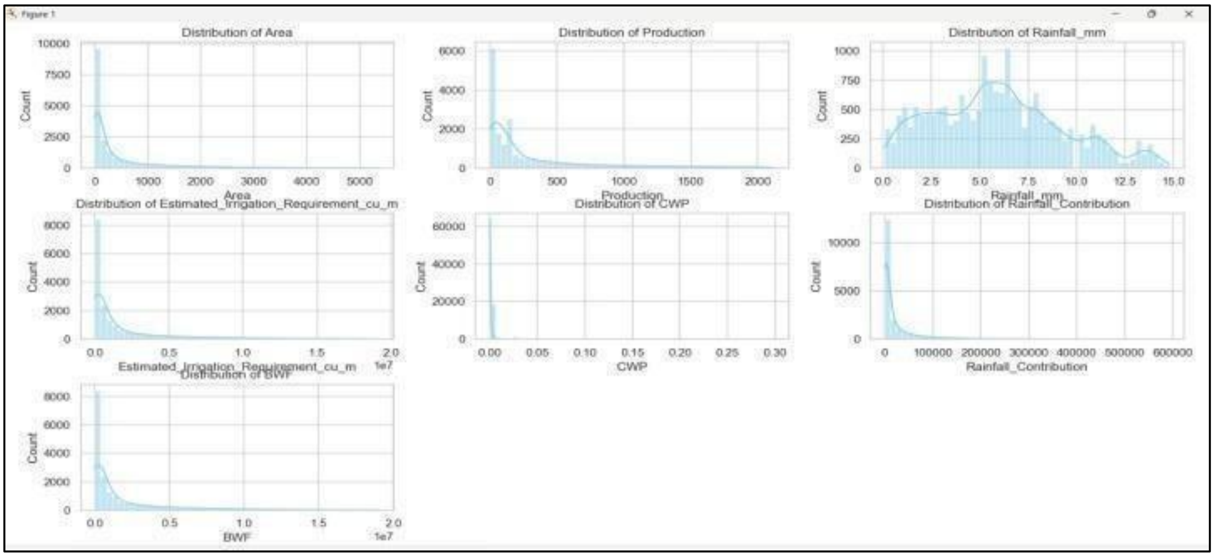

Figure 3 Distribution of Key Numerical Features

The Actual vs. Predicted Irrigation Requirement plot (Figure 4) shows how well the machine learning based model performs in determining irrigation water demand (in cubic meter). Each point plotted here corresponds to a single prediction; the x value indicates the actual irrigation required and y is the predicated value. The points are tightly packed deviating from the diagonal line (ideal prediction line), indicating that there is much proximity being shown in the predicted to real values. This alignment means that the model learned from the training data the lower-level structures well and can make [fairly] good predictions with little error. There are a few outliers which can be seen which is more at higher values which indicate a possible area for improvement or perhaps they are just worth looking into these particular instances. In summary, the visual correlation supports the model's robustness and its appropriateness for real-world forecasting of agricultural planning irrigation needs.

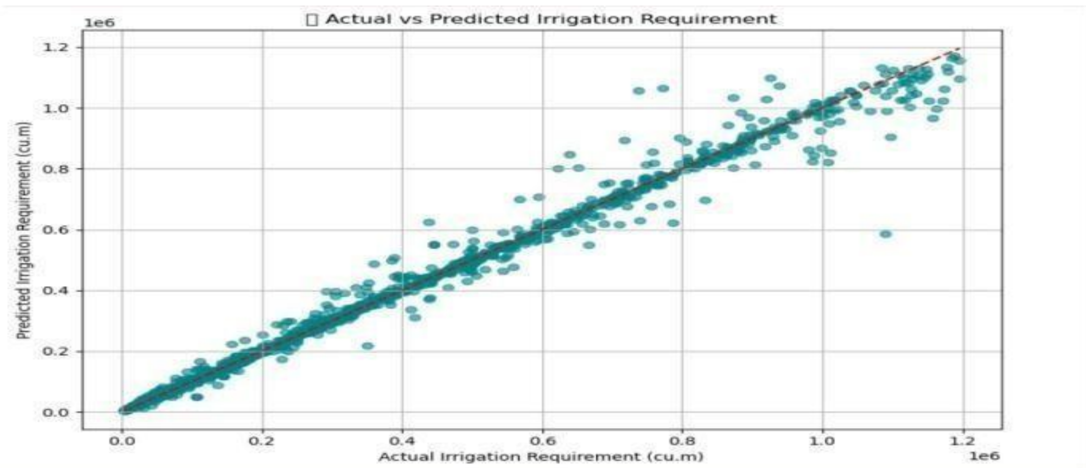

Figure 4 Actual vs Predicted Irrigation requirement

The system is successfully developed to compute the crop-specific water footprint for irrigation by employing a machine learning-based regression method. A Flask-based web interface is provided to accept input parameters like the name of the crop, state, district and hectares of area and display prediction of the amount of irrigation water needed in cubic meters. As can be observed in Figure 5, the user interface is simple and intuitive to facilitate a smooth interaction between farmers and planners.

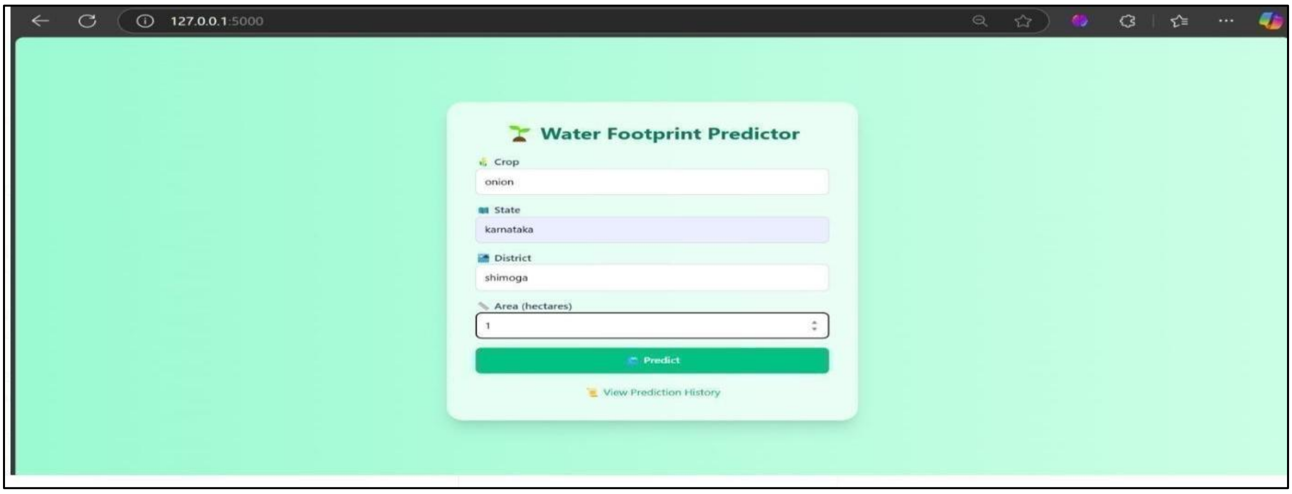

Figure 5 Input from user

Figure 6 is a picture of the final output screen of the Flask-based web-application developed for estimating the water footprint of agricultural crops. The interface shows predicted water footprint in cubic meters depending on the user-input considering such parameters as crop type, state, district and area expressed in hectares. In this case, the user had chosen onion as crop in Shimoga district of Karnataka state in 10 hectares of area. The system computed a water footprint of 4944.29cubic meters. The interface also has two visual components to make user engagement better. An image about the selected crop that is a visual image of the crop. A state map image showing the selected state, pointing to a contextualized geographic territory. Moreover, water is precious message reminder “Water is precious. Let's not waste it." is added for better awareness of the practice of the sustainable use of water. Users have the choice of prediction again and viewing history making the system interactive and useful for iterative usages by farmers and planners.

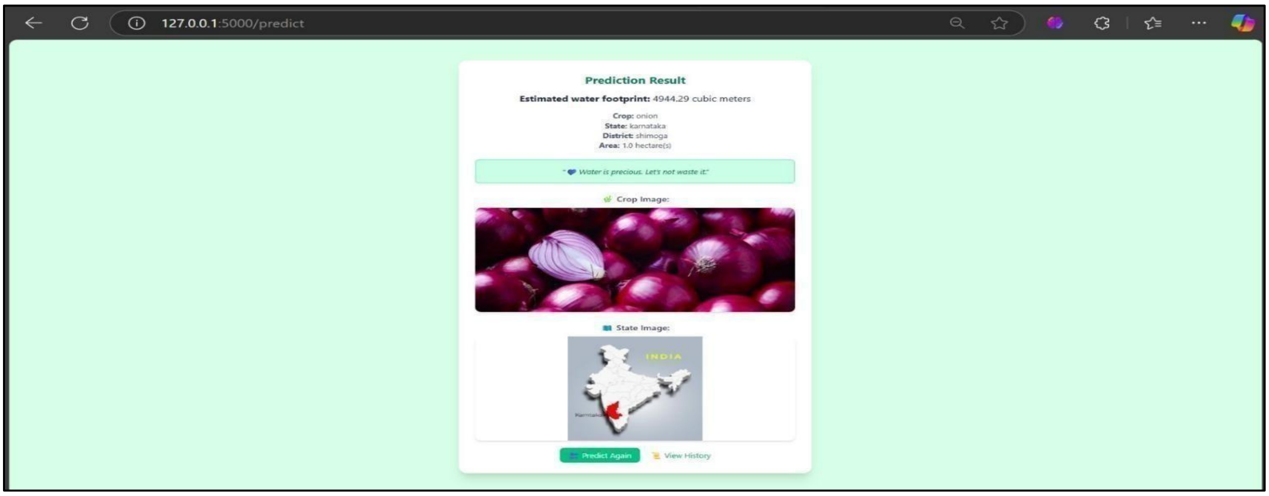

Figure 6 Water Prediction of Onion

The performance of the classification models was evaluated using confusion matrices for both the Decision Tree and Random Forest algorithms, as shown in Figure 7. Each confusion matrix provides insight into the model's classification accuracy across four categories: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). In the Decision Tree confusion matrix, the model achieved 1909 true negatives (TN), correctly identifying non-irrigation-requiring instances and 4 true positives (TP), accurately predicting irrigation-requiring instances. Importantly, there were no false positives (FP) and no false negatives (FN), indicating a perfect classification performance on this dataset.

In contrast, the Random Forest confusion matrix also recorded 1909 TNs and 3 TPs, but it had 1 false negative (FN) and 1 false positive (FP). Despite this very slight deviation from the perfect score of the Decision Tree, the Random Forest model was chosen for deployment due to its greater generalization ability and robustness to overfitting. Decision Trees are prone to overfitting, particularly with small or imbalanced datasets, as they can memorize training data patterns too precisely. Random

Forest, by aggregating multiple decision trees and introducing randomness, tends to offer better performance on unseen data, making it a more reliable model in real-world applications. This trade-off between marginal loss in perfect classification and improved model stability strongly justifies the selection of Random Forest for this project.

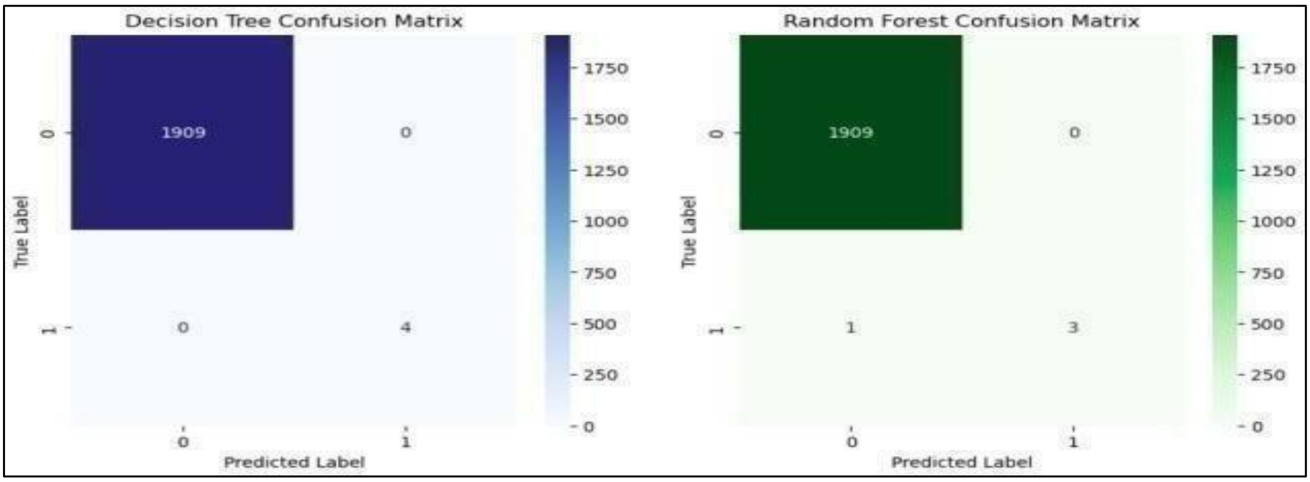

Figure 7 Confusion Matrix

The Table 1 compares Random Forest Regressor and Decision Tree Regressor were also matched based on crucial performance measurements and characteristics of models. The Random Forest Regressor scored an incredibly high accuracy of R^2 0.9999, defeating yet again the Decision Tree which had an earlier R^2 score of 0.9732. Regarding error, Random Forest evidenced extreme low Mean Absolute Error, whereas the Decision Tree proved to have significantly high error in comparison. When talking about generalization, Random Forest turned out to be excellent and manage to minimize overfitting by merging results of multiple trees, whereas the Decision Tree is more prone to overfitting as it relies on a single tree. Interpretability plays in favor of the Decision Tree since the latter's simpler structure makes it easier to interpret, and the Random Forest endows only moderate interpretability. However, the network's improved performance on Random Forest comes with a higher inference timing in contrast to Decision Tree models which deliver rapid evaluations thereby better fitting timing sensitive applications, presents the detailed comparison between the Random Forest Regressor and the Decision Tree Regressor based on several criteria used in the crop water footprint prediction system. By comparing the both Regressor techniques Random Forest Regressor is more effective and efficient than Decision Tree Regressor.

Table 1 Random Forest Regressor vs Decision Tree Regressor

CRITERIA	RANDOM FOREST REGRESSOR	DECISION TREE REGRESSOR
Accuracy R^2 score	0.9999	0.9732
Mean Absolute Error	Very low	Higher than RF
Generalization	Excellent flow overfitting	Moderate prone to overfitting
Interpretability	Moderate	High
Inference Time	High	Low

The Result of the project demonstrate that the proposed system accurately predicts crop-specific water requirements using Random Forest regression model. Overall, the system proves to be a fast, user-friendly and practical tool for sustainable irrigation planning.

4. CONCLUSION

The Machine learning tool is used to identify the accurate estimation of water footprint to maintain the environmental-safety surrounding and also it calculates water for particular crop with specified geographical region. Machine Learning utilizes the Random Forest Regressor to predict the water demand and it shows high accuracy with 0.9999 score. The model analyses the most significant attributes to accurately estimate blue and green water footprints per hectare. The Flask-based web application provided in the package provides an easy-to-use interface where users can input region and crop details to receive instant water requirement predictions. It also provides the history of previous predicted water footprint which is useful for farmers, researchers and policymakers. Specialized though it is on farm irrigation planning, principles and the model may apply to other areas like forestry, urban water management, or climate-based resource forecasting in future. Through its usability, scalability and accuracy, the project is a resilient testament to the ways in which the Machine Learning tools are used and support for the sustainable agricultural practices that includes the water management.

REFERENCES

[1] S. Kulkarni, M. Shah and K. Patel, “Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning” IEEE Access, vol. 12, no3, pp. 45-58,2023.

[2] A. Badshah, B. Y. AL Kazemi, F. Din, K. Z. Zamli and M. Haris, “Crop Classification and Yield Prediction Using Robust Machine Learning Models For Agricultural Sustainability” IEEE Access, vol. 12, pp. 127799– 127812, 2024.

[3] G. Mohyuddin, M. A. Khan, A. Haseeb, S. Mahpara, M. Waseem and A. M. Saleh, “Evaluation of Machine Learning Approaches for Precision Farming in Smart Agriculture System: A Comprehensive Review” IEEE Access, vol. 12, pp. 60155–60184, 2024.

[4] Alakananda Mitra, Sahila Beegum, David H. Fleisher, Vangimalla R. Reddy, Wenguang Sun, Chittaranjan Ray, Dennis Timlin and Arindam Malakar, “Cotton Yield Prediction: A Machine Learning Approach With Field and Synthetic Data,” IEEE Access, vol. 12, pp. 101273–101288, 2024.

[5] M. Louta, K. Banti, I. Karampelia, “Emerging Technologies for Sustainable Agriculture: The Power of Humans and the Way Ahead” IEEE Access, vol. 12, pp. 1– 1, 2024.

[6] P. K. Kashyap, S. Kumar, A. Jaiswal, M. Prasad and A. H. Gandomi, "Towards Precision Agriculture: IoT- Enabled Intelligent Irrigation Systems Using Deep Learning Neural Network" IEEE Sensors Journal, vol. 21, no.16, pp. 17608–17615, 2021.

[7] A. Jain, A. Asati, A. Kumar, S. Mahajan and M. Dhone, "Agro-Inundation for Maximizing Crop Yield and Water Efficiency" in Proc. INSCIRD-2023, SSGM Journal of Science and Engineering, vol. 1, no. 1, pp. 10–14, 2023.

[8] R. Turaev, A. Seytov, S. Kuldasheva and U. Nortojiev, "Algorithms for calculating limits in water management in irrigation systems" in E3S Web of Conferences, vol. 401, 5th Int. Sci. Conf. on Construction Mechanics, Hydraulics and Water Resources Engineering (CONMECHYDRO 2023), Tashkent, Uzbekistan, 2023, Art. no. 02016

[9] R. Kanmani, D. JeyaBharathi, G. Siva Suriya, A. C. Jeba Malar and G. [Last Name], "Distribution of Water Using Machine Learning and Data Analytic Techniques for Agricultural Purposes" in Proceedings of the 2021 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 944–947

[10] O. Thigale, R. Suryawanshi, S. Telsang, M. Waghmode, M. Thakare and V. Wankhade, "Hydro-Sense: Precision Agriculture and Water Conservation" in Proc. Int. Conf. Innovations in Engineering and Technology (ICIET), Vishwakarma Institute of Technology, Pune, India, 2023.

[11] C. I. Rodríguez, M. M. Arrien, S. H. Silva and M. M. Aldaya, "Global relevance of Argentinean rainfed crops in a climatic variability context: A water footprint assessment in Buenos Aires province" Science of The Total Environment, vol. 171946, 2024.

[12] B. B. Monchusi, A. T. Kgopa and T. I. Mokwana, "Integrating IoT and AI for Precision Agriculture: Enhancing Water Management and Crop Monitoring in Small-Scale Farms" in Proc. Int. Conf. Innovative Computing Applications (ICONIC), Mauritius, 2024.C.

[13] Suwannakhot, N. Pumijumnong, S. Sereenonchai, W. Chaowiwat and N. Arunrat, "Assessment of Climate Change Impact on Water Footprint of Glutinous Rice (RD6) Production and Adaptation Strategies: A Case Study of Udon Thani Province, Thailand" in Proc. 2022 Int. Conf. and Utility Exhibition on Energy, Environment and Climate Change (ICUE), Pattaya City, Thailand, 2022.

[14] M. R. Díaz, A. M. García and R. J. Pérez, "Profit and Sustainable Water Optimization for Irrigated Crop Planning Considering Environmental Conditions and Crop Seasonality" in Proceedings of the 2024 IEEE Biennial Congress of Argentina (ARGENCON), San Nicolás, Argentina, 2024.

[15] B. T. Agyeman, M. Naouri, W. Appels, J. Liu and S. L. Shah, "Integrating machine learning paradigms and mixed-integer model predictive control for irrigation scheduling" arXiv preprint arXiv :2306.08715, 2023.

[16] G. Geerthana, V. Hamsavarthini, A. Hariprasath and R. Anuradha, "AI Based Water foot Print of Crop" in Proceedings of the 2024 5th International Conference on Smart Electronics and Communication (ICOSEC).

[17] V. Saicharan and H. R. Shwetha, "Quantification and Assessment of the Virtual Water Content of Rice Crop: A Case Study of Mysore District, India" in Proceedings of the 2023 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), Bangalore, India, 2023, pp. 1–4