Evaluating the Effectiveness of Soil Amendment Practices on Groundwater Recharge and Irrigation Sustainability in Semi-Arid Farmlands

AHMED MOHAMED IDRIS ABIDAT
College of Agricultural Technology and Animal Health, Wadi Ataba. Libya. Soil and water Department
E-Mail-ID: ahmeedmahameed1@gmail.com

Abstract-- Irrigation requirement prediction plays a crucial role in increasing the efficiency of farming and sustainable utilization of groundwater resources in semi-arid lands. Prediction of irrigation requirements is difficult because of various factors involved in the process like interaction between soil type, climate, and crop type. In this research, Hybrid Feature Selection-Ransom Forest (HFS-RF) scheme is proposed to improve the irrigation requirement prediction using the Irrigation Water Requirement Prediction Dataset. Firstly, pre-processing is done in the form of dealing with missing values, duplicate removal, label encoding, and Min-Max normalization. Further, HFS scheme, which is based on the use of correlation analysis and RF feature importance, is utilized to select the optimal features having an impact on irrigation demand. The selected features are used to train thevRFclassifier, which classifies the irrigation requirements into Low, Medium, and High. The overall accuracy obtained from the proposed scheme is 97.00% with 98.00% recall, 97.00% precision, and 97.00% F1-score. Thus, the proposed scheme outperforms other schemes. The experimental results clearly reveal the effectiveness of the HFS-RF scheme in enhancing the prediction accuracy and decreasing the computational complexity.

Keywords-- Hybrid Feature Selection (HFS),Random Forest, Irrigation Requirement Prediction, Sustainable Irrigation Management, Precision Agriculture.

I. INTRODUCTION

Water shortage and groundwater depletion are important issues that affect agricultural production in semi-arid areas, and in such regions, irrigation management becomes an important task for sustainable agriculture. Accurate estimation of irrigation needs is helpful to achieve optimal utilization of water resources and increase crop yield. Machine Learning (ML) can be considered an effective tool in order to discover complicated correlations between various soil, climate, and crop parameters. The accurate prediction of water needs of crops plays an important role in proper irrigation planning in protected agriculture. Existing traditional techniques like FAO Penman-Monteith models use many input factors that make it difficult to apply in practice. The current paper suggests a new multi-source data fusion method combining canopy cover obtained from images with environment data and ML algorithms. The proposed stacking data fusion technique shows better accuracy of prediction. [1]. The accurate computation of the reference evapotranspiration (ETo) value is very important to help us manage irrigation and water resources effectively, particularly in areas where meteorological information is scarce. The ML algorithms Decision Tree, K-Nearest Neighbor, and RF were used in this study to predict the daily values of the reference ETo based on various input variables. The RF algorithm performed the best and determined that the temperature and solar radiation were the most important variables.[2]. Correct estimations of evapotranspiration fraction (ETf) and Evapotranspiration Actual (ETr) are important in ensuring that irrigation is carried out effectively and precisely. In this research, ML models have been developed to estimate ETf and ETr using spectral bands from Sentinel-2, where no use of thermal bands was involved. The model used that had the highest efficiency in predicting ETf and ETr was Cubist, while the Eto Brazil model was more accurate than Hargreaves-Samani in predicting ETr.[3].

The smart irrigation is extremely crucial in sustainable agriculture and effective utilization of water resources. In this research paper, a new concept of smart irrigation is presented which makes use of the internet of things (IoT). The system works on data collected through sensors and ML/deep learning algorithms to predict the irrigation requirement. The proposed algorithm uses RF and BiLSTM for irrigation prediction. [4]. Smart precision farming makes use of IoT technologies, ML, and remote sensing techniques for efficient crop production and water saving. A model incorporating IoT is used in this study to accomplish the task of land mapping, crop prediction and irrigation management by using satellite data, LSTM, RF and fuzzy logic.. The model successfully predicts crop suitability and irrigation needs with a considerable decrease in water usage.[5]. Soil moisture is an important aspect of water management and irrigation systems in agriculture. This research explores some advanced ML methods to estimate soil moisture based on Sentinel-1 SAR data for wheat cultivation areas. Of all the methods evaluated in this research, RF proved to be the most accurate and robust, giving the best results compared to other algorithms during training and testing phases.[6]. An accurate estimation of water requirements of crops is important for optimal irrigation and water management. The following research study will involve the use of ML techniques to forecast water requirement in relation to agricultural and environmental parameters. Through identification of critical parameters that affect irrigation requirement, these models will help achieve better prediction accuracy, effective water use, and precision agriculture.[7].

A. *Key contributions*

- Presented the HFS-RF approach for accurate prediction of irrigation needs through soil, climate, and agriculture-related features.
- Enhanced the accuracy of irrigation prediction through feature selection, avoiding redundant information, and improving RF classifier performance.
- Designed a decision support system for optimal irrigation scheduling, water saving, and sustainable use of groundwater in semi-arid agriculture environments.

II. RELATED WORKS

Li et al., suggested the integration of feature selection techniques using Pearson Correlation (PC), RF Importance, and Recursive Feature Elimination (RFE). This integration would be used to select the best vegetation indices for predicting winter wheat yields. The selection of these features resulted in better performance in the models, especially for the Cubist model, which outperformed the RNN model. [8]. Tausif et al., proposed research has provided a model for federated learning to estimate reference ETo for different agricultural lands under varying climatic conditions. RF Regressor proved to perform the best when compared to other models like Support Vector Regression and Decision Trees. It has been found that maximum temperature and wind speed play an important role in ETo estimation.[9].Hendy et al., explores the possibility of estimating reference ETo values in arid areas through the use of ML models and time series analysis. The performance of RF, LSTM, and ARIMA models was compared using weather data obtained from Egypt. The results revealed the high performance of RF and LSTM, whereas ARIMA performed well for time series predictions.[10]. eddy et al. discusses the development of an IoT smart irrigation system that aims to effectively control the usage of water in agriculture. Using soil moisture sensors, the NodeMCU module, and wireless communication, this IoT-based system is capable of making use of weather predictions and ETo in order to control the irrigation process. [11]. Taueatsoala et al., proposed an IoT solution for precision irrigation that is powered by TinyML. This framework aims to help out communities that are facing resource constraints. The IoT framework combines sensors, ESP32, and Raspberry Pi in order to provide decision-making locally and without any dependence on cloud computing. The machine learning model designed using Gradient Boosting gave the highest accuracy result and was used in TinyML to predict irrigation accurately.[12].

BP et al., suggest an innovative approach for developing an intelligent irrigation system through the integration of IoT and ML technology for the effective utilization of water in agriculture. Using the information available on various environmental factors such as moisture content, temperature, pH levels, and types of soil, the authors have constructed. The suggested model provided high accuracy and better adaptability, thus optimizing irrigation scheduling.[13]. Ali et al., examines the techniques used in ML algorithms for soil moisture prediction for the purpose of making precision farming and efficient water utilization possible. It is found that using the climate, soil, and vegetation databases, the machine learning algorithms including RF and XGBoost are better than conventional algorithms; in particular, RF showed the highest accuracy. The findings reveal that soil depth, leaf area index, and soil quality affect the forecasting accuracy. [14]. Roy et al., suggests an ML-based approach supported by the IoT paradigm for predicting water pumping needs in irrigation systems. The classification process is done using algorithms such as CatBoost, Random Forest, SGD, and KNN to classify either as "need for pumping" or "no need for pumping." This process uses cross-validation to evaluate the system, where the classification accuracy is the measure of performance.[15]

B. Problem statement

The conventional irrigation prediction models rely on using all the features available. The redundancy associated with the model leads to inefficiencies in computation and prediction accuracy. This has posed challenges in effective irrigation scheduling and water management in semi-arid agricultural zones. The HFS-RF algorithm tackles the problem through appropriate selection of the features.

III. PROPOSED HFS-RF METHODOLOGY FOR IRRIGATION REQUIREMENT PREDICTION

The workflow of HFS-RF Model involves collection of the irrigation dataset having attributes related to soil, climate, crop, and irrigation. Preprocessing of the dataset is done using missing values treatment, duplication, label encoding, Min-Max normalization, and exploratory data analysis (EDA). Feature selection involves hybridization of correlation feature selection method and RF feature importance method. These selected features are then provided to RFclassifier to perform the task of classifying the irrigation need as low, medium, and high based on majority voting. The model is built using 80% of the dataset, and 20% is used for evaluation of the model using precision, accuracy, F1 score, and recall. Figure.1 illustrate the proposed HFS-RF methodology for irrigation requirement prediction.

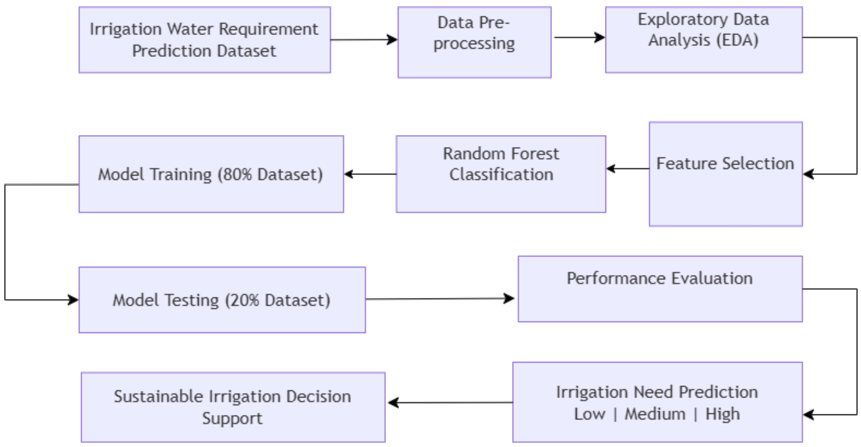


Figure.1 Proposed HFS-RF framework

A. Dataset Collection

The dataset of Irrigation Water Requirement Prediction that was utilized for the study was sourced from the Kaggle repository. It contains 10,000 records and 20 features that provide extensive details about different attributes such as soil characteristics, weather conditions, crop type and growth phase, and irrigation techniques. The set of predictors includes Soil pH, Soil Type, Soil Moisture, Organic Carbon, Electrical Conductivity, Temperature, Humidity, Rainfall, Sunlight Hours, Wind Speed, Crop Growth Stage, Crop Type, Irrigation Type, Season, Water Source, Field Area, Mulching, Previous Irrigation, and Region. The response variable is Irrigation Need, whose values fall into three categories: Low, Medium, and High. The dataset was chosen since it is quite recent, well-organized, and includes key environmental and agricultural factors needed for irrigation prediction. It was into the Python environment through the use of the Pandas library.[16]

B. Data Pre-processing

The data was pre-processed to ensure that it was of high quality and ready for ML analysis. The first step involved the identification of missing values, duplicates and inconsistency in the data. Missing numerical data was imputed using the mean value of that feature, while missing categorical values were imputed using the mode. Duplicates were also eliminated from the data. In addition, since some of the variables such as Soil Type, Crop Type, Season, Irrigation Type, Water Source, Mulching Used, and Region are categorical, they had to be converted into numeric data using Label Encoding. Label Encoding is performed using the formula as shown in (1)

$$LE(X) = \{0,1,2,\dots,n-1\} \quad (1)$$

Where $LE(X)$ is the encoded categorical variable. Following that, all numeric features were normalized via Min-Max Normalization in order to neutralize the scale disparity among the features. This procedure is described by the formula as shown in (2)

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Where X is the feature value whereas X_{min} and X_{max} denote the minimum and maximum values of that same attribute respectively. After normalization, the data was split up into 80 percent training and 20 percent testing data.

C. Proposed methodology of HFS-RF methodology for irrigation requirement prediction

The methodology comprises developing a methodology for predicting the irrigation requirements based on soil, climatic, crop, and irrigation management information using ML models. The approach involves five stages, including EDA, Feature Selection, Modelling, Model Training & Testing, and Performance Evaluation. Irrigation needs are categorized into three levels; low, medium, and high. Correlation between features is calculated using the expression as shown in (3)

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (3)$$

Whereas x_i and y_i are observations of two variables, $\bar{x}$ and $\bar{y}$ are the mean values of those variables respectively. The correlation coefficient can have a value between -1 and +1, reflecting the linear relationship between the variables. Feature selection follows after data exploration for selecting the best set of features that could be used in irrigation prediction. Feature selection helps to simplify computation, enhance accuracy in prediction, and reduce overfitting. The selected features include soil moisture, rainfall, temperature, humidity, organic carbon content in soil, soil type, crop type, crop growth stage, last irrigation amount, and field size because they are the features that impact the need for irrigation the most.

The selected set of attributes is used to train the RF algorithm, which is a kind of ensemble learning algorithm that generates multiple decision trees through bootstrapping and attribute sampling technique. In comparison with the individual decision tree algorithm, RF makes use of multiple decision trees and combines their output in order to increase accuracy and avoid over-fitting. The training of RF requires the generation of multiple decision trees, each of which is trained using a random subset of data and attributes. Node splitting is done by means of Gini Impurity Index, which is expressed as shown in (4)

$$Gini = 1 - \sum_{i=1}^{k} p_i^2 \quad (4)$$

Where p_i refers to the probability of class i at a certain node, while k refers to the total number of classes. A smaller Gini index implies greater separation between classes. At each stage, the decision tree algorithm chooses the attribute with the least impurity until the termination conditions are met. RF creates many decision trees before combining their results using majority voting. The irrigation class that will be predicted is derived as shown in (5)

$$\hat{y} = \text{Mode}(T_1(\alpha), T_2(\alpha), \dots, T_n(\alpha)) \quad (5)$$

Where $\hat{y}$ is the final irrigation class prediction obtained from the majority vote and $T_i(\alpha)$ is the i^{th} prediction of the decision tree, whereas the Mode gives the class having the maximum votes. This method enhances the robustness of prediction and minimizes classification mistakes. The preprocessed data set is partitioned into testing and training data set in an 80:20 ratio. Training data set is used to generate the RF model through the process of learning the relationship between soil, climate, irrigation variables, and crops while testing data set checks the generalization capability of the model. Ultimately, the performance of the model will be analysed by means of accuracy, recall, precision, and F1 measure, where accuracy is calculated using equation (6).

AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.3 (2026)
www.ajmimc.com

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Where TN, TP, FP, and FN stand for True Negatives, True Positives, False Positives, and False Negatives, respectively. The Precision measure determines the percentage of correct positive predictions among all positive predictions and is calculated as shown in (7)

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

The Recall score assesses the performance of the classifier on correctly predicting positive instances and is calculated as shown in (8)

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

F1-Score is the harmonic mean of Precision and Recall, which is calculated as shown in (9)

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

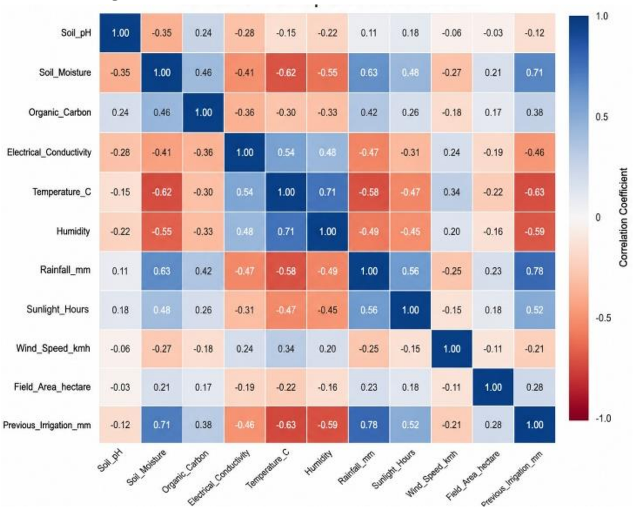
The combination of evaluation measures determines the accuracy and reliability of the RF algorithm modelThe classifier that produces the highest results in accuracy, precision, recall, and F1-score is identified as the optimal prediction model. After training, the model makes predictions of the need for irrigation in the new areas depending on the parameters such as soil type, climatic conditions, crops used, and irrigation records. The predicted value of the need for irrigation is then categorized as either Low, Medium, or High.

IV. RESULT AND DISCUSSION

The results of the experiment are presented in this part, where the HFS-RF framework was applied to predict the irrigation requirements. The performance of the proposed framework is evaluated using correlation analysis, feature importance analysis, confusion matrix analysis, classification metrics, and performance comparison. It can be clearly seen from the results that the proposed method for selecting features proves successful in selecting influential features and improving the predictive power of the RF algorithm. The proposed framework is evaluated using standard evaluation metrics such as Recall, Accuracy, Precision, and F1-Score to prove the effectiveness and reliability of the proposed model in predicting the requirements for irrigation. Moreover, the proposed framework is compared with other frameworks.

A. Correlation Analysis of Numerical Features

Figure 2 shows the correlation heatmap of the numeric features considered in the proposed irrigation prediction model. The correlation heatmap shows the level and nature of the relationship between soil-related, climate-related, and irrigation-related features with the help of correlation coefficient values between -1 and +1. Highly positively correlated features are displayed in dark blue, while highly negatively correlated features are displayed in dark red. It is found through the analysis that soil moisture, rainfall, temperature, and past irrigation have significant relationships, thus making them suitable features for the HFS-RF model.



AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.3 (2026)
www.ajmimc.com

Figure.2 Analysis of Selected Features

B. Feature Selection Analysis

The Table 1 below highlights the selection of features using the HFS technique through the importance scores. These selected features are the ones that have the most significant impact in the prediction of irrigation requirement and have been sorted based on the contribution of each feature to the RF classifier. Out of the chosen features, Soil Moisture was the one having the highest importance score of 0.243, which means that it is the one that influences irrigation requirement the most, followed by Rainfall (0.197) and Temperature (0.154). Some of the other important features are Previous Irrigation, Humidity, Soil Type, Crop Type, and Organic Carbon.

TABLE 1. SELECTED FEATURES USING HYBRID FEATURE SELECTION

Rank	Selected Feature	Importance Score
1	Soil Moisture	0.243
2	Rainfall (mm)	0.197
3	Temperature (°C)	0.154
4	Previous Irrigation (mm)	0.126
5	Humidity (%)	0.098
6	Soil Type	0.073
7	Crop Type	0.064
8	Organic Carbon	0.045

C. Analysis of Confusion Matrix of the Proposed HFS-RF Model

Figure 3 displays the confusion matrix of the proposed HFS-RF model to classify irrigation needs based on Low, Medium, and High levels. It is clear from the confusion matrix that the HFS-RF model has accurately classified 5795 Low, 3684 Medium, and 312 High irrigation requirements, making only a few errors in all three classes. Therefore, it can be concluded that the proposed HFS-RF model has achieved high classification accuracy.

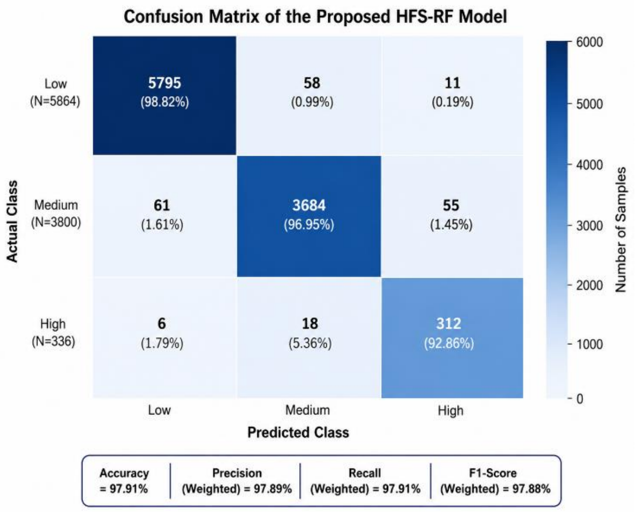


Figure. 3 Confusion Matrix of the Proposed HFS-RF Mode

D. Performance of the Classification of the Proposed HFS-RF Model

Table 2 illustrates the performance of the classification of the proposed HFS-RF model through Precision, Recall, F1-score, and Support. The model exhibited an outstanding level of accuracy of 97.00%, indicating a very strong ability of prediction. The Low category of irrigation was observed to yield the best results in terms of its performance with 99.00% recall, 98.00% precision, and 98.00% F1-score, while Medium and High irrigation also performed well in terms of classification. In addition, the Macro Average and Weighted Average of the model showed balanced prediction among all the irrigation categories.

TABLE 2 CLASSIFICATION REPORT OF THE PROPOSED HFS-RF MODEL

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Low	98.00	99.00	98.00	5864
Medium	97.00	96.00	96.00	3800

High	94.00	91.00	92.00	336
Accuracy (%)	97.00	–	–	10000
Macro Average (%)	96.00	95.00	95.00	10000
Weighted Average (%)	97.00	98.00	97.00	10000

E. *Comparison of Performance of Proposed HFS-RF Model and Existing Models*

The performance of the HFS-RF model versus existing irrigation prediction models such as HydroPredictor, RF, and XGBoost is given in Table 3. The proposed HFS-RF model offers the superior performance in terms of 97.00% precision, 97.00% accuracy, 98.00% recall, and 97.00% F1-score. It proves the ability of Hybrid Feature Selection to enhance the prediction capability of the RF classifier by selecting the most important features and removing redundant data.

TABLE 3 COMPARATIVE ANALYSIS

Models	Precision (%)	Recall (%)	Accuracy (%)	F1-score (%)
HydroPredictor [17]	-	-	89.23	89.37
RF [18]	91.9	97.1	94.3	94.4
XGBoost [18]	94.4	97.1	95.7	95.8
Proposed HFS-RF Model	97.00	98.00	97.00	97.00

The findings from the experiments conducted show that the proposed HFS-RF approach provides a reliable prediction of the amount of irrigation needed. Through feature selection, it eliminates redundancy, and therefore enhances classification efficiency to allow for effective irrigation scheduling, water resource management, and precision farming in semi-arid areas

F. *Discussion*

Experimental results show that the developed HFS-RF model accurately predicts irrigation needs by selecting the most significant soil, climatic, and irrigation variables. Soil moisture, precipitation, temperature, and past irrigation are some of the most significant variables according to the correlation analysis and the variable importance analysis. The confusion matrix and classification report show that the model is very accurate since it has an accuracy of 97.00%. The precision of the model is 97.00%, recall is 98.00%, and F1-score is 97.00%. It can be seen that the developed HFS-RF model has performed better than the current models.

V. CONCLUSION

This paper proposed the HFS-RF scheme for the prediction of irrigation requirement in semi-arid farming lands. In this framework, the combination of correlation analysis and feature importance by the RF was applied in order to select the most informative features of soil, climate, and crop. Thus, redundant features and high computational cost were reduced and eliminated. The selected features were employed in training the RF classifier to categorize irrigation requirements into Low, Medium, and High classes. From experimental results, it was evident that the performance of the suggested HFS-RF model is way better than other models such as HydroPredictor, Random Forest, and XGBoost with accuracy of 97.00%. Moreover, experimental results validated that soil moisture, rainfall, temperature, and previous irrigation are the most significant factors influencing the irrigation requirement. As the proposed framework provides irrigation predictions with high accuracy, it can be helpful for effective irrigation scheduling and water usage.

REFERENCE

[1] X. Lv *et al.*, "Maximizing multi-source data integration and minimizing the parameters for greenhouse tomato crop water requirement prediction," *Sci. Rep.*, vol. 15, no. 1, p. 29161, 2025.
[2] A. A. Farag, "ML approaches for enhanced estimation of reference evapotranspiration (ETo): a comparative evaluation," *Sci. Rep.*, vol. 15, no. 1, p. 38485, 2025.
[3] T. S. Costa, R. Filgueiras, R. A. Dos Santos, and F. F. da Cunha, "Actual evapotranspiration by ML and remote sensing without the thermal spectrum," *Plos One*, vol. 18, no. 5, p. e0285535, 2023.
[4] S. Neduncheliyan, T. Kazuaki, R. Krishnamoorthy, R. Thiagarajan, B. Swarna, and A. Kumar, "Comparative analysis of ML and novel deep learning models for IoT enabled smart irrigation water management system," *Discov. Internet Things*, 2026.
[5] G. Saha, F. Shahrin, F. H. Khan, M. M. Meshkat, and A. A. M. Azad, "Smart IoT-driven precision agriculture: Land mapping, crop prediction, and irrigation system," *Plos One*, vol. 20, no. 3, p. e0319268, 2025.

[6] D. Lakra, S. Pipil, P. K. Srivastava, S. K. Singh, M. Gupta, and R. Prasad, “Soil moisture retrieval over agricultural region through ML and sentinel 1 observations,” *Front. Remote Sens.*, vol. 5, p. 1513620, 2025.

[7] M. Nagappan, S. Jaya Murugan, and A. Kudiarasumani, “Advanced Prediction of Crop Water Requirements Using ML Models,” *Int. J. Innov. Sci. Res. Techno*, pp. 2518–2523, 2025.

[8] Z. Li *et al.*, “An integrated feature selection approach to high water stress yield prediction,” *Front. Plant Sci.*, vol. 14, p. 1289692, 2023.

[9] M. Tausif, M. W. Iqbal, R. N. Bashir, B. AlGhofaily, A. Elyassih, and A. R. Khan, “Federated learningbased reference evapotranspiration estimation for distributed crop fields,” *PloS One*, vol. 20, no. 2, p. e0314921, 2025.

[10] Z. M. Hendy, M. A. Abdelhamid, Y. Gyasi-Agyei, and A. Mokhtar, “Estimation of reference evapotranspiration based on ML models and timeseries analysis: a case study in an arid climate,” *Appl. Water Sci.*, vol. 13, no. 11, p. 216, 2023.

[11] V. S. Reddy, S. Harivardhagini, and G. Sreelakshmi, “IoT and cloud based sustainable smart irrigation system,” in *E3S Web of Conferences*, EDP Sciences, 2024, p. 01026.

[12] K. Taueatsoala, C. Daniels, A. J. Ramsunar, P. Bronkhorst, and A. E. Ezugwu, “TinyML-Enabled IoT for Sustainable Precision Irrigation,” *ArXiv Prepr. ArXiv260113054*, 2026.

[13] U. R. BP, K. N. A. Sattar, and A. A. Elngar, “A smart irrigation system using the IoT and advanced ML model,” *J. Smart Internet Things*, vol. 2024, no. 2, pp. 13–25, 2024.

[14] A. Ali, “ML approaches for soil moisture prediction: enhancing agricultural water management with integrated data,” *Model. Earth Syst. Environ.*, vol. 12, no. 1, pp. 1–22, 2026.

[15] A. K. Roy, K. K. Jena, and D. Mohapatra, “Water Pumping Requirement Prediction in Irrigation System Using Internet of Things-Assisted Machine Intelligence-Based Approach,” in *International Conference on Advances in Computational Intelligence and Informatics*, Springer, 2023, pp. 131–138.

[16] “Irrigation Water Requirement Prediction Dataset.” Accessed: Jul. 02, 2026. [Online]. Available: https://www.kaggle.com/datasets/miadul/irrigation-water-requirement-prediction-dataset

[17] A. Elmotawakkil *et al.*, “HydroPredictor a hybrid ML model for addressing data scarcity in groundwater prediction,” *Sci. Rep.*, 2025.

[18] K. Halder *et al.*, “Application of bagging and boosting ensemble ML techniques for groundwater potential mapping in a drought-prone agriculture region of eastern India,” *Environ. Sci. Eur.*, vol. 36, no. 1, p. 155, 2024.