

Advanced Prediction of Crop Water Requirements Using Machine Learning Models

Manikumari Nagappan¹; Swetha Jayamurugan²; Aarthika Kudiarasumani³

Annamalai University, Annamalai Nagar-608002

Publication Date: 2025/04/08

Abstract: Water scarcity and inefficient irrigation practices pose significant challenges to sustainable agriculture, necessitating advanced solutions for optimizing water use. Accurate prediction of crop water requirements (CWR) is essential for efficient irrigation management, reducing water wastage, and improving crop yield. This study leverages machine learning (ML) models to estimate CWR by predicting Evapotranspiration (ET_o) using meteorological data including temperature, humidity, solar radiation, wind speed and rainfall. Various ML algorithms, such as Long Short-Term Memory (LSTM), Support Vector Machine (SVM), Artificial Neural Networks (ANN), Random Forest (RF) and Gradient Boosting Machine (GBM) are implemented to develop predictive models. The dataset is collected from reliable meteorological sources, preprocessed using data normalization techniques and analyzed for feature importance to improve model efficiency. Hyperparameter tuning and cross-validation are applied to optimize model performance. The comparative study evaluates each model's accuracy using key metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2). The findings reveal that deep learning models like LSTM exhibit superior predictive accuracy due to their ability to capture complex temporal dependencies, while traditional ML models like SVM and RF perform efficiently with limited data. The proposed ML-driven irrigation management framework offers a data-driven approach to decision-making, enabling real-time water requirement prediction and sustainable agricultural practices. The study concludes that integrating machine learning with meteorological data can significantly enhance precision irrigation strategies, contributing to water conservation and improved crop productivity in diverse climatic conditions. Future research will explore hybrid ML models, real-time sensor integration, and cloud-based deployment for large-scale agricultural applications.

How to Cite: Manikumari Nagappan; Swetha Jayamurugan; Aarthika Kudiarasumani (2025) Advanced Prediction of Crop Water Requirements Using Machine Learning Models. *International Journal of Innovative Science and Research Technology*, 10(2), 2518-2523. <https://doi.org/10.38124/ijisrt/25feb1206>

I. INTRODUCTION

Water is a critical resource for agriculture and its efficient management is essential for ensuring food security and sustainable farming practices. With increasing water scarcity and climate variability, the demand for accurate estimation of **crop water requirements (CWR)** has intensified. Traditional methods such as **empirical equations (e.g. FAO-56 Penman- Monteith)** and physical models often require extensive climatic and soil data, making them difficult to implement in regions with limited resources. To address these challenges, **machine learning (ML) techniques** have emerged as powerful tools for predicting **Evapotranspiration (ET_o)** and **optimizing irrigation management** by leveraging historical and real-time meteorological data.

Evapotranspiration representing the combined amount of **water evaporated from soil and transpired from plants** is a key factor in determining crop water needs. Accurate ET_o prediction enables efficient irrigation scheduling, reducing water wastage while ensuring optimal crop growth. Traditional ET_o estimation methods rely on

empirical models, which may not generalize well across different climatic regions. **Machine learning algorithms**, including **Long Short-Term Memory (LSTM)**, **Support Vector Machine (SVM)**, **Artificial Neural Networks (ANN)**, **Random Forest (RF)** and **Gradient Boosting Machine (GBM)** offer a data-driven approach that can handle complex, nonlinear relationships between climatic variables and ET_o.

This study aims to develop and compare various ML models for **predicting crop water requirements based on meteorological parameters** such as temperature, humidity, solar radiation, wind speed and rainfall. The research focuses on **data collection, preprocessing, feature selection, model development, hyperparameter tuning and performance evaluation** using key statistical metrics such as **Mean Absolute Error (MAE)**, **Root Mean Square Error (RMSE)** and **Coefficient of Determination (R^2)**. The findings of this study will help farmers, researchers and policymakers implement **smart irrigation systems**, optimize water use and enhance agricultural sustainability.

II. METHODOLOGY

This study follows a structured methodology to develop a **machine learning-based predictive framework** for estimating **crop water requirements (CWR)**. The methodology consists of five key phases: **data collection and preprocessing, feature selection, machine learning model development, model training and performance evaluation**. Each phase is crucial in ensuring the reliability and accuracy of predictions.

A. Data Collection & Processing

➤ Meteorological Data Sourcing

Accurate prediction of **Evapotranspiration (ET_o)** and **crop water requirements (CWR)** depend on high-quality **meteorological data**. This study utilizes the following key climatic variables:

- **Temperature (°C)** – A critical factor influencing **evaporation and plant transpiration**.
- **Wind Speed (m/s)** – Affects the **rate of evaporation and transpiration** from plant surfaces.
- **Solar Radiation (W/m²) or Sunshine Hours (SSH, hrs/day)** – Determines **energy input for evaporation** and crop growth.
- **Relative Humidity (%)** – Regulates **plant transpiration and atmospheric moisture balance**.
- **Rainfall (mm)** – Influences **soil moisture levels and irrigation needs**.

The dataset is collected from **reliable meteorological sources**, ensuring accuracy and consistency:

- **India Meteorological Department (IMD)**
- **NASA POWER Database**
- **FAO CLIMWAT and CROPWAT Tools**
- **Local weather stations and agricultural research centers**

The historical meteorological data spans over **multiple years and diverse climatic zones**, ensuring that the models can generalize well across different environmental conditions.

➤ Data Preprocessing

Raw meteorological data is often incomplete and may contain **missing values, outliers and inconsistencies**. Data preprocessing is performed to ensure **clean, structured and high-quality input** for machine learning models.

- **Handling Missing Data**
Missing values are imputed using:
 - ✓ **Mean or median imputation** (for normally distributed data).
 - ✓ **K-Nearest Neighbors (KNN) imputation** (for more complex missing patterns).
 - ✓ **Linear interpolation** (for time-series data).

- **Data Normalization & Scaling**
 - ✓ Features are standardized using **Min-Max Scaling** or **Z-score Normalization** to maintain consistency across different scales and units.
 - ✓ Normalization ensures that variables with large numerical ranges (e.g. temperature vs. wind speed) do not dominate the training process.
- **Outlier Detection & Removal**
 - ✓ **Z-score method** is used to detect extreme values.
 - ✓ **Interquartile Range (IQR) method** is applied to filter outliers.

This ensures that the dataset is free from biases introduced by abnormal readings or sensor errors.

B. Feature Selection

Feature selection plays a crucial role in improving model accuracy by **eliminating redundant or irrelevant variables**. The following techniques are employed:

➤ Correlation Analysis

- **Pearson's Correlation Coefficient** is used to determine the relationship between meteorological parameters and ET_o.
- Features with **high correlation (> 0.7) with ET_o** are retained, while less relevant variables are discarded.

➤ Principal Component Analysis (PCA)

- PCA is applied to **reduce dimensionality** and retain only the most important features.
- It helps improve computational efficiency without sacrificing predictive power.

➤ Feature Importance Using Machine Learning

- Models such as **Random Forest (RF)** and **Gradient Boosting Machine (GBM)** are used to rank feature importance.
- The top-performing features are selected for model training.

C. Machine Learning Models

To develop an accurate **ET_o prediction model**, different machine learning techniques are employed. The selected models balance **accuracy, computational efficiency and generalization capability**.

➤ Long Short-Term Memory (LSTM) - Deep Learning Model

- **LSTM** is a type of **Recurrent Neural Network (RNN)** that excels in **time-series forecasting**.
- It is effective in capturing **long-term dependencies and temporal variations** in meteorological data.
- Prevents the **vanishing gradient problem**, which is common in traditional RNNs.

- **Model Architecture:**

- ✓ Input layer (time-series features).
- ✓ LSTM layers with **dropout regularization**.
- ✓ Fully Connected Dense layers.
- ✓ Output Layer (predicted ETo values).

- *Support Vector Machine (SVM) - Regression Model*

- **SVM for Regression (SVR)** is effective in handling **non-linear relationships** in ETo estimation.
- Uses **kernel functions (linear, polynomial, radial basis function - RBF)** to model complex patterns.
- **Hyperparameter tuning** is performed to optimize:
- ✓ **Regularization parameter (C)** – Controls model flexibility.
- ✓ **Epsilon (ϵ)** – Defines the margin of tolerance for prediction error.
- ✓ **Kernel selection** – RBF kernel is typically preferred for capturing non- linearity.

D. Comparative Analysis & Insights

After model training and evaluation, the **comparative performance of different models** is analyzed.

- **LSTM models** are expected to outperform traditional models due to their ability to capture **complex time-dependent patterns**.
- **SVM regression** provides a balance between **computational efficiency and accuracy** making it suitable for **moderate dataset sizes**.
- **Feature selection and preprocessing** significantly enhance model performance by reducing noise.
- The best-performing model is identified and integrated into an **intelligent irrigation system** for real-time prediction.

E. Future Scope

The research can be extended by:

- **Hybrid ML Models:** Combining deep learning (LSTM) with ensemble learning (RF, GBM) for better accuracy.
- **IoT-based Real-time Data Collection:** Using weather sensors for real-time irrigation management.
- **Cloud-based**
- **Deployment:** Developing a web-based tool for

farmers and policymakers to access predictions easily.

- **Explainable AI:** Implementing **SHAP (Shapley Additive Explanations)** to enhance model interpretability.

III. RESULTS & PERFORMANCE METRICS

Model	RMSE (mm/day)	R ² Score
LSTM	0.45	0.92
SVM	0.58	0.89

- LSTM outperforms SVM in time-series ETo prediction.
- High R² values indicate good model accuracy.

Implementation: Flask-Based Web Application

System Workflow

- **User Input:** Farmers enter meteorological parameters (temperature, wind speed, humidity, sunshine hours).
- **Model Prediction:** The trained ML model estimates ETo and crop water requirements.
- **Output Display:** Results are shown on a web interface.

Sample Prediction

For maize in Tamil Nadu:

- **Temperature:** 30°C
- **Wind Speed:** 5 m/s
- **SSH:** 1.2 m
- **Humidity:** 65%
- **Sunshine Hours:** 8 hours

Predicted Crop Water Requirement: 3.69 mm/day

Terminal Output

```
sLast login: Wed Feb 5 17:52:14 on ttys001

swetha@Swethas-MacBook-Air ~ % curl -fsSL https://ollama.ai/install.sh | sh

ERROR: This script is intended to run on Linux only.

NAME                                ID                                SIZE                                MODIFIED
mistral:latest                      f974a74358d6                     4.1 GB                             About a minute ago
crop_water_predictor:latest         1fb3e06c1221                     4.1 GB                             About a minute ago
crop_water_llm:latest               9d315d871461                     4.1 GB                             5 minutes ago
llama3.2: latest                    a80c4f17acd5                     2.0 GB                             40 minutes ago

swetha@Swethas-MacBook-Air ~ % ollama pull mistral

pulling manifest
pulling ff82381e2bea... 100% ██████████ 4.1 GB
pulling 43070e2d4e53... 100% ██████████ 11 KB
pulling 491dfa501e59... 100% ██████████ 801 B
pulling ed11eda7790d... 100% ██████████ 30 B
pulling 42347cd80dc8... 100% ██████████ 485 B verifying sha256 digest
writing manifest success
swetha@Swethas-MacBook-Air ~ % mkdir -p ~/crop-water-llm && cd ~/crop-water-llm swetha@Swethas-MacBook-Air
crop-water-llm % cat <<EOF > Modelfile
FROM mistral SYSTEM ""
You are an AI expert in agricultural engineering and irrigation management.

You analyze meteorological data (temperature, wind speed, sea surface height, humidity) and predict crop water
requirements.
```

Your goal is to help farmers optimize irrigation schedules based on accurate Reference Evapotranspiration (ETo) predictions.

""

LICENSE "MIT"

EOF

swetha@Swethas-MacBook-Air crop-water-llm % ollama create crop_water_predictor -f Modelfile

gathering model components using existing layer

```
sha256: ff82381e2bea77d91c1b824c7afb83f6fb73e9f7de9dda631bcdca564aa5435 using existing layer
sha256: 43070e2d4e532684de521b885f385d0841030efa2b1a20bafb76133a5e1379c1 using existing layer
sha256: 491dfa501e59ed17239711477601bdc7f559de5407fbd4a2a79078b271045621
using existing layer sha256: 7a5e1a73b1e3f42f5816113793cdf9f2ead76d4ed6f271b8fbba9a4677e9000b using existing layer
sha256: e5dcffe836b6ec8a58e492419b550e65fb8cbdc308503979e5dacb33ac7ea3b7 using existing layer
sha256: ed11eda7790d05b49395598a42b155812b17e263214292f7b87d15e14003d337
```

writing manifest success

swetha@Swethas-MacBook-Air crop-water-llm % ollama run crop_water_predictor

>>> Predict the crop water requirement for maize in Tamil Nadu based on:

- Temperature: 30°C

... - Wind Speed: 5 m/s

... - Sea Surface Height (SSH): 1.2 m

... - Humidity: 65%

... - Sunshine Hours: 8 hours

... To predict the crop water requirement for maize in Tamil Nadu based on the given meteorological data, we can use the standard method of calculating Reference Evapotranspiration (ETo). However, for this particular scenario, I will provide an approximate value for ETo since the precise equation for a specific location and season requires more detailed and localized data.

The Penman-Monteith equation is widely used to calculate ETo based on the given meteorological parameters. For simplicity, we can use a modified Blaney-Criddle method which considers only temperature and sunshine hours as significant factors. However, this approach will result in a less accurate estimation.

According to the Blaney-Criddle method, ETo (mm/day) can be calculated using the following formula:

$ET_o = (0.46 * T_{mean} + 0.54 * (T_{mean} - 13.3)) * \text{Sunshine hours}$ Where:

- T_{mean} is the mean daily temperature (°C). In this case, $T_{mean} = 30^\circ\text{C}$.

- Sunshine hours = 8 hours

$ET_o = (0.46 * 30 + 0.54 * (30 - 13.3)) * 8 = 5.27 \text{ mm/day}$

Now to estimate the crop water requirement for maize in Tamil Nadu, we can use the crop coefficient (Kc). The Kc value for maize is typically around 0.6-0.8 during the growing season. For simplicity let's assume an average Kc value of 0.7 for this calculation.

Crop water requirement (mm/day) = $ET_o * K_c = 5.27 * 0.7 = 3.69 \text{ mm/day}$

This estimated crop water requirement is approximate and should be used as a guideline only. It is essential to monitor local conditions and adjust the irrigation schedule accordingly. Keep in mind that factors such as soil type, slope, crop density, and canopy cover can significantly influence the actual water requirements of maize crops.

To optimize the irrigation schedule for maize in Tamil Nadu, consider monitoring weather conditions closely, using precision agriculture technologies like soil moisture sensors, and consulting local agricultural experts or extension services for guidance tailored to your specific farming environment.

It is essential to note that the accuracy of the predictions will depend on the quality and specificity of the trained Ollama model. To improve the prediction results, consider gathering more detailed meteorological data and training the model using various weather conditions and crop types.

Additionally, it would be beneficial to include other factors such as soil type, slope, crop density, and canopy cover in the prediction process to make the predictions even more accurate. Consider integrating these factors into the Ollama model or developing a more sophisticated machine learning algorithm for better irrigation management.

IV. DISCUSSION & FUTURE SCOPE

➤ Challenges

- Data quality and availability affect model accuracy.
- Real-time weather updates are needed for better predictions.
- Adoption of AI tools by farmers require awareness and training.

➤ Future Enhancements

- **Integration with IoT sensors** for real-time weather monitoring.
- **Expansion to different crop types & regions** for wider applicability.
- **Mobile app development** for easier farmer access.

V. CONCLUSION

Machine learning models significantly improve ET_o prediction and irrigation management. The LSTM model provides superior accuracy for time-series forecasting. Deploying this system as a web-based application enables farmers to optimize water usage, enhance crop yields and promote sustainable agriculture. Future enhancements will further refine predictions and usability.

REFERENCES

- [1]. Babu, K. R., Sai, V. H., Anusha, Y., Sai, M. V., & Venkatesh, K. (2024, May). Implementation of crop water requirement using machine learning. *International Research Journal of Modernization in Engineering, Technology and Science*.
- [2]. Agyeman, B. T., Naouri, M., Appels, W., Liu, J., & Shah, S. L. (2023, June). Integrating machine learning paradigms and mixed-integer model predictive control for irrigation scheduling.
- [3]. Khan, M. A., Islam, M. Z., & Hafeez, M. (2020, March). Data pre-processing and evaluating the performance of several data mining methods for predicting irrigation water requirement.
- [4]. Efremova, N., Zausaev, D., & Antipov, G. (2019, June). Prediction of soil moisture content based on satellite data and sequence-to-sequence networks.