Machine learning-based water stress detection using wireless sensor network data

Abdelaziz El Yazidi[1], *Abdallah* Rhattoy[1], and *Moulay Lahcen* Hasnaoui[2]

[1]ISIC-TEAM, L2ISEI Laboratory – ESTM, Moulay Ismail University, Meknes, Morocco
[2]ACTE-TEAM, EDS Laboratory – FSE, Mohammed V University of Rabat, Morocco

Abstract. Smart agriculture is increasingly recognized as a key solution for improving agricultural productivity while reducing environmental impact and water consumption. The integration of Internet of Things (IoT) technologies and Wireless Sensor Networks (WSNs) enables continuous monitoring of environmental conditions in agricultural fields. These sensing infrastructures generate large volumes of environmental data that can be exploited using Machine Learning techniques in order to support intelligent decision-making.

This study investigates the application of supervised Machine Learning algorithms for the detection of soil water stress using real environmental data. The objective is to identify the most suitable algorithm for integration into intelligent irrigation systems deployed in smart agriculture environments.

A real dataset collected at the Yanco Agricultural Research Institute in Australia was used in this study. The dataset contains 6,623 time-stamped observations recorded between June and August 2018. The measurements include soil moisture and soil temperature values captured at a depth of 0–5 cm by sensors deployed in a Wireless Sensor Network.

Three Machine Learning algorithms were evaluated and compared: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression (LR). These models were selected because they represent lightweight and computationally efficient methods suitable for integration in IoT-based agricultural systems.

Data preprocessing included cleaning erroneous measurements, converting decimal values, and defining a stress threshold derived from the 25th percentile of soil moisture values combined with temperature conditions. The dataset was divided chronologically into training and testing sets using a 70%–30% split. A Time-Series-Split validation strategy was adopted in order to preserve the temporal characteristics of the data.

The performance of the models was evaluated using several classification metrics including accuracy, precision, recall, F1-score, Area Under the ROC Curve (AUC), and confusion matrices.

Experimental results show that the KNN model significantly outperforms the other methods, achieving an accuracy of 99.6%, precision of 99.0%, recall of 98.4%, and an AUC value of 0.998. The SVM model obtained a perfect recall score but produced a very high number of false positive predictions, which may lead to unnecessary irrigation events. Logistic Regression demonstrated moderate accuracy but failed to detect most water stress events due to its very low recall.
The results demonstrate that KNN provides the best balance between detection reliability and false alarm reduction. This makes it particularly suitable for integration in IoT-based irrigation management systems where both precision and energy efficiency are critical.
The findings of this research contribute to the development of data-driven decision support systems for sustainable agriculture. Future work will focus on integrating additional environmental variables, expanding the dataset, and deploying the model on edge computing platforms for real-time irrigation control.

Keywords: Smart Agriculture, Wireless Sensor Networks (WSN), Machine Learning, Water Stress Detection, KNearest Neighbors (KNN).

1 Introduction

Agriculture is facing increasing pressure due to climate change, population growth, and the limited availability of water resources. Efficient water management has therefore become a critical challenge for modern farming systems. Smart agriculture has emerged as a promising approach that combines digital technologies, sensor networks, and artificial intelligence to optimize agricultural processes [1, 13].

Wireless Sensor Networks (WSNs) enable the deployment of distributed sensors in agricultural fields to continuously monitor environmental parameters such as soil moisture, temperature, humidity, and solar radiation. These sensors generate high-frequency data streams that provide valuable information about crop conditions and environmental dynamics.

However, the raw data collected by sensors cannot directly support decision-making. Advanced analytical methods are required to transform this data into actionable information. Machine Learning techniques have demonstrated significant potential for extracting patterns and relationships from environmental data and supporting predictive agricultural systems [2].

One important application of Machine Learning in agriculture is the detection of water stress in crops. Water stress occurs when plants receive insufficient water to sustain optimal growth. Early detection of such stress conditions allows farmers or automated systems to apply irrigation more efficiently and prevent crop damage [13].

Although deep learning models have achieved remarkable results in many domains, their computational complexity often limits their applicability in resource-constrained environments such as embedded agricultural sensor systems. Consequently, lightweight Machine Learning algorithms remain highly relevant for real-time decision support in IoT-based agricultural infrastructures [2, 6].

This research aims to evaluate the performance of three classical supervised Machine Learning models—K-Nearest Neighbors, Support Vector Machine, and Logistic Regression—for the detection of soil water stress using real data collected from agricultural sensors. The main objective is to identify the model that provides the best balance between accuracy, reliability, and computational simplicity for integration into smart irrigation systems.

2 Related Work

Machine Learning has increasingly been applied to agricultural data analysis in order to improve crop management and resource optimization [6, 15]. Previous studies have explored various algorithms for tasks such as irrigation scheduling, crop yield prediction, and disease detection. In particular, classification algorithms have been widely used to identify environmental stress conditions affecting crops [12, 15].

Several researchers have demonstrated the effectiveness of Machine Learning models for predicting soil moisture and irrigation needs using sensor data. Algorithms such as Random Forest, Support Vector Machine, and Neural Networks have been applied to agricultural datasets with promising results [7, 8, 11].

Despite the success of complex models, lightweight algorithms remain particularly attractive for deployment in IoT-based agricultural systems. These systems often operate with limited computational power and energy resources, making efficient algorithms essential [3, 13]. Therefore, comparative evaluations of classical Machine Learning models remain valuable for determining the most suitable techniques for smart agriculture applications [4, 15].

3 Dataset and Data Preparation

3.1 Data Source

The dataset used in this study was collected at the Yanco Agricultural Research Institute in Australia. The measurements were obtained using sensors deployed in a Wireless Sensor Network installed in experimental agricultural plots. The dataset contains 6,623 observations recorded between June 1 and August 31, 2018, with measurements taken approximately every 20 minutes. Two main environmental variables were considered: Soil Moisture (SM) and Soil Temperature, both at a depth of 0–5 cm. These parameters are widely recognized as key indicators of plant water status.

3.2 Data Cleaning

Several preprocessing operations were performed to improve data quality: conversion of decimal commas to decimal points, removal of erroneous sensor values such as −99°C, and verification of time consistency in the dataset. These steps ensured that the dataset was suitable for Machine Learning analysis.

3.3 Stress Label Definition

To transform the problem into a classification task, a binary stress label was defined based on soil moisture and temperature values. A stress condition was defined as: Stress = 1 if soil moisture < threshold and temperature > threshold. The threshold values were derived from the statistical distribution of the dataset, particularly the 25th percentile of soil moisture measurements.

4 Methodology

4.1 Logistic Regression

Logistic Regression is a linear classification model that estimates the probability of an observation belonging to a particular class. Due to its simplicity and interpretability, it is often used as a baseline model [14].

4.2 K-Nearest Neighbors

K-Nearest Neighbors is a non-parametric algorithm that classifies observations based on the majority class among their nearest neighbors in the feature space. This method can capture non-linear relationships between variables [9, 14].

4.3 Support Vector Machine

Support Vector Machine aims to determine an optimal decision boundary that maximizes the margin between different classes. Kernel functions allow the model to handle non-linear separations between classes [8, 14].

4.4 Training and Validation

The dataset was divided chronologically: 70% for model training and 30% for testing. A TimeSeriesSplit validation approach was used to preserve the temporal order of the data. Model performance was evaluated using Accuracy, Precision, Recall, F1-Score, and Area Under the ROC Curve (AUC).

5 Experimental Results

After evaluating the three models using Python in the Google Colab environment, with Matplotlib used for data visualization, the experimental evaluation produced the results shown in Table 1. The ROC curves of the evaluated models are illustrated in Fig. 1.

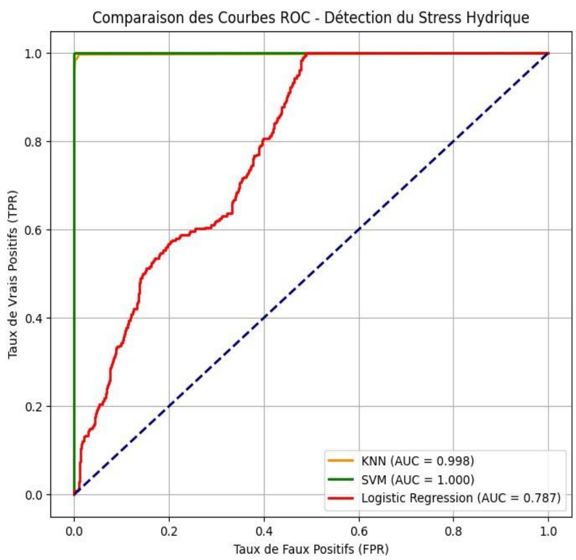

Fig. 1. ROC curves comparison for KNN, SVM and Logistic Regression models for water stress detection.

Table 1. Performance comparison of KNN, SVM, and Logistic Regression classification models.

Model	Accuracy	Precision	Recall	F1-Score	AUC
KNN	0.996	0.990	0.984	0.987	0.998
SVM	0.517	0.246	1.000	0.395	0.786
LR	0.844	0.548	0.073	0.129	0.787

6 Discussion

The results demonstrate that KNN provides the most reliable predictions for water stress detection. The model achieves an excellent balance between precision and recall, ensuring that stress conditions are detected accurately without generating excessive false alarms [4, 11].

In contrast, the SVM model, despite detecting all stress cases, produced a very large number of false positives. Such behavior could lead to unnecessary irrigation actions in automated systems. Logistic Regression performed moderately in terms of overall accuracy but failed to detect most stress cases, which could result in serious agricultural losses.

These findings highlight the importance of selecting appropriate Machine Learning models when designing decision support systems for smart irrigation [3, 15].

7 Conclusion and Future Work

This study presented a comparative analysis of three Machine Learning algorithms for soil water stress detection using real sensor data from a Wireless Sensor Network deployed in an agricultural environment. The experimental results show that the K-Nearest Neighbors algorithm significantly outperforms Support Vector Machine and Logistic Regression in terms of classification accuracy and reliability.

Due to its high performance and simplicity, KNN appears to be particularly suitable for integration into IoT-based irrigation management systems [9, 13]. Future research directions include: integrating additional environmental variables such as humidity and rainfall; extending the dataset to multiple seasons; deploying the model on edge computing platforms; and combining sensor data with satellite imagery. These developments will contribute to the advancement of intelligent and sustainable agricultural systems [2, 15].

8 References

1. Md.N. Mowla, N. Mowla, A.F.M.S. Shah, K.M. Rabie, T. Shongwe, Internet of Things and Wireless Sensor Networks for Smart Agriculture Applications: A Survey, IEEE Access **11**, 145813–145852 (2023)
2. A. Kamilaris, F.X. Prenafeta-Boldu, Comput. Electron. Agric. **147**, 70–90 (2018)
3. G.S. Prasanna Lakshmi, P.N. Asha, G. Sandhya, S. Vivek Sharma, S. Shilpashree, S.G. Subramanya, An intelligent IoT sensor coupled precision irrigation model for agriculture, Meas. Sens. **25**, 100608 (2023)
4. M.A. Osama, A.A. El-Tawil, A.S. Saad, Soil moisture prediction using machine learning for smart irrigation, in Proc. IEEE ITCE, Aswan, Egypt (2019), pp. 432–436
5. C. Zhang et al., Remote Sens. Environ. **221**, 173–187 (2019)

6. S. Wolfert, L. Ge, C. Verdouw, M.-J. Bogaardt, Agric. Syst. **153**, 69–80 (2017)
7. Y. Liang, S. Kou, Y. Liu, W. Liu, H. Zhou, Soil Moisture Prediction Using Remote Sensing and Machine Learning Algorithms: A Review on Progress, Challenges, and Opportunities, Remote Sens. **17**(14), 2397 (2025)
8. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing **408**, 189–215 (2020)
9. P.K. Syriopoulos, N.G. Kalampalikis, S.B. Kotsiantis, M.N. Vrahatis, kNN Classification: a review, Ann. Math. Artif. Intell. **91**, 1547–1581 (2023)
10. J.N. Ndunagu, O.C. Onuoha, R.U. Ndubuisi, J.B. Ogar, T.B. Eze, C.N. Owuama, Development of a Wireless Sensor Network and IoT-based Smart Irrigation System, Appl. Environ. Soil Sci. **2022**, 7678570 (2022)
11. S. Ju et al., IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. **10**(2), 573–581 (2017)
12. S. Uddin, I. Haque, H. Lu, M.A. Moni, E. Gide, Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction, Sci. Rep. **12**, 6256 (2022)
13. H. Ullah, M. Gohar, A. Iqbal, M. Yasin, H. Mahmood, G.A. Shah, Integration of smart sensors and IoT in precision agriculture: trends, challenges and future prospectives, Front. Plant Sci. **16**, 1587869 (2025)
14. L. Buitinck et al., API design for machine learning software: experiences from the scikit-learn project, ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pp. 108–122 (2013)
15. 15. A. Sharma, E. Jain, P. Gupta, V. Chowdary, IEEE Access **9**, 4843–4873 (2021)