International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, Delaware, USA
Volume 12– No. 47, June 2025 – www.ijais.org

Comparative Analysis of Machine Learning Models for Irrigation Technique Classification in Precision Agriculture

Saviour Inyang
Department of Computing, Topfaith University, Mkpatak, Nigeria

Daniel Essien
Department of Computer Science, Dalhousie University, Halifax, Canada

Ezichi Ndudirim
Department of Computer, Abia State University

ABSTRACT
Efficient water management in agriculture is crucial for sustainable food production, especially in regions facing water scarcity and climate variability. This research investigates the application of machine learning (ML) techniques to classify irrigation methods which includes: Overhead, Surface, and Precision Irrigation based on relevant agricultural data. The study evaluates the performance of four widely used ML models: Random Forest (RF), Support Vector Machine (SVM), XGBoost, and K-Nearest Neighbors (KNN), with the aim of identifying the most suitable model for accurate and consistent classification. Each model was trained and tested using labeled datasets and assessed through performance metrics including Precision, Recall, F1-Score, Accuracy, and Cohen's Kappa. Confusion matrices and ROC-AUC curves were also utilized to visualize class-specific performance. The results indicate that XGBoost outperformed all other models, achieving the highest classification accuracy (86%) and a Kappa score of 0.79. It demonstrated superior performance across all irrigation classes, particularly excelling in the Precision Irrigation category. Random Forest followed closely, with an accuracy of 80% and Kappa of 0.70, notably achieving a perfect precision score for Overhead Irrigation. SVM delivered moderate performance with 72% accuracy and a Kappa of 0.57, while KNN lagged behind, scoring 64% accuracy and 0.46 Kappa. The comparative analysis highlights the effectiveness of ensemble-based methods, particularly XGBoost, for handling diverse and potentially non-linear agricultural datasets. The findings support the integration of advanced ML models in agricultural decision support systems, enabling more precise irrigation management and optimizing water resource utilization.

Keywords
Irrigation Techniques, Machine Learning, Agricultural Data Analysis, Classification Models

1. INTRODUCTION
Irrigation is crucial in the shaping of agricultural sector today. There have been many advancements in irrigation techniques from traditional techniques like surface, furrow, and flood irrigation to more modern approaches like sprinklers, drip, and subsurface irrigation. Recently, there has been an influx of automated smart irrigation systems that give more precision by analyzing soil moisture content or environmental variables, e.g., humidity, radiation levels, etc. A review in [1], measured the impact of applied machine learning techniques on solving certain irrigation issues and how productivity was directly affected. The study analyzed techniques related to soil and water management. This showed the recent attempts at utilizing irrigation techniques to improve water usage. Also, the importance of frequent modernization of irrigation techniques to address issues of sustainability in agriculture shows that traditional methods like surface irrigation have low water use efficiency (30% - 40%), while modern methods such as drip and sprinkler irrigation achieve efficiencies up to 90-95%. Advanced practices like precision land leveling and subsurface irrigation significantly improved water distribution, crop yield, and resource management [2]. These findings are proof that innovation systems optimize water usage, thereby ensuring more sustainable agricultural practices. The role of data is crucial when migrating to the newer smart irrigation systems because these give room for more precise use of the water resources. Smart Irrigation Systems make decisions based on data such as soil moisture, available water resources, crop type, or crop health. A perfect application of this is shown in [3]. The study introduced a smart irrigation system called IoTML-SIS, which was used to achieve effective water usage and automated irrigation. This model made decisions based on the environmental conditions of the farm. It involved different IoT-based sensors to track soil moisture, humidity, temperature, and light. The data was then transferred to the cloud for processing and decision-making. The classification made use of the artificial algae algorithm (AAA) and least squares-support vector machine (LS-SVM). This resulted in 97% accuracy, even higher compared to other ML models. Nevertheless, machine learning models are typically divided into three subgroups: Supervised, Unsupervised, and Reinforcement Learning. Many studies have incorporated these three main types of machine learning models. Each type is used to solve different problems. A study by [4] investigated the potential of ML models in predicting the effect of evapotranspiration (ETo), which is very crucial in efficient irrigation water management, this evapotranspiration is simply the sum of all processes by which water moves from the land to the atmosphere. They study compared traditional methods like Penman-Monteith, Hargreaves, and Blaney-Criddle with advanced ML models such as Random Forest (RF), Support Vector Regression (SVR), XGBoost, and Decision Trees, highlighting their predictive accuracy and efficiency. Using 38 years of climate data from Egypt, the study highlighted significant climate trends e.g., changes in solar radiation, and rising temperatures. The ML models gave high accuracy with the R2 (R squared) values ranging from 0.91 to 0.99. This study was impactful when developing adaptive strategies to mitigate climate change's impact on water resources. Again, the

possibilities are endless when incorporating machine learning with irrigation. [5] used machine learning to estimate water use in China at a national scale. They study utilized satellite remote sensing, meteorological data, and economic statistics to address the limitations of traditional irrigation techniques. Key findings in their study indicate that Irrigation Water Use (IWU) is expected to increase significantly under higher emission scenarios. This calls for better water practices not only in China but also in the world. This can be accomplished by optimizing crop selection and adopting efficient irrigation techniques. Further study in [6] has been carried out to know the quality of water used in irrigation systems. This further improves the quality of water which the crops use. The model was used to measure the pH, Total Dissolved Solids (TDS), Electrical Conductivity (EC), and Sodium (Na). The study aimed to analyze the water quality at Ele River Nnewi, Anambra State for irrigation purposes with a view of predicting a one-year water quality index using an Artificial Neural Network (ANN). Farmers and Stakeholders can benefit from this technology. Aside from all the theoretical research, there is a need for a real-life approach for farmers to benefit from this technology. [7] addressed this issue by creating an innovative, smart, low-cost irrigation system that spots anomalies and problems in water usage using Artificial Intelligence. The device was targeted towards smallholder farming communities to enhance irrigation decisions, it also predict important environmental conditions e.g. soil moisture, and humidity, using its field sensory data. This irrigation system consisted of sensor layers, fog layer, and cloud layer. It took the approach of making a "smart irrigation in a box system". The study carried out in [8] highlights several specific advancements in the application of AI, ML, and IoT for precision irrigation. One notable innovation is the application of Artificial Neural Network (ANN) to develop smart irrigation scheduling systems that analyze key environmental parameters such as humidity, radiation, and solar radiation to determine optimal water requirements for each crop. This approach improves water usage and reduces waste. [8] also highlighted several challenges and limitations when implementing AI, ML, and IoT in irrigation systems. One of the major limitations is the high initial cost of deploying these advanced technologies e.g., sensors, IoT devices, and automated control systems. This inflated cost will prevent small-scale farmers from having access to these technologies. Maintenance of these systems in turn becomes more expensive, further increasing the burden. Additionally, there is a significant skill gap between the farmers and AI engineers, as farmers have inadequate technical knowledge to operate these innovative technologies. Contrary, AI Engineers do not have enough farming/agricultural experience to determine what parameters should be a priority. Another difficulty highlighted is the Limited Internet Connectivity and Infrastructure in this remote area. Typically, farmlands are in the outskirts of towns and cities, this challenge will affect data collection by the sensors and the latency will affect the data processing. Nevertheless, to the best of our knowledge, there has not been any study that classifies or recommends which irrigation techniques should be implemented in farms, taking into account the farm area, crop type, fertilizer amount used, etc. This study will bridge this gap by classifying the several types of irrigation techniques a farm should use based on these parameters. This solution will help entry-level farmers make better use of their water resources without waste.

2. LITERATURE REVIEW

2.1 Overview of Machine Learning in Agriculture

Machine Learning has been gaining traction in the field of agriculture, it has been utilized to address challenges like food security, resource management & optimization, and sustainable farming. In crop management, ML Models like Artificial Neural Network (ANN) and SVMs are used to predict yield, detect weeds and diseases, classify species and optimize usage of resources [9]. Meanwhile, in Livestock Farming, Machine Learning has been used to improve their workflow and setup by utilizing wearable sensors and computer vision to monitor animal behaviors, which enables early disease detection. In [10], the author iterated that Adaboost algorithms produce high accuracy in identifying behaviors that include rumination, and locomotor play in calves. These findings and information generally improve the welfare of cows and calves. Furthermore, when ML models are incorporated with ML models, they can be used to optimize water usage and assess soil quality [9] [11]. CNN and LSTM models have been effectively utilized in recognizing nuance and complex behaviors including mounting, feeding, and nursing. Machine Learning can be utilized in a larger and more robust solution, for example, it can analyze multi-modal data like spanning images, accelerometer data, and audio data, to recognize patterns and make predictions. These solutions have been used in Precision Livestock farming where ML models like YOLO and Faster R-CNN are utilized for cattle detection and health monitoring. Studies emphasize advancements in feature extraction such as using visual cues like muzzle prints and coat patterns for livestock identification. In contrast, limitations are still prevalent when researchers and engineers want to deploy ML systems in agriculture. These challenges occur due to the complexity and variation of agricultural environments, imbalance of datasets, and ethical concerns when collecting animal data. However, newer solutions are coming up, for example, adaptive learning techniques and hybrid models. With the field of Machine learning constantly evolving and advancements in sensor technologies, we hope to see more efficient and sustainable agricultural practices [12][13].

2.2 Sustainable Agriculture Goals and Challenges

This study in [14] describes several goals and challenges which concern sustainable agriculture. The main goals consist of achieving environmental sustainability by reducing resource depletion, promoting social equity, and maintaining economic viability for farmers. These goals are set to meet the demands of current and future generations balancing local and global needs. Implementation still faces challenges that are categorized into theoretical, methodological, personal, and practical limitations. Theoretical limitations occur due to the vague definitions of sustainability goals, while methodological challenges relate to insufficient or inconsistent assessment frameworks. Personal issues comprise of knowledge, attitude, and capacity of farmers and practical barriers include resource challenges and societal constraints. A multi-faceted approach is necessary to solve these issues, with support from institutions, stakeholder corporations, etc. A study in [15] addressed the importance of sustainable agriculture and its challenges. The primary goal is balancing nutritional needs while preserving the environmental quality and the economic capability of these agricultural systems. However, there are several challenges which affect these sustainable practices.

Scalability of sustainable practices across various agricultural ecosystems is one major hurdle highlighted in their paper. Economic barriers in the short term, usually discourage farmers from implementing these practices. Typically, the initial cost is the deciding factor for farmers who chose traditional methods over sustainable methods. Their study identifies several important practices including crop rotations, polycultures, pest management, and renewable resource management. These practices can reduce water scarcity, prevent soil degradation, and control climate change but their global adoption of these practices requires a joint effort to overcome societal and technical barriers.

2.3 Machine learning in Irrigation Management

Irrigation has been a very crucial in the field of agriculture, it is simply the artificial application of water to the soil in order to assist growth, revegetation etc. They typically consist of types like Drip, Sprinkler, Smart and Surface irrigation. These methods have been used across times and have been recently evolving. The evolution of these irrigation techniques now incorporates machine learning to further optimize water management and timing. The application of machine learning span across different components from water quality forecasting to scheduled irrigation based on soil water content and many more. In this study [16] , the research was able to utilize machine learning to recommend which irrigation should be used. The machine learning models made use of different types of data like soil, weather, yield and irrigation characteristics. Their study employed models like Gradient Boosted Regression trees and Boosted Tree classifier for this study. This shows how machine learning can also be used in irrigation. Other studies [17] try to mitigate the high cost which have been used in the testing of water quality that is used for irrigation. The study made use of machine learning models like Adaptive Boosting (Adaboost), Random Forest (RF), Artificial Neural Network (ANN), and Support Vector Regression (SVR) to analyze 520 samples of data to recognize the quality of the water. The paper aimed to forecast the Total Dissolved Solid (TDS), Potential Salinity (PS), Sodium Adsorption Ratio (SAR), Exchangeable Sodium Percentage (ESP), Magnesium Adsorption Ratio (MAR), and the Residual Sodium Carbonate (RSC) parameters through Electrical Conductivity (EC), Temperature (T), and pH as inputs. It resulted in Adaboost and Rf obtained the higher performance. Their study still proved successful showing the wide array of solutions that machine learning provides. In order to support year-round farming, solutions need to implemented that can handle the variety of seasons and weather year-round. A study in [18] utilize machine learning techniques which consider factors such as soil moisture, temperature, humidity, and time to aid in smart irrigation. The study used machine learning models like logistic regression, random forest, support vector machine, and convolutional neural network. Random Forest ended up being the best performing with 99.98% accuracy. The need for real time and practical solutions are equally crucial, in order to see the sustainable agriculture that is needed. A study [9] , was conducted in order to create a real time irrigation system which consisted of wireless sensors and actuators network, a mobile application that offers the user the capability of consulting not only the data collected in real time but also their history and also make decisions in accordance with the data it analyses.

3. METHODOLOGY

This section shows the necessary steps and procedures taken for this research. The method steps are chronologically explained in this section, consisting of dataset summary, preprocessing, feature selection, classification process and model evaluation as shown in Figure 1.

3.1 Data Collection

This research made use of secondary data, which was downloaded from the Kaggle Repository [19]. The data set provided a comprehensive view into the agricultural and farming sectors by capturing key variables and trends which are critical for analysis and decision making. The dataset was gotten in Comma Separated Values (CSV) and R was used to extract and view the datasets.

3.2 Data Summary

According to figure 2, the dataset contains 500 entries with 10 total columns. the dataset contains 9 columns which are considered as features, the output is expected to be a categorical value giving the predicted irrigation type.

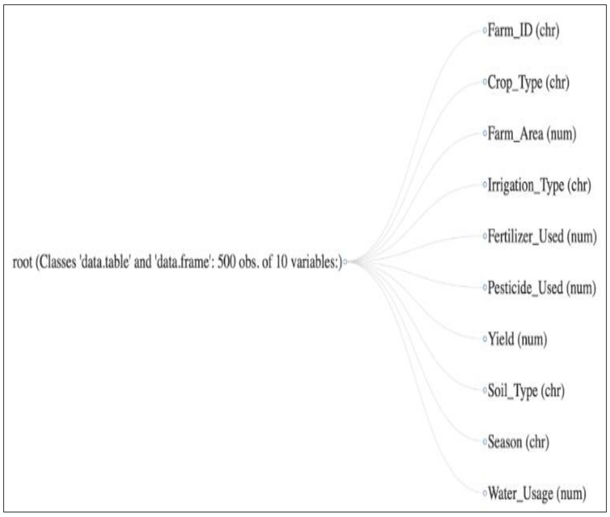

Figure 2: Data Summary

The datasets attributes and descriptive statistical summaries of the observations in the datasets are presented in table 1 and table 2.

Table 1: Dataset Attributes

Column	DataType	Description
Farm_ID	Integer	Unique Identifier for each column
Crop_Type	Factor	Type of crop grown in the farm
Farm_Area(Acres)	Numeric	Size of the farm
Irrigation_Type	Factor	The type of Irrigation used
Fertilizer_Used(tons)	Numeric	The amount of fertilizer used on the farm.
Pesticide_Used(kg)	Numeric	The amount of pesticide used on the farm.
Yield(tons)	Numeric	The amount of yield in tons

Soil_Type	Factor	The soil type used in the farm
Season	Factor	Season
Water_Usage (Cubic Meters)	Numeric	Amount of water used in Cubic Meters

Table 2: Descriptive Statistics of Agricultural Data

Continuous Variable	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Farm_Area (Acres)	12.50	115.3	246.3	242.5	360.5	483.90
Fertilizer_Used(tons)	0.50	2.81	4.96	5.10	7.35	9.96
Pesticide_Used(kg)	0.14	1.30	2.59	2.58	3.77	4.99
Yield(tons)	3.86	15.62	27.62	27.01	38.16	48.02
Water_Usage (Cubic Meters)	5870	27610	49934	50431	73079	94755
Factored Variable	Count					
Crop_Type	10					
Irrigation_Type	5					
Soil_Type	5					
Season	3					

3.3 Exploratory Data Analysis

The bar chart shown Figure 3 shows the distribution of different irrigation techniques across the dataset. It highlights which irrigation techniques were commonly used in the data collection process. By understanding the frequency of each method, we can further understand the data and how to approach it when implementing the machine learning models.

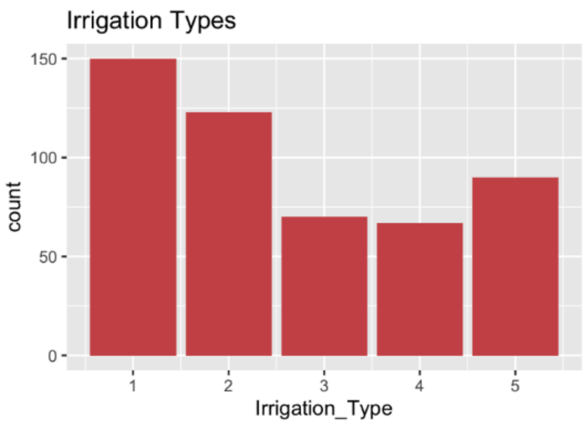

Figure 3: Bar Chart of Irrigation types

The bar chart Figure 4 shows the distribution of crop types, and the irrigation techniques used in the data set. It provides an overview of the crop diversity and helps identify which crops benefited from the individual irrigation type.

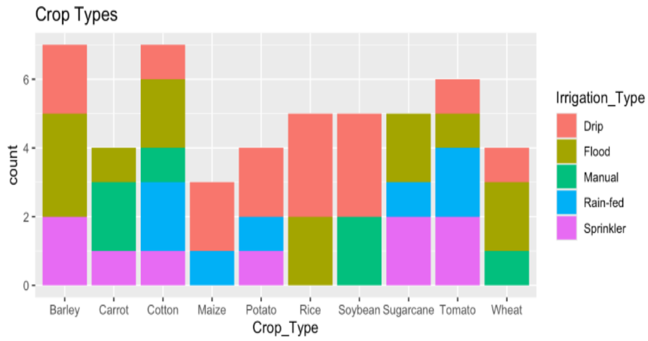

Figure 4: Bar Chart of Crop Types in Relation to Irrigation Types

3.4 Data Preprocessing

The data gotten was already cleaned from the Kaggle repository so minimal steps were taken using min max algorithm for normalization and missing values where handle appropriately. The first step we took was to change the names of each column so that the computer can recognize it better. The names initial had brackets to some columns e.g. “Farm_Size(Area)” was changed to “Farm_Area”. We performed a train and test split for the data, splitting it randomly using 70 to 30 splits.

3.5 Feature Selection

Feature selection plays a very important role in optimizing machine learning models by prioritizing the most relevant variables in the data that will be utilized in the training process. In this study, a couple of features were selected based on their potential influence on the classification of irrigation techniques. These features are Farm_Area (the size of the land), Water_Usage (the amount of water used for irrigation), Fertilizer_Used (quantity of fertilizers applied), and Pesticide_Used (number of pesticides applied). In this study, a couple of features were selected based on their potential influence on the classification of irrigation techniques using Pearson correlation method. These features are Farm_Area (the size of the land), Water_Usage (the amount of water used for irrigation), Fertilizer_Used (quantity of fertilizers applied), and Pesticide_Used (amount of pesticides applied). In addition, categorical variables like Soil_Type, Season, and Crop_Type were included as the factors which could determine irrigation technique decision. Also, based on the imbalance nature of observation distribution, this necessitate the reduction of the irrigation techniques to three classes which are: surface irrigation, precision irrigation and overhead irrigation techniques which are the three classes that will be considered in this study.

3.6 Model Implementation

Three machine learning models were implemented in this study: Artificial Neural Network (ANN), Support Vector Machine (SVM) and Random Forest (RF). The models were independently trained using Caret package in R with 10-fold cross Validation to ensure good performance and avoid overfitting.

3.6.1 Random Forest

Random forest was chosen for this study because of its ability to manage high dimensional data and determine the relative importance of features [20]. This model is built on multiple decision trees and has capabilities to aggregate their predictions to enhance accuracy and reduce variance. The random forest package in the Caret was used to train the model. The hyperparameters were optimized during the training process and the model is evaluated using the confusion matrix. The random forest model combines several trees, with the final prediction gotten from majority vote in classification problems or the average of the individual trees for regression problems. The equation for the Random Forest Model is given by:

$$yRF = \frac{1}{T} \sum_{t=1}^{T} f_t(x) \qquad 1$$

Where T is the total number of trees, $f_t(x)$ is the prediction of the t-th tree for input x and yRF is the final prediction.

3.6.2 Support Vector Machine

This study selected SVM for its robust ways of handling both linear and nonlinear classification tasks. The svmRadial method from the caret package in R was employed for the training of the model with a radial basis function kernel (RBF). The radial basis function (RBF) kernel was employed to map input data into higher-dimensional space. The SVM decision function is represented by

$$f(x) = sgn\left(\sum_{i=1}^{N} \alpha_i \, y_i \, K(x_i, x) + b\right) \qquad 2$$

Where;

x is the input vector,

α_i are the lagrange multipliers,

y_i are the class labels,

$K(x_i, x)$ is the kernel function and

b is the bias term

Parameters e.g. cost (C) and kernel width were tuned to optimize performance. The evaluation metric used are accuracy, precision, confusion matrix and ROC curve.

3.6.3 XGboost

Extreme Gradient Boosting is a model which is an efficient implementation of gradient boosting machines [21]. It is based on the boosting principle where models are sequentially trained, with each new model attempting to correct errors made by the previous one. This model is typically known for speed, scalability and performance.

3.6.4 KNN

K nearest Neighbors (KNN) is an algorithm used for classification tasks. It works by grouping data points based on the majority class of its k nearest neighbors in the feature space. This model is simple and effective yet it can require high computational power for larger datasets [22]. The KNN algorithm calculates the distance between the query points and all the points in the dataset. The widely used distance metric is the Euclidean distance which is given by

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} \left(x_{i,k} - x_{j,k}\right)^2} \qquad 3$$

Where;

$d(x_i, x_j)$ is the Euclidean Distance between xi and xj,

$x_{j,k}$and $x_{i,k}$ are the features of point xi and xj

n is the number of features

3.7 Performance Evaluation Metrics

The performance evaluation metrics used in this study will evaluates the effectiveness of the ML model using the a test set data in this study. The performance metrics used in this research are Accuracy, Recall, Precision, and F-1 score. A confusion matrix was also used to visualize misclassification between cow behaviours.

$$Accuracy = \frac{TP + TN}{total\, number\, of\, predictions} \qquad 4$$

$$Precision = \frac{TP}{TP + FP} \qquad 5$$

$$Recall = \frac{TP}{TP + FN} \qquad 6$$

$$F1 - Score = \frac{precision}{precision + recall} \qquad 7$$

$$kappa = \frac{p_0 - p_e}{1 - p_e} \qquad 8$$

4. RESULTS

This section presents the results evaluation for this study based on each ML models that was utilize in this work.

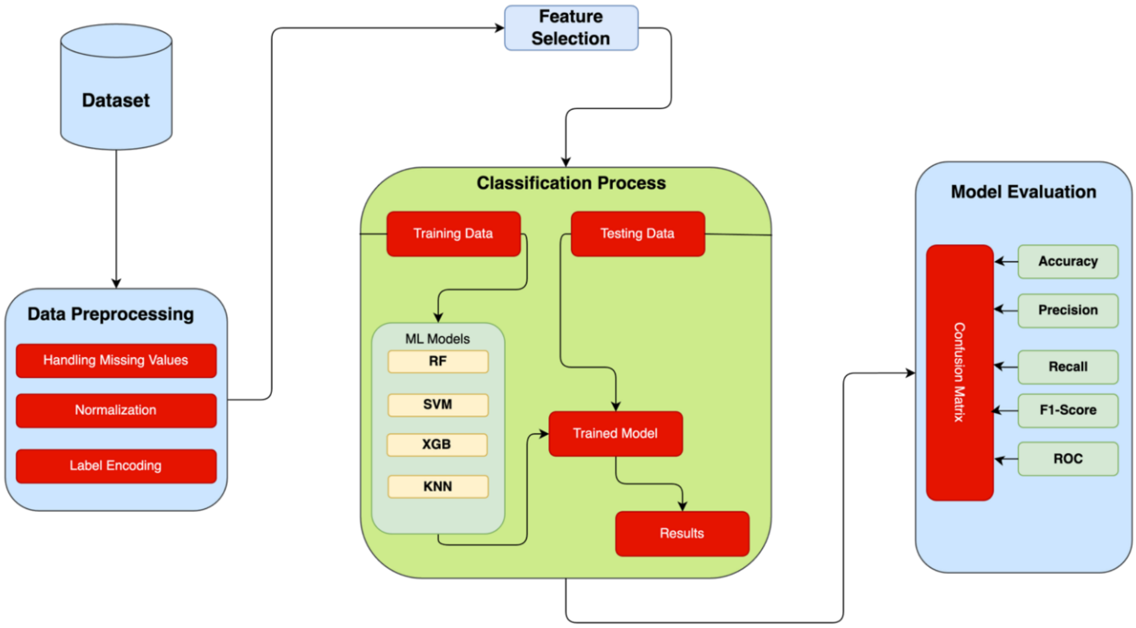

Figure 1: Methodological Diagram of the proposed Framework

4.1 Random Forest

Figure 5 shows the Confusion Matrix (CM) for the Random Forest (RF) classification of Irrigation techniques.

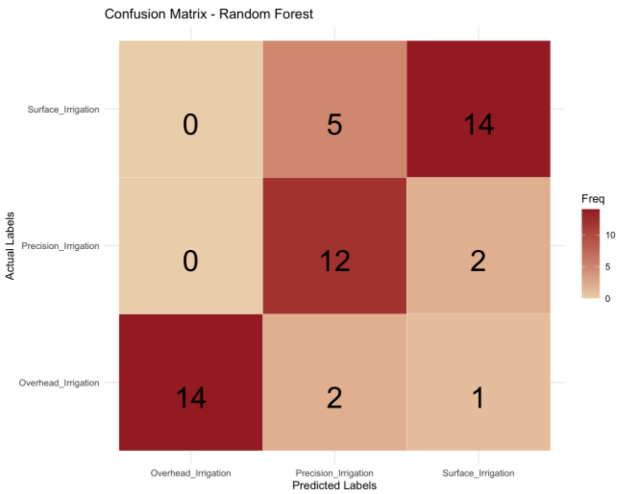

Figure 5: CM for RF

Based on figure 5, table 3 presents the performance evaluation of the RF model based on the three classes of irrigation techniques with each performance.

Table 3: RF Model Performance Evaluation

Metric	Overhead Irrigation	Precision Irrigation	Surface Irrigation
Precision	1.0	0.63	0.82
Recall	0.82	0.86	0.73
F1-Score	0.90	0.73	0.77
Accuracy		0.80	
Kappa		0.70	

Also, the roc curve for the random forest model is presented in figure 6 alongside its area under the cure for each of the three classes of irrigation colour in read blue and green colours respectively.

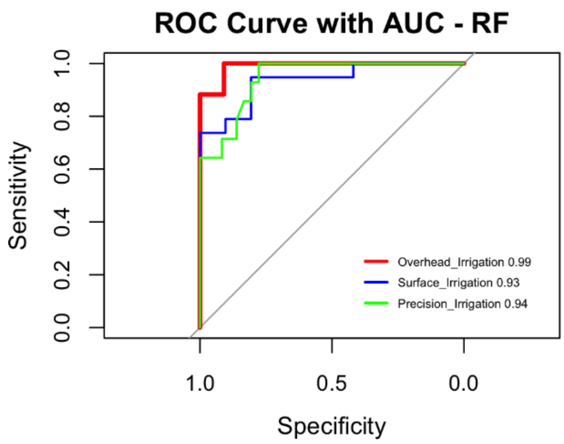

Figure 6: ROC - AUC for RF

4.2 Support Vector Machine

Figure 7 shows the Confusion Matrix (CM) for the SVM classification of Irrigation techniques. Also, the roc curve for the SVM model is presented in figure 8 alongside its area under the cure for each of the three classes of irrigation colour in read blue and green colours respectively

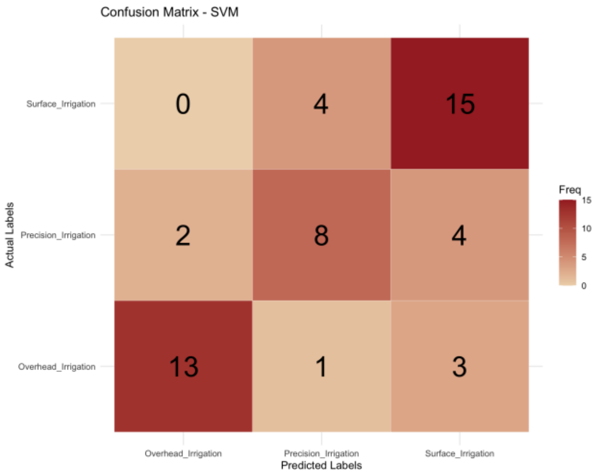

Figure 7: - CM for SVM

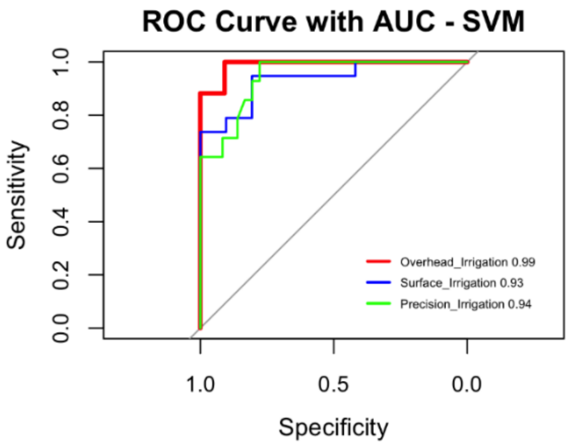

Figure 8: ROC - AUC for SVM

Based on figure 7, table 4 presents the performance evaluation of the SVM model based on the three classes of irrigation techniques with each performance.

Table 4: SVM Model Performance Evaluation

Metric	Overhead Irrigation	Precision Irrigation	Surface _Irrigation
Precision	0.87	0.62	0.68
Recall	0.76	0.57	0.79
F1-Score	0.81	0.59	0.73
Accuracy		0.72	
Kappa		0.57	

4.3 XGBoost

Figure 9 shows the Confusion Matrix (CM) for the XGBoost classification of Irrigation techniques. Also, the roc curve for the SVM model is presented in figure10 alongside its area under the cure for each of the three classes of irrigation colour in read blue and green colours respectively.

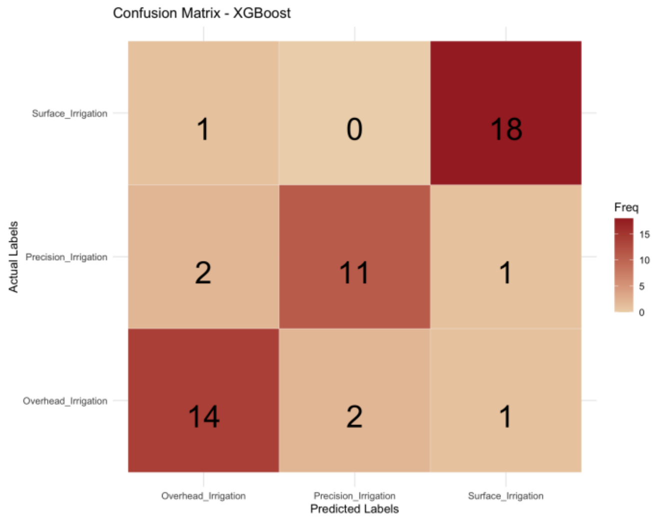

Figure 9: CM for XGBoost

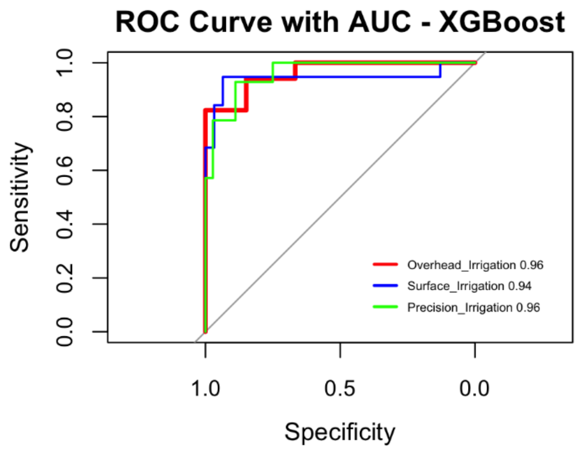

Figure 10: ROC - AUC for XGBoost

Based on figure 9, table 5 presents the performance evaluation of the XGBoost model based on the three classes of irrigation techniques with each performance.

Table 5: XGBost Model Performance Evaluation

Metric	Overhead Irrigation	Precision Irrigation	Surface Irrigation
Precision	0.82	0.85	0.90
Recall	0.82	0.79	0.95
F1-Score	0.82	0.82	0.92
Accuracy		0.86	
Kappa		0.79	

4.4 K-Nearest Neighbor

Figure 11 shows the Confusion Matrix (CM) for the KNN classification of Irrigation techniques. Also, the roc curve for the KNN model is presented in figure 12 alongside its area

International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, Delaware, USA
Volume 12– No. 47, June 2025 – www.ijais.org

under the cure for each of the three classes of irrigation colour in read blue and green colours respectively.

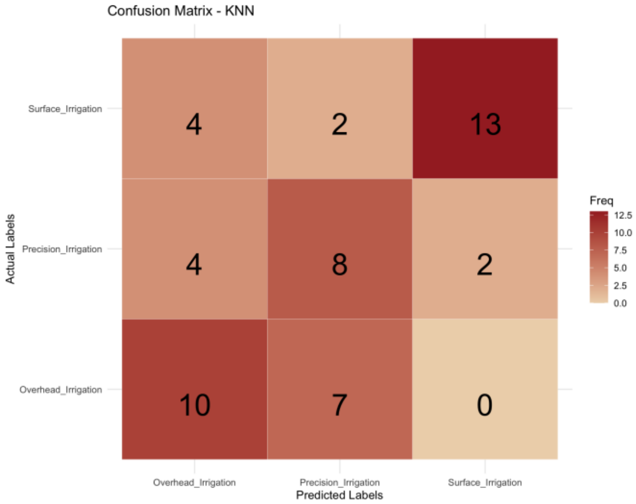

Figure 11: - CM for KNN

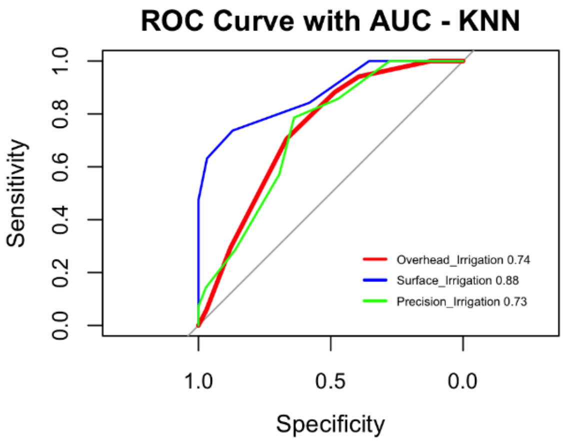

Figure 12: - ROC - AUC for KNN

Based on figure 11, table 6 presents the performance evaluation of the KNN model based on the three classes of irrigation techniques with each performance.

Table 6: KNN Model Performance Evaluation

Metric	Overhead_Irrigation	Precision_Irrigation	Surface_Irrigation
Precision	0.58	0.50	0.87
Recall	0.58	0.64	0.68
F1-Score	0.58	0.56	0.78
Accuracy		0.64	
Kappa		0.46	

4.5 Comparative Analysis of ML Models

In this section the presentation of the comparative evaluation of the models utilize in this study is presented in figure 13 and 14 respectively.

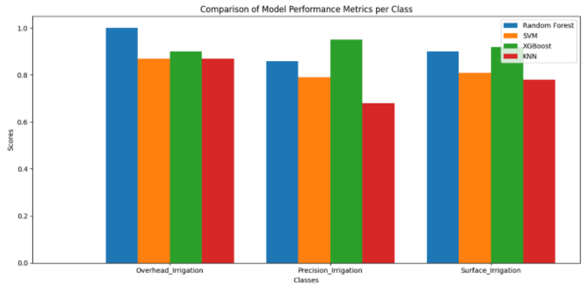

Figure 13: Model Preformace Per Class

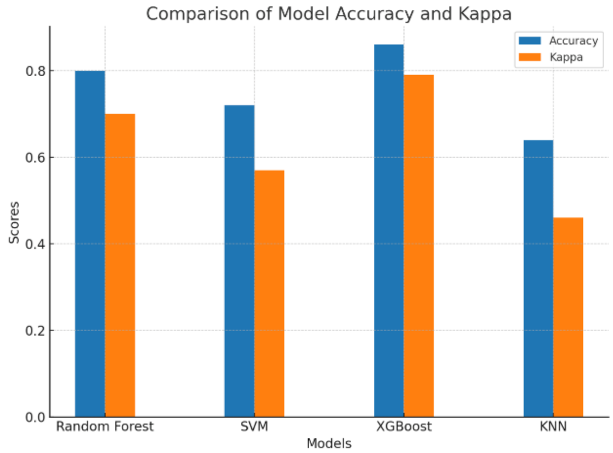

Figure 14: Comparison based on Accuracy and Kappa

The comparative analysis of the four machine learning models Random Forest, SVM, XGBoost, and KNN illustrates a clear distinction in performance across different evaluation metrics and irrigation classes. From figure 13, which depicts model performance per class, XGBoost stands out consistently, especially in the Precision Irrigation class where it achieves the highest score among all models. It also performs strongly in the Surface Irrigation and Overhead Irrigation categories, reflecting its robust adaptability and precision across varying data distributions. Random Forest follows closely, performing exceptionally well in the Overhead Irrigation class, where it achieves a perfect score of 1.0. Its performance across the other two classes remains commendable, suggesting strong generalization capability. SVM and KNN, in contrast, show relatively weaker performance, with KNN particularly underperforming in the Precision Irrigation category. SVM maintains a more balanced but modest performance across all classes, trailing slightly behind Random Forest and XGBoost. KNN, while occasionally competitive in the *Overhead* Irrigation class, significantly drops off in others, indicating potential issues with scalability or sensitivity to feature distributions. Further insights are gained from figure 14, which compare overall model accuracy and Cohen's Kappa scores. XGBoost again leads the pack with the highest accuracy (86%) and Kappa (0.79), reinforcing its reliability not only in correct predictions but also in maintaining agreement beyond chance. Random Forest also shows strong performance with 80% accuracy and a Kappa of 0.70, suggesting it is a solid choice for general classification tasks in this context. SVM achieves a moderate 72% accuracy and a lower Kappa of 0.57, indicating fair classification ability but reduced consistency. KNN exhibits the lowest performance with 64% accuracy and a Kappa of just 0.46, highlighting its limitations and potential unsuitability for more complex or diverse datasets. Overall, XGBoost emerges as the most effective model, followed by Random Forest, while SVM and especially KNN lag behind in both class-specific and general performance metrics.

5. CONCLUSION

This study explored the application of supervised machine learning algorithms to classify irrigation techniques using agricultural data. By employing Random Forest, Support Vector Machine, XGBoost, and K-Nearest Neighbors models, the research aimed to identify the most effective model for accurate irrigation prediction based on key farm-level features such as water usage, fertilizer and pesticide application, farm size, soil type, crop type, and season. The dataset, sourced from Kaggle, was preprocessed and analyzed through exploratory data analysis, feature selection via Pearson correlation, and model evaluation using metrics like accuracy, precision, recall, F1-score, and Cohen's Kappa. Among the four models, XGBoost emerged as the most effective, achieving the highest accuracy (86%) and Kappa score (0.79), followed closely by Random Forest with 80% accuracy and a Kappa of 0.70. SVM demonstrated moderate performance, while KNN showed the least predictive power across all evaluation metrics. The findings suggest that XGBoost is highly reliable for irrigation classification tasks in precision agriculture, offering strong generalization across diverse irrigation types. This study underscores the potential of machine learning to enhance decision-making in agricultural water management and supports the integration of data-driven approaches for sustainable and efficient irrigation practices. Future research could explore model scalability, real-time implementation, and the incorporation of remote sensing data for further performance enhancement. Future research can further the scalability of the models developed to process larger and more diverse datasets for other geographical areas to gain broader application in global agricultural systems. Real-time implementation of the models on IoT sensors and edge computing can revolutionize irrigation management via real-time decision-making over live data. In addition, the incorporation of remote sensing information, including satellite images and drone-based monitoring, could significantly enhance model accuracy and facilitate predictive analytics at increased spatial resolution. Further, the incorporation of temporal information along with climatic variables would help to render these models resilient to seasonal and environmental fluctuations and hence more robust and apt for real-world application in actual environments.

6. ACKNOWLEDGMENTS

All authors in this study discussed the results and contributed to the manuscript.

7. REFERENCES

[1] R. Jebakumar, "A Study on Smart Irrigation Using Machine Learning," Cell & Cellular Life Sciences Journal, vol. 4, no. 1, 2019, doi: 10.23880/cclsj-16000141.

[2] M. Attri et al., "Improved Irrigation Practices for Higher Agricultural Productivity: A Review," International Journal of Environment and Climate Change, pp. 51–61, Apr. 2022, doi: 10.9734/ijecc/2022/v12i930737.

[3] K. I. M. Abuzanouneh et al., "Design of Machine Learning Based Smart Irrigation System for Precision Agriculture," Computers, Materials and Continua, vol. 72, no. 1, pp. 109–124, 2022, doi: 10.32604/cmc.2022.022648.

[4] M. A. Youssef et al., "Enhancing irrigation water management based on ETo prediction using machine learning to mitigate climate change," Cogent Food Agric, vol. 10, no. 1, 2024, doi: 10.1080/23311932.2024.2348697.

[5] K. Liu, Y. Bo, X. Li, S. Wang, and G. Zhou, "Uncovering Current and Future Variations of Irrigation Water Use Across China Using Machine Learning," Earths Future, vol. 12, no. 3, Mar. 2024, doi: 10.1029/2023EF003562.

[6] J. I. Ubah, L. C. Orakwe, K. N. Ogbu, J. I. Awu, I. E. Ahaneku, and E. C. Chukwuma, "Forecasting water quality parameters using artificial neural network for irrigation purposes," Sci Rep, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-04062-5.

[7] R. Benameur, A. Dahane, B. Kechar, and A. E. H. Benyamina, "An Innovative Smart and Sustainable Low-Cost Irrigation System for Anomaly Detection Using Deep Learning," Sensors, vol. 24, no. 4, Feb. 2024, doi: 10.3390/s24041162.

[8] N. Dawn et al., "Implementation of Artificial Intelligence, Machine Learning, and Internet of Things (IoT) in revolutionizing Agriculture: A review on recent trends and challenges," International Journal of Experimental Research and Review, vol. 30, 2023, doi: 10.52756/ijerr.2023.v30.018.

[9] A. Glória, J. Cardoso, and P. Sebastião, "Sustainable irrigation system for farming supported by machine learning and real-time sensor data," Sensors, vol. 21, no. 9, May 2021, doi: 10.3390/s21093079.

[10] C. Carslake, J. A. Vázquez-Diosdado, and J. Kaler, "Machine learning algorithms to classify and quantify multiple behaviours in dairy calves using a sensor–moving beyond classification in precision livestock," Sensors (Switzerland), vol. 21, no. 1, pp. 1–14, Jan. 2021, doi: 10.3390/s21010088.

[11] D. Hernandez, L. Pasha, D. Arian Yusuf, R. Nurfaizi, and D. Julianingsih, "The Role of Artificial Intelligence in Sustainable Agriculture and Waste Management: Towards a Green Future," International Transactions on Artificial Intelligence (ITALIC), vol. 2, no. 2, pp. 150–157, Jun. 2024, doi: 10.33050/italic.v2i2.552.

[12] A. Rohan, M. S. Rafaq, M. J. Hasan, F. Asghar, A. K. Bashir, and T. Dottorini, "Application of deep learning for livestock behaviour recognition: A systematic literature review," Sep. 01, 2024, Elsevier B.V. doi: 10.1016/j.compag.2024.109115.

[13] M. E. Hossain, M. A. Kabir, L. Zheng, D. L. Swain, S. McGrath, and J. Medway, "A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions," Jan. 01, 2022, KeAi Communications Co. doi: 10.1016/j.aiia.2022.09.002.

[14] N. Siebrecht, "Sustainable Agriculture and Its Implementation Gap Overcoming Obstacles to Implementation," Sustainability, vol. 12, no. 9, p. 3853, May 2020, doi: 10.3390/su12093853.

[15] D. Dönmez, M. A. Isak, T. İzgü, and Ö. Şimşek, "Green Horizons: Navigating the Future of Agriculture through

Sustainable Practices," Sustainability, vol. 16, no. 8, p. 3505, Apr. 2024, doi: 10.3390/su16083505.

[16] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, and G. Ravid, "Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge," Precis Agric, vol. 19, no. 3, pp. 421–444, Jun. 2018, doi: 10.1007/s11119-017-9527-4.

[17] A. El Bilali, A. Taleb, and Y. Brouziyne, "Groundwater quality forecasting using machine learning algorithms for irrigation purposes," Agric Water Manag, vol. 245, Feb. 2021, doi: 10.1016/j.agwat.2020.106625.

[18] A. IORLİAM, S. BUM, I. S. AONDOAKAA, I. B. IORLIAM, and Y. SHEHU, "Machine Learning Techniques for the Classification of IoT-Enabled Smart Irrigation Data for Agricultural Purposes," Gazi University Journal of Science Part A: Engineering and Innovation, vol. 9, no. 4, pp. 378–391, Dec. 2022, doi: 10.54287/gujsa.1141575.

[19] https://www.kaggle.com/datasets/bhadramohit/agriculture-and-farming-dataset

[20] S. Inyang and I. Umoren, "Semantic-Based Natural Language Processing for Classification of Infectious Diseases Based on Ecological Factors," International Journal of Innovative Research in Sciences and Engineering Studies (IJIRSES), vol. 3, no. 7, pp. 11-21, 2023.

[21] I. Umoh, V. Essien, and S. Inyang, "Optimizing Hypertension Risk Classification through Machine Learning," *International Journal of Computer Applications*, vol. 186, no. 14, pp. 21–29, 2024.

[22] S. Inyang and I. Umoren, "From Text to Insights: NLP-Driven Classification of Infectious Diseases Based on Ecological Risk Factors," Journal of Innovation Information Technology and Application (JINITA), vol. 5, no. 2, pp. 154–165, 2023.