European Journal of Computer Science and Information Technology, 14(4),12-23, 2026
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Predictive Modeling of Agricultural Land Performance Using Digital Twin and Artificial Intelligence Techniques

Vu Ngoc Kien[1*], Hoang Nam Anh[2]
[1]Academy of Military Science and Technology, Vietnam; Kvu3009@gmail.com
[2]Hanoi Amsterdam High School for the Gifted;
Corresponding author: Vu Ngoc Kien

doi: https://doi.org/10.37745/ejcsit.2013/vol14n41223 Published June 15, 2026

Citation: Kien V.N. and Anh H.N. (2026) Predictive Modeling of Agricultural Land Performance Using Digital Twin and Artificial Intelligence Techniques, *European Journal of Computer Science and Information Technology*, 14(4),12-23

Abstract: *Agricultural land performance is a key factor in ensuring productivity, resource efficiency, and sustainable investment in modern farming. Traditional methods fail to account for complex interactions between soil properties and climatic variability, leading to suboptimal decisions. In order to facilitate data-driven decision-making in agriculture, the goal is to provide a precise forecasting model for agricultural land performance. This research develops a model for agricultural land performance by integrating Digital Twin technology with advanced Artificial Intelligence (AI) techniques. A digital twin was created using 10,000 data points to represent physical agricultural land, combining sensor data and remote sensing inputs that included soil moisture, weather parameters, and crop conditions. Normalization was used to maintain data uniformity. To improve efficiency and accuracy, Principal Component Analysis (PCA) was used for feature extraction to reduce dimensionality, remove redundant information, and identify the most important aspects driving land performance. The hybrid Magnetotactic Bacteria Optimized K-Nearest Neighbour-Tuned Tree Algorithm (MBO-KNN-TA) accurately predicts agricultural land performance using crop yield. MBOA optimizes parameters by exploring solution space; KNN identifies patterns in multi-dimensional feature data; and XGBoost predicts the agricultural land performance. Experimental results conducted in Python 3.10 indicate that the proposed model outperforms existing approaches, achieving 98.2% accuracy. The digital twin enables scenario simulation, risk-informed decision-making, and optimized resource allocation, making it a robust tool for precision agriculture. These findings demonstrate that integrating artificial intelligence (AI) with a digital twin framework can significantly enhance predictive accuracy, sustainability, and data-driven management of agricultural land.*

Keywords: digital twin, agricultural land performance, predictive modeling, machine learning, precision agriculture, decision support system

INTRODUCTION

Digital Twins (DT) provide for high-frequency data interchange, guaranteeing synchronization between real and virtual systems, as opposed to simulations or decision-support systems that require regular updates [1]. DTs were first introduced by Grieves' Manufacturing 1 (M1) and are currently used in agriculture, healthcare, construction, industry, automotive, and aviation. For system modeling and analysis, they use the Internet of Things (IoT), remote sensing, data analytics, and artificial intelligence [2]. While Global Navigation Satellite System (GNSS) is extensively utilized in affluent countries, less than 20% of farmers embrace advances such as Variable Rate Technology (VRT). High prices, extended payback times, poor technical maturity, limited field data, and rough terrains are some of the challenges [3].

Digital farming uses smart technologies to collect and examine data, therefore producing better decisions, increased output, cost savings, and process optimization. Artificial intelligence and large data help with monitoring, sensor analysis, crop health, yield enhancement, and risk reduction; that helps with farm management and supply chains [4]. Computer-based methods enhance accuracy, reduce crop cycles, and overcome the constraints of traditional phenotyping by forecasting crop yields and processing image data, given rising food demand and climate change. Crop prediction is challenging due to the connection of soil, temperature, and crops; standard methods are imprecise; and existing models have issues with basic presumptions, complexity, and high costs [5].

The model requires high-quality data, might have problems with generalizability and scalability, and might be computationally demanding with little interpretability. To overcome these limitations, this research aims to develop a Magnetotactic Bacteria Optimized K-Nearest Neighbour-Tuned Tree Algorithm (MBO-KNN-TA)-based digital twin framework for accurate agricultural land performance prediction using Artificial Intelligence (AI) techniques. The bKey Contribution of this research is follows.

- A novel Digital Twin–AI integrated model is developed to model agricultural land performance using multi-source data.
- A hybrid MBO-KNN-TA model is proposed, combining optimization, pattern recognition, and predictive learning techniques.
- The research demonstrates high predictive accuracy and supports sustainable agriculture through efficient resource utilization and risk-aware planning.

The research has the following sections. Section 1 provide introduction for digital twin framework. Section 2, Overview of previous research. Part 3 talks about Digital Twin. In Section 4, the datasets and methodology are discussed. Section 5: performance indicators/simulation outcomes/experimental results. The discussion of the research is in Section 6. Section 7 provides the conclusion to the research, as well as limitations and directions.

RELATED WORK

IoT sensors, Long Short-Term Memory (LSTM) networks, and Digital Twin (DT) architecture were used to monitor the soil moisture, temperature, and salinity [6]. This also rendered irrigation and drainage management smarter and more efficient, which allowed the making of saline soils sustainable for agriculture. Though generalization could not be done as site-specific soil stratification, coordinated irrigation facilitated better utilization of water and minimized salt. Even though challenges of data integration, complexity and scalability continued, DT applications in agriculture [7] enhanced decision-making through semantic web technologies, cloud-fog computing, and multi-agent systems to simulate irrigation, fertilization, nutrient and pest management. A combination with Reinforcement Learning (RL) and DTs [8] allowed crop management with recommendations of fertilizer and pesticides, although scaling and more general application were nonetheless problematic. This was effective in tests on apple scab and wheat.

Although the scalability of multi-crop further needs validation, the research [9] tried to minimize energy use and increase greenhouse efficiency with the help of a Digital Twin system that controlled the intensity of Light Emitting Diode (LED) and regulated with quantum sensors, RL, and adaptive Proportional Integral Derivative (PID) control, leading to the 23.6% decrease in energy use. Combining Light Use Efficiency (LUE) and Random Forest (RF) improved prediction of Winter Wheat (WW) and Oilseed Rape (OSR) to improve R^2 by 14.3% on meteorological data [10].

Research Gap

Current techniques were limited by their failure to record the complicated interactions between soil, climate, and management variables, which results in incorrect forecasts and poor decision-making. While DT-based systems relying on cloud-fog and semantic technologies have data integration, implementation complexity, and scalability issues [6], IoT and LSTM with DT together enhance irrigation efficiency but have restricted generalizability because of site-specific soil stratification [8]. Moreover, although combining RL with DT enhances adaptive decision-making, there were limits to large-scale implementation and more use [10].
The proposed method MBO-KNN-TA addresses these shortcomings by creating a strong, data-driven DT along with the sophisticated AI models that work with diverse data, enhance scaling, and respond to the changing environmental conditions. It enhances generality, data integration is easier, and accurate predictions regarding resource management and informed agricultural decision-making can be made.

Digital Twin (DT)

A Digital Twin (DT) is a dynamic virtual self-copy of a physical system that models physical conditions and functioning. The fusion of historical data on soil and crop, remote sensing data, and the data provided by IoT sensors results in the reproduction of agricultural land by DT. The DT assists through regularly aligning the virtual and physical systems, performance assessment, scenario modelling, and monitoring. Implementing the DT would allow us to accurately monitor the soil health, water consumption, and crop development and optimize resource allocation, as

European Journal of Computer Science and Information Technology, 14(4),12-23, 2026
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

well as give insights into predicting information. State-of-the-art AI algorithms, such as MBO-KNN-TA, are capable of accurately forecasts crop yields based on this point.

METHODOLOGY

The research used soil, ambient, and time-varying sensor data to simulate agricultural land performance, which was normalized and denoised to ensure accuracy. A virtual representation was built using a Digital Twin, and the MBO-KNN-TA model caught spatial-temporal patterns, resulting in high prediction accuracy as measured by RMSE. Figure 1 illustrates the system flow diagram.

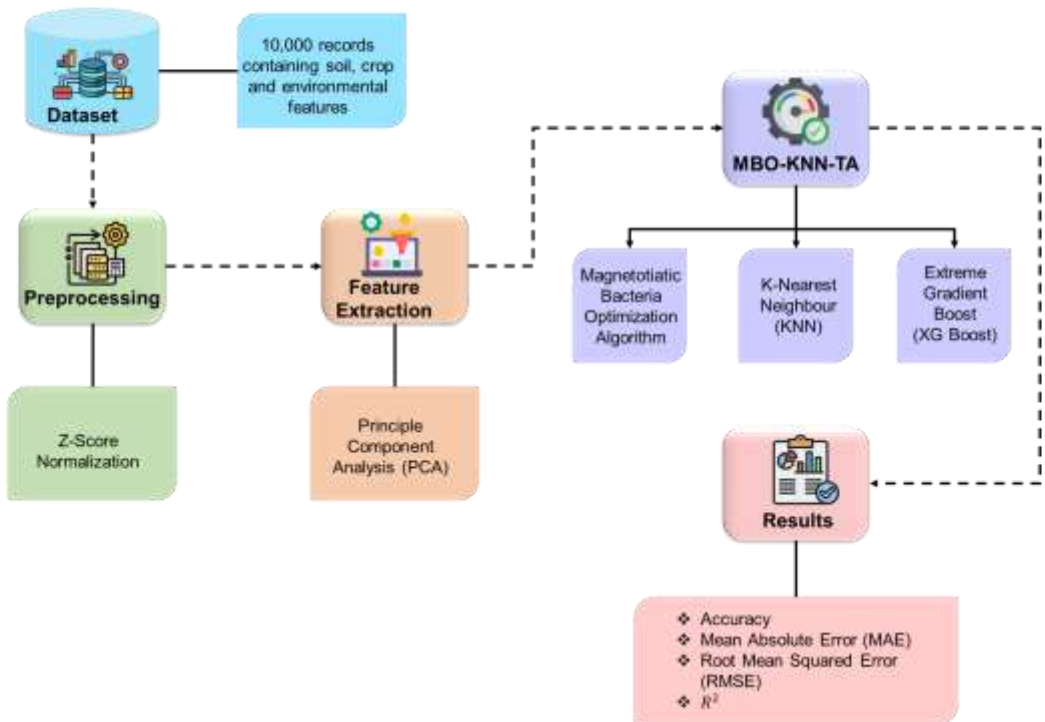

Figure 1. Method workflow of the proposed system

Dataset

The Smart Crop yield prediction dataset is publicly available (https://www.kaggle.com/datasets/miadul/smart-crop-yield-predication-dataset). For crop yield forecasting and soil health assessment, the dataset contains 10,000 data points, which include records of environmental, weather, crop, and soil variables. Divided into 70% (7000 data) training and 30% (3000 data) testing, it offers labeled data for constructing and assessing predictive models in precision agriculture.

Preprocessing

Remote sensing was coupled with IoT. Features were normalized using Z-score normalization, and noise was filtered.

Z-Score Normalization for Data Standardization

Z-score normalization transforms data by putting it on a single scale with a mean of 0 and a standard deviation of 1. It removes scale bias in soil, crop, and environmental variables so KNN can detect spatial-temporal patterns and offer more precise predictions of crop yield, soil condition, and water usage efficiency.

Dimensionality Reduction of Agricultural Data Using Principal Component Analysis (PCA)

After standardizing the data using z-score normalization, PCA is used to extract features. PCA minimizes the data in a farm containing the most significant aspects and also makes it smaller. This enhances the MBO-KNN-TA computational economy and predictive ability, as there are lower correlated variables like soil moisture, fertilizer concentration, temperature, rainfall, and crop status index. PCA was used to reduce dimensionality and identify essential features, which were then fed into the KNN model to capture spatial-temporal patterns in the data.

K-Nearest Neighbors (KNN) for Pattern Recognition

To find patterns, KNN is a machine learning technique that looks for the k closest data points in d-dimensional space. It helps to analyze crop, soil, and environmental factors in the Digital Twin model, therefore accurately predicting crop yield, soil health, and water usage efficiency. Equation (1) shows how the KNN algorithm is used to identify patterns.

$$I(D_l) = \Sigma_{w_j} ||w_j - \mu_l||^2 \tag{1}$$

For cluster D_l, $I(D_l)$ is the sum of intra-cluster variation. Σ_{w_j} represents the summation of all data points. The vector w_j represents the $j - th$ feature within the cluster, while μ_l is the cluster's average vector. The formula $||w_j - \mu_l||^2$ sums the squared Euclidean distances to gauge feature dispersion inside the cluster, therefore helping to maximize agricultural land performance.

Early estimates of crop production, soil health, and water use efficiency, along with the creation of spatial-temporal similarity patterns from standard soil, crop, and environmental data using KNN, contribute to the development of accurate predictive model. These outputs are supplied into MBOA for the last MBO-KNN-TA design to iteratively fine-tune model parameters, hence lowering error and enhancing prediction accuracy.

Optimization of Model Predictions Using Magnetotactic Bacteria Optimization Algorithm (MBOA)

Inspired by bacteria that change magnetic moments to reduce energy, the MBOA enhances model predictions of agricultural production, soil health, and water efficiency. MBOA continuously improves solutions by simulating magnetosome manufacture, amplification, and replacement. This enhances resilience and accuracy in the Digital Twin model for precise agricultural land performance prediction under many environmental conditions. Coefficient $C(W_j, W_q)$ measures feature magnitude via the Euclidean norm of C_j in Equation (2).

$$C(W_j, W_q) = \|C_j\|_2 \tag{2}$$

Where $C(W_j, W_q)$ represents the magnetotactic bacteria-inspired coefficient for the $j-th$ feature with respect to a reference W_q. The related feature vector, C_j, has a Euclidean (L2) norm, $\|C_j\|_2$, that shows how big the feature is. This helps to make the estimates of the performance of agricultural land better. Therefore, MBOA optimizes prediction parameters by efficiently exploring solution spaces, improving agricultural land performance accuracy, reliability, and computational efficiency.

Magnetotactic System (MTS) Amplification

The next position is updated toward the best-known position with a random factor to guide, as shown in Equation (3).

$$W_j(s+1) = U_{best}(s) + rand(1,K) * (U_{best}(s) - U_j(s)) \tag{3}$$

Step $(s+1)$ updated the $j-th$ feature position is $W_j(s+1)$. The best location discovered so far is $U_{best}(s)$, the current one is $U_j(s)$, and $rand(1,K)$ is a random element adding variation to direct the search for agricultural land performance optimization.

MTS Replacement

Equation (4) shows how the updated position is calculated using the reference factor and random variables.

$$W_j(s+1) = /n_{or}(s) * ((rand(1,K) - 1) * rand(1,K)) \tag{4}$$

Where $W_j(s+1)$ represents the new location of the $j-th$ feature at step $s+1$, K represents the number of dimensions. In the MBOA optimization for agricultural land performance, $/n_{or}(s)$ stands for reference or scaling factor; $rand(1,K)$ are random numbers.

After using MBOA to refine the model's parameters, KNN is used to identify spatial-temporal patterns in the information.

Prediction of Agricultural Land Performance Using Extreme Gradient Boosting (XGBoost)

XGBoost is an efficient and scalable boosting system that predicts crop yields, soil quality, and water use. Using the residual data, XGBoost is repeatedly fitted, so it can forecast crop output. It combines soil, crop, and environmental variables with a set of weak learners through loss gradients and regularization to avoid overfitting, thus achieving a strong predictive model that increases the correctness and reliability of predictions related to agricultural land performance.

$$\Theta = \sum_{j=1}^{N} K(z_j, \hat{z}_j) + \sum_{i=1}^{I} \Omega(e_i) \tag{5}$$

In Equation (5), Θ, overall objective function for the predictive model. N is the total number of data samples used for training. j is the index representing each individual data sample (from 1 to N). $K(z_j, \hat{z}_j)$, similarity or error between the actual z_j and expected $\hat{z}_j$ values for the $j-th$ feature, $\Omega(e_i)$ evaluates the $i-th$ error term e_i, ensuring accurate and optimized agricultural land performance predictions. I is the total number of model components. i is the **i**ndex representing each model component (from 1 to I). $\sum_{j=1}^{N}$ is the summation over all N is the data samples, adding the specified term for each sample. $\sum_{i=1}^{I}$ is the summation over all I model components, adding the specified term for each component. However, XGBoost predicts agricultural land performance accurately by analyzing environmental, soil, and crop-related features efficiently.

Hybrid MBO-KNN-TA model for Predictive Modeling of Agricultural Land Performance

The hybrid model is formed by utilizing KNN, MBO-KNN-TA, and XGBoost to develop predictive models using spatial and temporal datasets such as soil, crop, and environmental attributes. In the first stage, KNN conducts the analysis of the patterns, whereas MBOA is used to tune the parameters of KNN to minimize errors and ensure reliability. The tuned hybrid model is then integrated with XGBoost to decrease prediction errors. Figure 2 mentions the predictive hybrid MBO-KNN-TA model's architecture for agricultural land performance and the procedure of the hybrid model MBO-KNN-TA.

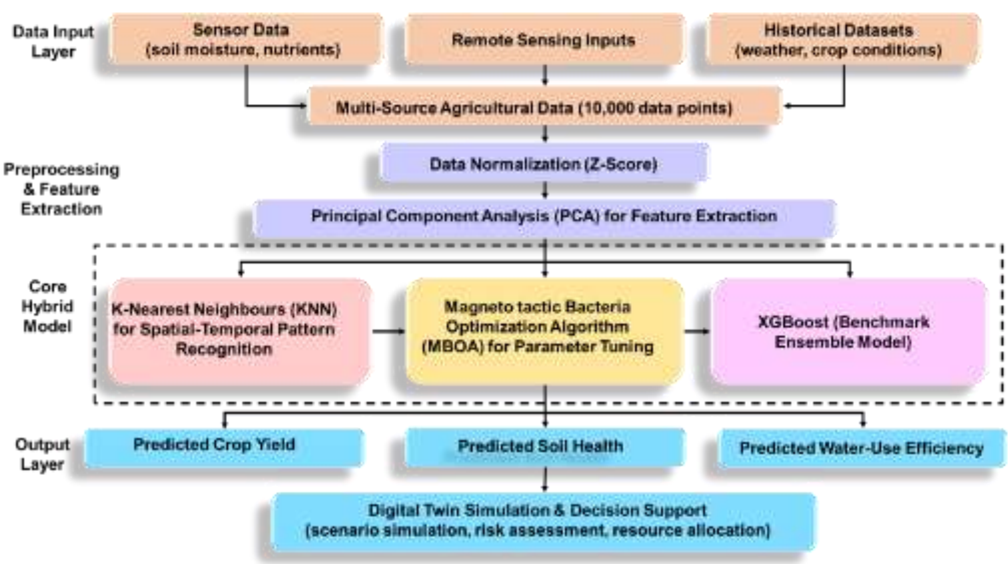

Figure 2. Architecture of the MBO-KNN-TA model for agricultural land performance

RESULT

This proposed model was implemented using Python 3.10 with NumPy, Pandas, Scikit-learn, and XGBoost libraries, along with a custom MBOA implementation. Experiments were conducted on a system equipped with an NVIDIA RTX 3080 GPU, 32 GB RAM, and an Intel Core i7 processor. Such an arrangement allowed the training of the MBO-KNN-TA model; it was possible to tune the hyperparameters and offered an opportunity to make accurate and reproducible outputs in terms of crop output, soil health, and water efficiency.

Performative Analysis

A steady decrease in error and a rapid increase in the Accuracy Curve indicates that great accuracy is seen in the Loss Curve and Accuracy Curve, respectively. Figure 3 shows the results favor a consistent and extended generalized model with no major overfitting in agricultural outcome prediction. A candlestick chart presents the area and the movement of data over time. The Digital Twin focuses on the volatility of agricultural yield, emphasizing high, low, and closing values using bodies and wicks to point them out so that it can provide better insights into the risk the investment might take and generate predictability.

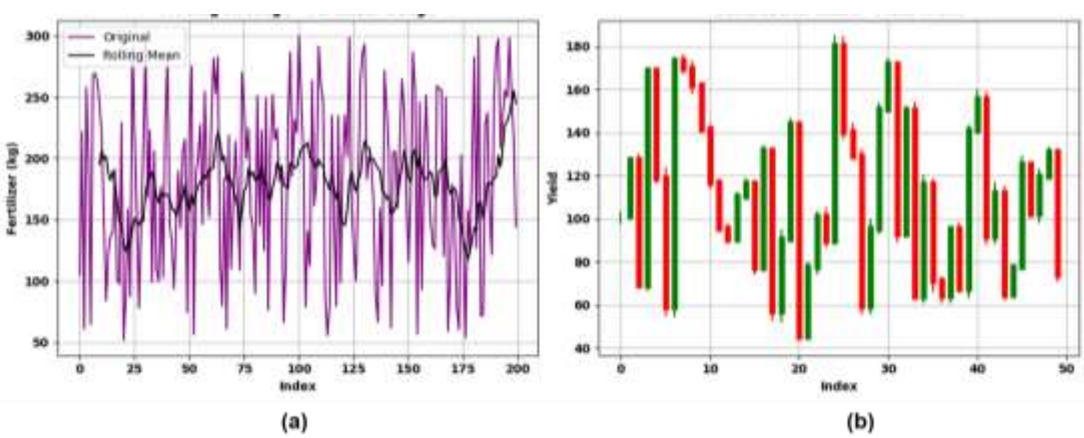

Figure 3. Illustration of (a) Moving Average Analysis for Fertilizer Usage Smoothing and Trend Detection, and (b) Candlestick Analysis for Crop Yield Volatility and Performance Trends.

Figure 3 (a) illustrates how a rolling mean (black) smoothes the original volatile data (purple) to eliminate high-frequency swings in fertilizer data. This reveals underlying patterns, reduces noise, improves the prediction accuracy of the Digital Twin, and allows for solid long-term projections of soil health and resource optimization. This enables the Digital Twin to recognize current trends, optimize resource consumption, and assist the AI model in predicting long-term soil health needs.

Figure 3 (b) depicts crop yield volatility, with green and red bodies indicating positive or negative growth. Wicks show the highest and lowest yields, emphasizing extreme swings. This approach enables the Digital Twin to assess risks, capture variability, and improve prediction accuracy beyond simple average trends.

Evaluation Metrics

Metrics such as Accuracy, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Median Absolute Error (MedAE) were used to empirically evaluate how effectively the constructed MBO-KNN-TA model predicted agricultural land performance, including crop yield, soil health, and water-use efficiency under varying environmental conditions.

Accuracy: Measures how accurately the MBO-KNN-TA model predicts agricultural land performance outcomes such as crop yield, soil health, and water-use efficiency, indicating the overall effectiveness of the predictive framework. **Mean Absolute Error (MAE):** measures the average absolute difference between actual and predicted agricultural land performance values, indicating the prediction accuracy of the MBO-KNN-TA framework. **Root Mean Square Error (RMSE):** quantifies the variance of prediction errors by penalizing larger deviations between actual and predicted agricultural performance outcomes. **Median Absolute Error (MedAE):** determines the median of absolute prediction errors, providing a robust evaluation of typical model error without being significantly affected by extreme outliers.

Comparative Analysis

The MBO-KNN-TA model outperforms Deep Neural Network (DNN), Long Short-Term Memory (LSTM), Extreme Gradient Boosting (XGBoost) [11], Random Forest (RF), and Decision Tree (DT) [12] in forecast accuracy and consistency, providing more reliable agricultural forecasts (Table 1). Both the Existing model and the proposed model were trained on the Smart Crop yield prediction dataset and environment. The proposed MBO-KNN-TA model achieved better results with the suggested model.

Table 1. Comparative Analysis of Trained Models for Agricultural Land Performance Prediction Using the Proposed Dataset

Model	Accuracy (%)	MAE	RMSE	R^2 (%)
FMIG-RFE-SVM	96.8	0.041	0.234	94.2
Random Forest	93.5	0.078	0.348	87.8
Decision Tree	87.2	0.149	0.542	70.1
XGBoost	95.3	0.058	0.287	91.5
Proposed Model	98.2	0.020	0.142	97.9

All models were trained on the same dataset to ensure a fair comparison. The proposed MBO-KNN-TA model achieved the highest accuracy (98.2%) and R^2 (97.9%), with higher MAE (0.020), RMSE (0.142), showing better overall performance than Fuzzy Mutual Information Gain–Recursive Feature Elimination–Support Vector Machine (FMIG-RFE-SVM), XGBoost, RF, and DT in predicting crop yield, soil health, and water efficiency.

DISCUSSION

The agricultural land performance model is an integration of DT and the more advanced MBO-KNN-TA model. Some methods like FMIG-RFE-SVM [11], are also sensitive to feature selection, and are incapable of dealing with high-dimensional effects, or changing temporal behaviour. Although RF [12] research well, it is difficult to comprehend and has trouble with overfitting in noisy, correlated data as well as capturing temporal patterns. Though they are easy to understand, DT [12] tends to overfit, and are not effective with time, complex and nonlinear data. Even though XGBoost [12] is effective, it interprets time changes with a lot of difficulty, it requires much readjustment, and is not interpretable.

To enhance agricultural prediction this research developed a hybrid MBO-KNN-TA. MBOA optimizes model parameters, KNN identifies complex spatial relationships and the altered tree form will enhance predictions. The model improves the accuracy, precision and dynamism of the simulations of agricultural yield, soil health and water-use efficiency.

CONCLUSION

An accurate agricultural land performance prediction model integrating Digital Twin technology with the MBO-KNN-TA. To calculate crop production, soil status, and water efficiency, it had to normalize the datasets, PCA to reduce dimensions, training the model. The results provided good performance in precision agriculture with an accuracy of 98.2%, MAE of 0.02, an RMSE of 0.142 and an R^2 of 97.9%.

The reliance on quality sensors and historical data, scalability to bigger areas and interpretation of the data related to the complexity of MBO-KNN-TA are some of the limitations. Future research will focus on real-time agricultural monitoring using IoT-enabled Digital Twin systems, integrating satellite analytics, adaptive DL, and edge computing for scalable precision farming.

REFERENCES

1. Badshah, A., Alkazemi, B.Y., Din, F., Zamli, K.Z. and Haris, M., 2024. Crop classification and yield prediction using robust machine learning models for agricultural sustainability. *IEEe Access*, *12*, pp.162799-162813. https://doi.org/10.1109/ACCESS.2024.3486653
2. Kim, S. and Heo, S., 2024. An agricultural digital twin for mandarins demonstrates the potential for individualized agriculture. *Nature Communications*, *15*(1), p.1561. https://doi.org/10.1038/s41467-024-45725-x
3. Cesco, S., Sambo, P., Borin, M., Basso, B., Orzes, G. and Mazzetto, F., 2023. Smart agriculture and digital twins: Applications and challenges in a vision of sustainability. *European Journal of Agronomy*, *146*, p.126809. https://doi.org/10.1016/j.eja.2023.126809
4. Nasirahmadi, A. and Hensel, O., 2022. Toward the next generation of digitalization in agriculture based on the digital twin paradigm. *Sensors*, *22*(2), p.498. https://doi.org/10.3390/s22020498
5. Wang, T., Tong, R., Xu, T., Li, Y. and Chen, Y., 2026. Artificial intelligence in plant science: From image-based phenotyping to yield and trait prediction. *Frontiers in Plant Science*, *16*, p.1732979. https://doi.org/10.3389/fpls.2025.1732979
6. Qin, M., Zhang, C., Ma, G., Ma, Y., Feng, X., Li, P., and Feng, K., 2025. Digital twin-enabled intelligent irrigation-drainage system for precision water-salt management in saline agroecosystems. *Agricultural Water Management*, *322*, p.109957. https://doi.org/10.1016/j.agwat.2025.109957
7. Kalyani, Y., Vorster, L., Whetton, R. and Collier, R., 2024. Application scenarios of digital twins for smart crop farming through cloud–fog–edge infrastructure. *Future Internet*, *16*(3), p.100. https://doi.org/10.3390/fi16030100
8. Kheir, A.M., Govind, A., Nangia, V., Devkota, M., Elnashar, A., Omar, M.E.D. and Feike, T., 2024. Developing automated machine learning approach for fast and robust crop yield prediction using a fusion of remote sensing, soil, and weather dataset. *Environmental Research Communications*, *6*(4), p.041005. https://doi.org/10.1088/2515-7620/ad2d02
9. Bua, C., Wang, Y., Ojo, M.O., Giordano, S., and Zahid, A., 2026. Digital Twin-Enabled Multi-Zone Adaptive Lighting Control in Greenhouses Using Reinforcement Learning Optimization. *Smart Agricultural Technology*, 14(2026), p.101984. https://doi.org/10.1016/j.atech.2026.101984

10. Dhillon, M.S., Dahms, T., Kuebert-Flock, C., Rummler, T., Arnault, J., Steffan-Dewenter, I. and Ullmann, T., 2023. Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oilseed rape. *Frontiers in Remote Sensing*, *3*, p.1010978. https://doi.org/10.3389/frsen.2022.1010978
11. Abdel-Salam, M., Kumar, N. and Mahajan, S., 2024. A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. *Neural Computing and Applications*, *36*(33), pp.20723-20750. https://doi.org/10.1007/s00521-024-10226-x
12. Sharma, P., Dadheech, P., Aneja, N., and Aneja, S., 2023. Predicting agricultural yields based on machine learning using regression and deep learning. *IEEE Access*, *11*, pp.111255-111264. https://doi.org/10.1109/ACCESS.2023.3321861