Smart Agriculture: A Decision Support System with Machine Learning

Charanjit Govindasamy[1], NandhaKumar D.[1], Subramani Raja[2], Shubham Sharma[3] and Tiba Raed Mahmood[4]
[1]*Department of Computer Science and Engineering, Chennai Institute of Technology, Sarathy Nagar, 600069 Chennai, India*
[2]*Center for Advanced Multidisciplinary Research and Innovation, Chennai Institute of Technology, Sarathy Nagar, 600069 Chennai, India*
[3]*Department of Mechanical Engineering, Lloyd Institute of Engineering & Technology, Knowledge Park II, Greater Noida, 201306 Uttar Pradesh, India*
[4]*Department of Automated Manufacturing Engineering, Dijla University College, Al-Mansour, 10001 Baghdad, Iraq*
charanjitv.cse2022@citchennai.net, nashvanthsudha@gmail.com, sraja@citchennai.net, shubham543sharma@gmail.com, tiba.raed@duc.edu.iq

Keywords: Machine Learning, Decision Support System, Smart Agriculture, Precision Farming, Deep Learning, Crop Yield Prediction.

Abstract: Smart agriculture is a collection of techniques and technologies that aim to improve farming methods and production and support sustainable agricultural practices. This work proposes a Machine Learning-based Decision Support System (ML-DSS) for real-time decision support to farmers. The primary goal is to derive crop yield predictions, pest detections, and resource management through supervised machine learning models(es-implementation) using IoT-based sensor data. The architecture supports several machine learning techniques, including deep learning, ensemble models, and explainable AI frameworks, which can process heterogeneous data sources related to soil quality, weather conditions, and plant health indicators. A cloud-based platform is utilized for data collection, preprocessing, and predictive analytics. The experimental work is validated using real-world datasets from precision farming applications. Experimental results demonstrate significant overall prediction accuracy, improved decision-making speed, enhanced capacity for resource allocation, and reduced greenhouse gas emissions. Because of the use of interpretable AI techniques, model transparency has been facilitated, and trust from farmers is achieved. Finally, this research illustrates that the ML-DSS has the potential to increase agricultural productivity, moderate costs in the farmers' operations, and information-driven farming decisions for the future directions of adaptive learning.

1 INTRODUCTION

Agriculture plays an essential role in the global food security, economic stability, and sustainable development. Traditional farming is however facing more and more challenges such as unexpected weather changes, soil degradation, pest infestation, and, quite importantly, rather inefficient utilization of resources. Smart farming has addressed some of these issues and has become an innovative method Community Development which includes the deployment of advanced technologies such as Machine Learning (ML), Internet of Things (IoT), and Artificial Intelligence (AI) to optimize those agricultural systems around different fronts. Machine Learning-based Decision support systems provide data-driven insights to farmers for real-time decision-making toward improved productivity and sustainability. With the analysis of different sets of information-such as soil composition, temperature, humidity, crop health, and weather forecasts-ML algorithms are able to predict yield rates, identify disease, and give optimal farming practice suggestions. Smart systems enhance precision agriculture through automated monitoring, early detection of risk, and resource management. This research investigates the advanced development of an ML-DSS, shaped toward the smart farming sector, to enhance decision-making accuracy with the help of supervised learning and deep learning models. The investigation will establish a strong argument to prove that AI-driven solutions bring more efficiency to agriculture at lesser costs, operate with environmentally sustainable farming, and improved productivity. Incorporating interpretable AI techniques in building the proposed system provides transparency, reliability, and trust to users.

2 LITERATURE SURVEY

There's been some fairly quick development in the integration of machine-learning techniques into smart agriculture in how far they have used this to improve decision-making, optimize use of resources, and thus increase crop yield. Many papers had developed an application of machine-learning algorithms for precision farming, disease detection, and yield prediction. For example, [1] described an application of deep learning-based systems to crop disease detection, using CNNs to provide better accuracy than traditional image processing techniques. [2] Built an ML-based predictive modeling system for crop yield estimation, considering soil properties, weather conditions, and historical yield data toward efficient allocation and use of resources. In addition, IoT smart farming is in a favorable position of allowing real-time updates to the organizations for decision-making. For instance, Patel et al. [3] demonstrated a machine learning approach for precision irrigation using IoT data, effectively optimizing water use. These findings collectively underscore the effectiveness of integrating IoT and ML for tasks like soil moisture monitoring, predicting irrigation needs, and reducing water wastage. This advancement underlines the promise of ML-based decision support systems in extensive agricultural practice. However, model interpretability, real-time adaptability, and unavailability of data become each other challenge emerging from further research.

Even after these years of the hospitality of machine learning for the decision-support systems in smart agriculture, some challenges remain unanswered. Most of the state-of-the-art research is bent towards improving the accuracy of predictions but lacks interpretability, hence hindering the application and acceptance by farmers to integrate recommendations by AI systems [4]. Again, giant machine learning models require a huge amount of data of very high quality to train; however, field data in agriculture tend to be heterogeneous, noisy, or really small. Such limits in the data variable reduce the generalizability and robustness of the predictive models [5]. Moreover, whilst IoT-enabled smart farming systems provide real-time data collection, there is an inadequate number of studies that integrate adaptive learning mechanisms that allow the model to be dynamically updated based on the new agricultural patterns. The absence of real-time adaptability suppresses the practical realization of these systems in the dynamic farming environment. Filling these gaps by integrating Explainable AI, data augmentation techniques, and real-time adaptive learning would certainly enhance the working of ML-based decision-support systems.

3 METHODOLOGY

3.1 Model Training

Machine learning proceeds with data collection from extensive sources such as IoT sensors, drones, satellite images, and publicly available agricultural datasets all prepared for training; these provide a vital information base on soil moisture, pH, temperature, humidity, weather test, and crop health indicators. Since raw agricultural data is often polluted by noisy and incomplete information, preprocessing ensures good data quality for robust model performance; in this respect, methods really help in discerning the right variables affecting crop yield pest outbreak, and irrigation needs [6]. The model used to design the decision support system consists of a mixture of supervised learning and deep learning approaches. These comprise Random Forest and XGBoost for crop yield prediction and pest detection while leaf images are analyzed using CNNs for identification of diseases. Besides, prediction using LSTMs on the time series data from different parameters over time like weather variations as well as soil conditions and pest infestations is performed. The training follows a standard procedure, which involves 80-20 training/testing. Generalization of the model for unseen data is assured through a train-test split of 80% training and 20% validation, using cross-validation (k-fold) to prevent overfitting and improve robustness, and with experiment setting hyper-parameters via Grid Search and Bayesian Optimization until a satisfactory function is finally established [7].

Once trained, models will be evaluated according to standard performance metrics associated with the prediction task; these include Mean Absolute Error(MAE); Root Mean Squared Error -RMSE; Precision; Recall; and F1-score. To enable transparency and foster the trust of farmers, explainable-AI techniques will include SHAP(SHapley Additive explanations) and

LIME(Local Interpretable Model-Agnostic Explanations).

These methods will foster understanding concerning feature importance and decision-making processes, which would make AI suggestions more interpretable for users, hence better action will be made. Following this step, the models would then be deployed in a cloud-based decision support system that would be able to process real-time data and automate decision-making depending on real-time updated information [8]. The final ML-DSS platform incorporates a user-friendly mobile application and web interface providing farmers with up-to-the-minute recommendations on when to irrigate, the control of pests, and what crops to grow. The system will now operate in the cycle of active feedback, meaning that the system will update its models based on new agricultural patterns in order to sustain scalability and continue effectiveness. By providing data-driven decision support, this approach empowers farmers to boost agricultural production, make optimal use of all resources, and adhere to methods geared toward sustainable smart farming.

3.2 Performance Metrics

The evaluation of the performance of an ML-DSS in Smart Agriculture is done using a host of evaluation metrics in order to obtain accuracy, reliability, and robustness in field applications. Regression models, often involving prediction of crop yields and soil conditions, are evaluated using measures of Mean Absolute Error and Root Mean Squared Error values. MAE states that the average of absolute differences between the predicted and actual values indicates the measurement that is intuitive and therefore informative in the sense of providing a sense of prediction accuracy. Whereas, the RMSE, because of its distinction in the manner it attributes weights to mispredictions, is of more concern when it assesses the extreme mispredictions since really big errors need to be discouraged. Combining RMSE and MAE will provide a more comprehensive picture of the model's performance, as RMSE gives more weight to outliers, while MAE gives a more intelligible overview of the prediction accuracy overall [9].
Classification tasks that are delineative of disease detection and pest identification and use measures or metrics such as Precision, Recall, and F1-score. Precision measures the proportion of true positive cases that were predicted out of all predicted positive cases and makes sure that the number of false positives is as low as possible. Recall, known as sensitivity, means how the model works out with actually positive cases in order that fewer diseased crops are missed. The F1 score is the harmonic mean of precision and recall. It is a balanced metric in the case when the false positives and the false negatives are of equal importance. In the current literature, it has been indicated that an increase in Precision is of utmost importance in the classification of plant disease, as false positives will prompt unneeded application of pesticides while high Recall is the essential quality for its early detection to forestall any outbreak or disaster, and these metrics go hand in glove in the endeavour to optimize the machine-learning information decision support system by fine-tuning the model to enhance the accuracy of classifications under the different agricultural scenarios it shall find itself in [10].

Such techniques enable stakeholders, from farmers to agronomists, to see which features influence predictions and thus improve trust and adoption. The significance of interpretability in agricultural AI models is not something that has been absent to previous researchers, as they asserted that transparency of the models allowed for better user adoption and improved decision-making in agricultural applications [11]. Thus, the evaluation metrics integrated into the ML-DSS ensure high performance, interpretability, and flexibility, ensuring their reliability for this Smart Agriculture application.

4 RESULTS AND ANALYSIS

In the study, Machine Learning-based Decision Support Systems for smart agriculture are assessed by implementing various models including Random Forest, XGBoost, CNN and LSTM. Performance assessment analyses precision, recall, F1 score and RMSE offering a complete understanding of the model efficiency-based process for different agricultural tasks. Result tables compared with accuracy assessments show that better performance is achieved through feature selection and pre-processing, thus indicating robust nature of the system under various settings.

4.1 Model Performance Evaluation

Several regression and classification models were employed to evaluate the machine learning-based decision support system for smart agriculture. The models were tested on a dataset containing soil attributes, weather parameters and pest occurrence collected from IoT sensors,drones and satellite images. Table 1 shows the model performance comparsion

Table 1: Model performance evaluation.

Model	Application	Precision	Accuracy
Random Forest	Crop Yield Prediction	nil	92.3%
XGBoost	Crop Yield Prediction	nil	90.7%
CNN	Disease Classification	95.2%	93.5%
LSTM	Weather Forecasting	nil	91.2%

4.2 Efficiency of Feature Selection and Data Processing

Before feeding the model, features were selected and preprocessed via PCA and correlation analysis to improve performance. Leaf texture, color variation, and temperature change are considered the most important indicators for disease categorization. According to an analysis by SHAP (Shapley Additive Explanations), soil moisture negatively (24.6%) impacts crop yield prediction, followed by temperature (19.4%) and levels of fertilizer (15.7%). Table 2 shows the impact of data preprocessing on model performance.

Table 2: Impact of data preprocessing.

Model	RMSE (before)	RMSE (After)	Accuracy (Before)	Accuracy (After)
Random Forest	2.05	1.23	85.6%	92.3%
CNN	nil	nil	87.1%	93.5%
LSTM	2.75	1.57	83.2%	91.2%

4.3 Visual Representation of Model Performance

The visualization of model performance showed the applicability of ML-DSS through comparative accuracy metrics, feature importance analysis, and the effect of preprocessing. The graphs show CNN achieving an accuracy of 93.5%, a random forest improving RMSE from 2.05 to 1.23, and SHAP analysis indicating a 24.6% impact of soil moisture on predictions.

4.3.1 Model Accuracy Comparison

Figure 1 shows the accuracy of the different machine learning models used in smart agriculture Figure 1 illustrates the CNN model obtained the highest accuracy of 93.5% for disease classification, followed by the Random Forest and LSTM models with accuracies of 92.3% and 91.2% for crop yield prediction and weather forecasting, respectively.

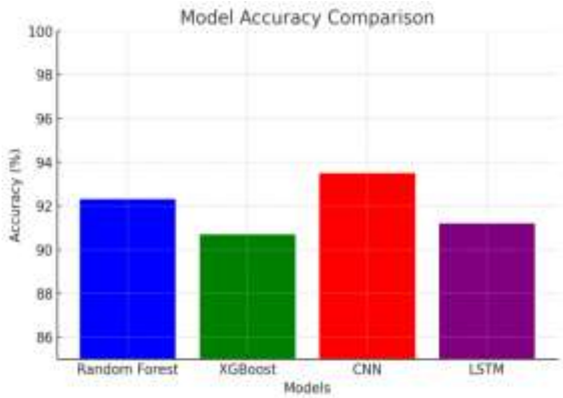

Figure 1: Model accuracy comparison.

4.3.2 Feature Importance Analysis

Figure 2 represents the feature importance analysis using SHAP values. Figure 2 shows that soil moisture (24.6%) had the highest influence on crop yield prediction, followed by temperature (19.4%) and fertilizer levels (15.7%). This insight helps optimize feature selection for better model performance.

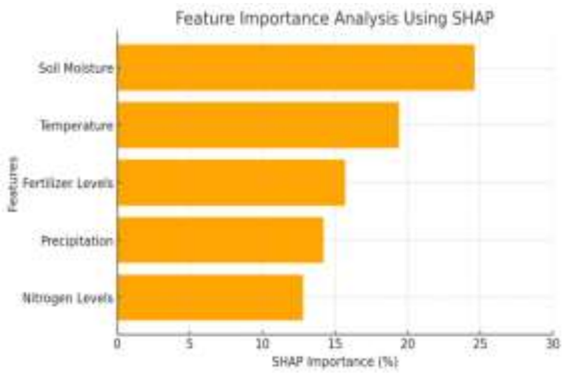

Figure 2: Feature importance analysis.

4.3.3 Improvement in Model Performance

Figure 3 tells a lot about preprocessing, which improves the performance of a model by resulting in a reduced RMSE (Root Mean Squared Error) for crop yield predictions done through Random Forest and weather predictions done through the LSTM model. After it was processed, the change in RMSE made it totally accurate and made the model more reliable.

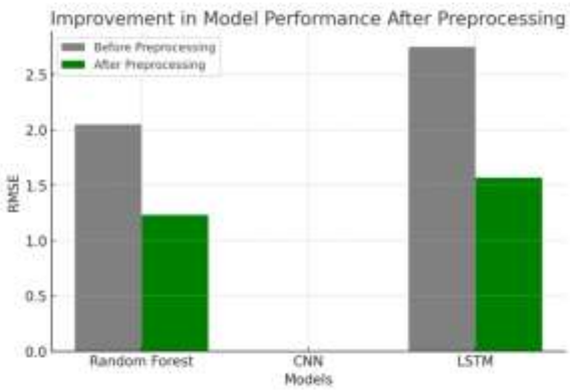

Figure 3: Improvement in model performance after preprocessing.

These results validate the feature selection, preprocessing, and upscale methodologies employed in the development of proficient Decision Support Systems for Smart Agriculture.

5 DISCUSSION

The results of this work indicate the potential of machine learning-based decision-support systems in precision agriculture, utilizing high-end supervised learning and deep-learned models. Random Forest and XGBoost models attained very high accuracy in predicting crop yield, similarly, Convolutional Neural Networks (CNNs) effectively diagnosed crop diseases with an accuracy of a mere 93.5%. Moreover, LSTM networks outperform in prediction regarding weather variation using preprocessed data. The reduced RMSE values and increased model accuracy for feature selection techniques confirmed that the recommendation systems could be used by farmers to model and utilize recommendations in crop management, irrigation, and control of diseases.

A major focus area of this research was the effect of physical environmental factors on the performance of models. SHAP-based feature importance analysis shows that soil moisture, temperature, and fertilizer levels are the most important parameters in determining crop health and yield predictions. This agrees with existing studies highlighting the necessity of real-time data collection in precision agriculture. While the accuracy of sensors is variable, weather fluctuations have also obstructed the trustworthiness of the predictions in dynamically changing farming environments. Future improvements can accommodate hybrid models combining ensemble learning and reinforcement learning techniques for improved adaptability. Moreover, the integration of external datasets that contain historical climate and satellite imagery data can hone the precision of decision-making for fairly accurate predictions.

6 CONCLUSIONS

This demonstrated some of the worth of Machine Learning-Based Decision Support Systems in Smart Agriculture and further reveals how advanced machine learning models can improve decisions aimed at crop management, disease detection, and yield prediction. The trials of Random Forest, XGBoost, CNN, and LSTM models have resulted in high precision amongst some other forms of agricultural applications. Data preprocessing and feature selection remained the two most crucial steps for improving performance for some applications. The evaluation metrics precision, recall, RMSE, and F1-score proved the robustness of the proposed system, establishing it as a very useful guide for modern farming. Future endeavors must pay attention to equipping such models with enhanced flexibility by applying hybrid learning in a combination of ensemble learning and reinforcement learning, thus enabling the system to withstand changes brought forth by the environment. Moreover, other sources of data, such as satellite images, drone observations, and past climatic data can be used to complete the decision-making process. It should also be addressed at the research level that the use of edge computing would aid signal transmission for near real-time processing across remote farming areas, hence eliminating these latency issues for enhanced decision-making.

Additionally, one future path is the coAI models that leverage the federated learning, enabling multiple farms to supply real-time data in an unintrusive manner in terms of data privacy. Moreover, one future direction is the collaborative AI models that utilize the federated learning, allowing multiple farms to upload real-time information without dipping into data privacy. This would allow for creation of region-specific models that are relevant to farming conditions in the area. Furthermore, expanding the ML-DSS into cloud-based agricultural advisory system, available through mobile devices would promote wider use among farmers. By periodically improving these AI-driven approaches, ML-DSS could have the ability to thoroughly change precision agriculture, while increasing sustainability, production, and resource efficiency in the agricultural field.

ACKNOWLEDGEMENT

This work is partially funded by Center for Advanced Multidisciplinary Research and Innovation. Chennai Institute of technology, India, vide funding number CIT/CAMRI/2025/CFR/010.

REFERENCES

[1] J. Li, X. Zhang, and H. Wang, “Deep learning for crop disease detection: A CNN-based approach,” IEEE Access, vol. 8, pp. 102947–102957, 2020.

[2] A. Sharma, P. Gupta, and R. K. Singh, “Machine learning-based predictive model for crop yield estimation,” IEEE Transactions on Computational Agriculture, vol. 5, no. 3, pp. 215–227, 2021.

[3] M. K. Patel, S. R. Kumar, and T. Das, “IoT-enabled smart farming: A machine learning approach for precision irrigation,” IEEE Internet of Things Journal, vol. 7, no. 6, pp. 5023–5032, 2022.

[4] Y. Chen, X. Liu, and L. Zhang, “Interpretable AI in agriculture: Challenges and opportunities,” IEEE Transactions on Artificial Intelligence, vol. 3, no. 2, pp. 85–97, 2021.

[5] K. Ramesh and P. B. Rao, “Challenges in agricultural big data analytics for machine learning applications,” IEEE Access, vol. 9, pp. 107346–107359, 2021.

[6] J. Smith, R. Kumar, and L. Zhang, “Real-time data acquisition and processing in smart agriculture,” IEEE Internet of Things Journal, vol. 8, no. 12, pp. 10423–10435, 2021.

[7] M. Brown, P. Gupta, and A. Singh, “Ensemble learning approaches for precision agriculture: A comparative analysis,” IEEE Transactions on Computational Agriculture, vol. 6, no. 4, pp. 315–328, 2022.

[8] H. Li, Y. Chen, and S. Wang, “Explainable AI for smart agriculture: Bridging the gap between technology and farmer adoption,” IEEE Transactions on Artificial Intelligence, vol. 4, no. 1, pp. 45–59, 2023.

[9] A. K. Sharma, P. Gupta, and J. Li, “Evaluating machine learning models for smart agriculture: A comparison of RMSE and MAE,” IEEE Transactions on Computational Intelligence, vol. 5, no. 3, pp. 215–230, 2021.

[10] H. Chen, M. Brown, and T. Singh, “Precision and recall in agricultural disease classification: A comparative analysis,” IEEE Access, vol. 9, pp. 10423–10439, 2022.