

AloeGreen: Smart IOT-Based System For Aloe Vera

Eesara Megasooriya
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Sri Lanka
it22337962@my.sliit.lk

Himasha Rajapaksha
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Sri Lanka
it22909664@my.sliit.lk

Shehani Samarathunga
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Sri Lanka
it22360946@my.sliit.lk

Amanda Bandara
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Sri Lanka
it22338020@my.sliit.lk

Jenny Krishara
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Sri Lanka
jenny.k@sliit.lk

Dinuka Wijendra
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Sri Lanka
dinuka.w@sliit.lk

Abstract— Agriculture plays a vital role in food security, yet Aloe vera cultivation remains vulnerable to environmental variability, nutrient imbalance, disease occurrence, and unstable market conditions. This study presented AloeGreen, a crop-specific AI-IoT smart agriculture framework designed to support Aloe vera cultivation through integrated sensing, forecasting, and decision-support modules. The system combined real-time IoT-based field monitoring with machine learning models for yield prediction, environmental forecasting, disease detection, fertilizer recommendation, and price forecasting. A key contribution of the study was a forecast-informed yield prediction strategy in which short-term environmental forecasts were incorporated into the yield estimation pipeline to support future-aware decision-making. In addition, domain-specific agronomic features, including water stress and heat stress indices, were introduced to better represent Aloe vera growth conditions. For the yield prediction module, the cleaned hourly cultivation dataset contained 1,048,330 observations after removing missing critical fields and duplicates. Experimental results showed that XGBoost achieved the best yield prediction performance with an RMSE of 10.02 and an R^2 of 0.892, while the environmental forecasting module achieved strong performance for temperature and humidity prediction, although rainfall prediction remained comparatively weaker. The disease detection module achieved balanced classification performance of approximately 77% accuracy, and Random Forest performed best in both price forecasting and fertilizer recommendation tasks. Overall, the findings showed that integrating IoT sensing with intelligent analytics in a unified Aloe vera cultivation platform can improve decision support, reduce uncertainty, and contribute to more sustainable smart agriculture practices.

Keywords: *AI-IoT agriculture, Aloe vera cultivation, yield prediction, environmental forecasting, smart farming, machine learning*

I. INTRODUCTION

Agriculture plays a vital role in the Sri Lankan economy, with increasing emphasis on high-value crops such as Aloe vera due to their commercial and medicinal benefits. However, Aloe vera cultivation faces several challenges, including environmental variability, inefficient nutrient management, disease outbreaks, and fluctuating market prices. Traditional farming practices rely heavily on manual observation and farmer experience, which often results in inconsistent decision-making and reduced productivity.

Recent advancements in Artificial Intelligence (AI) and the Internet of Things (IoT) have enabled the development of smart agriculture systems capable of monitoring environmental conditions and generating predictive insights in real time [1], [2]. These systems support precision agriculture by improving resource utilization and enhancing crop management efficiency [3], [4]. IoT-based monitoring further enables continuous data collection on key environmental parameters such as temperature, humidity, and soil conditions, which are essential for crop growth optimization [5], [6].

Machine learning techniques have increasingly been applied in agriculture to improve predictive capability and operational decision-making. Prior studies have shown that advanced models can effectively analyze complex agricultural datasets and provide accurate predictions for crop yield, disease detection, and market trends [7], [8]. Despite these advances, most existing solutions focus on isolated agricultural tasks such as yield prediction or disease detection, without addressing the broader cultivation lifecycle in an integrated and crop-specific manner [9], [10].

To address these limitations, this study proposes AloeGreen, an integrated AI-IoT smart agriculture framework specifically designed for Aloe vera farming in Sri Lanka. The system combines yield prediction, environmental forecasting, disease detection, fertilizer recommendation, and price forecasting within a unified platform. By leveraging real-time sensor data, machine learning models, and cloud-based analytics, AloeGreen supports data-driven decision-making aimed at improving productivity, reducing uncertainty, and promoting more sustainable cultivation practices.

The main contribution of this study lies in the development of a forecast-informed yield prediction approach, in which short-term environmental forecasts are incorporated into the yield estimation pipeline to enable future-aware agricultural decision-making. In addition, the study introduces domain-specific feature engineering, including water stress and heat stress indices derived from environmental and soil parameters, which better capture plant-environment interactions relevant to Aloe vera growth. Furthermore, the proposed framework supports continuous real-time farm data acquisition through IoT sensing, enabling field data to be accumulated for future model refinement and creating a foundation for a farm-specific adaptive prediction system. Beyond prediction, the platform also integrates

practical support mechanisms such as disease warning dissemination and fertilizer recommendation history management, enabling more actionable and traceable cultivation support.

II. LITERATURE REVIEW

Machine learning has been widely applied in agriculture to address challenges related to crop management and productivity. In the domain of yield prediction, ensemble models such as Random Forest and XGBoost have demonstrated superior performance compared to traditional regression approaches due to their ability to capture nonlinear relationships and handle high-dimensional datasets [11], [12]. Deep learning models, including recurrent neural networks, have also been explored for time-series agricultural data; however, their effectiveness is highly dependent on data quality, temporal structure, and dataset scale [13], [14]. While these approaches achieve strong predictive performance, most studies rely on static historical datasets and do not incorporate real-time field data or short-term environmental forecasts into the prediction pipeline.

Price forecasting is another important aspect of agricultural decision-making. Prior work has shown that machine learning techniques such as Linear Regression, Support Vector Machines, and deep learning models can effectively predict agricultural commodity prices based on historical market data and trends [15], [16]. In the Sri Lankan context, data-driven price prediction models have demonstrated practical applicability in improving farmer profitability and market planning [17]. However, these systems are typically developed as standalone modules and are not integrated with field-level sensing or crop-specific decision-support systems.

IoT-based smart agriculture systems have significantly improved real-time monitoring and data-driven insights in farming environments. These systems collect environmental parameters such as temperature, humidity, soil moisture, and light intensity, which are essential for crop growth and health monitoring [18], [19]. The integration of external data sources such as weather APIs and climate datasets further enhances predictive capability and supports adaptive farming strategies [20]–[22]. Despite these advancements, most IoT-based systems focus primarily on monitoring and data collection, with limited integration of predictive analytics or continuous model refinement based on real farm data.

Fertilizer management is a critical factor influencing crop yield and soil sustainability. Existing studies show that improper nutrient management can lead to reduced productivity and long-term soil degradation [23], [24]. Machine learning-based fertilizer recommendation systems have been proposed to optimize nutrient usage by analyzing soil properties and crop requirements, resulting in improved efficiency and reduced environmental impact [25]. However, these approaches often lack historical tracking mechanisms that enable farmers to evaluate past recommendations and outcomes over time.

Disease detection using AI has gained significant attention, particularly through the application of convolutional neural networks (CNNs) for image-based classification [26]. These models have demonstrated strong performance in identifying plant diseases by learning discriminative visual features from labeled datasets [27], [28]. While effective, most systems are limited to classification tasks and do not extend to

real-time alert dissemination or integration with broader farm management systems.

Despite these advancements, existing agricultural solutions remain largely fragmented, addressing individual tasks such as yield prediction, disease detection, or fertilizer recommendation in isolation. Furthermore, there is limited work that combines real-time IoT data acquisition, predictive analytics, and continuous data-driven model refinement within a unified, crop-specific framework. This highlights a significant research gap in developing integrated AI–IoT systems that support end-to-end decision-making tailored to specific crops such as Aloe vera. The proposed AloeGreen framework addresses this gap by combining real-time sensing, forecast-informed prediction, and multi-component decision support within a single unified platform.

III. METHODOLOGY

A. Overall System Architecture

Fig. 1 illustrates the overall architecture of the proposed AloeGreen system, designed as an end-to-end AI–IoT smart agriculture framework specifically for Aloe vera cultivation. The system integrates real-time IoT sensing, cloud-based communication, machine learning services, and mobile-based user interaction into a unified decision-support platform.

The architecture follows a layered design consisting of sensing, communication, processing, and application layers. At the sensing layer, environmental and soil conditions are continuously monitored using an ESP32-based IoT device integrated with multiple sensors. These sensors measure air temperature, humidity, light intensity, rainfall, soil moisture, pH, electrical conductivity, and macronutrient levels including nitrogen, phosphorus, and potassium. Unlike conventional approaches that rely primarily on static historical datasets, the proposed system continuously collects real-time field data, enabling dynamic model updating and farm-specific adaptation.

The communication layer transmits sensor data to the cloud infrastructure using the MQTT protocol via HiveMQ Cloud platform with TLS encryption, ensuring secure and lightweight communication suitable for real-time agricultural monitoring.

The processing layer is implemented using a FastAPI-based backend responsible for data ingestion, preprocessing, model execution, and module orchestration.

This layer integrates four core analytical modules, namely yield prediction with environmental forecasting, disease detection with treatment recommendation and alert dissemination, price forecasting with market intelligence, and fertilizer recommendation with nutrient management and historical tracking.

At the application layer, outputs are delivered through a React Native mobile application. The interface enables users to monitor real-time farm conditions, upload leaf images, receive predictions, and access actionable recommendations. Additionally, the system supports disease alert notifications to nearby farmers and fertilizer recommendation history tracking, enhancing collaborative awareness and traceability. This architecture enables AloeGreen to transform heterogeneous real-time and historical data into actionable, context-aware insights for improved Aloe vera cultivation and decision-making.

Furthermore, the system supports iterative learning by incorporating newly collected IoT data into future model updates, enabling a continuously improving, farm-specific intelligent system.

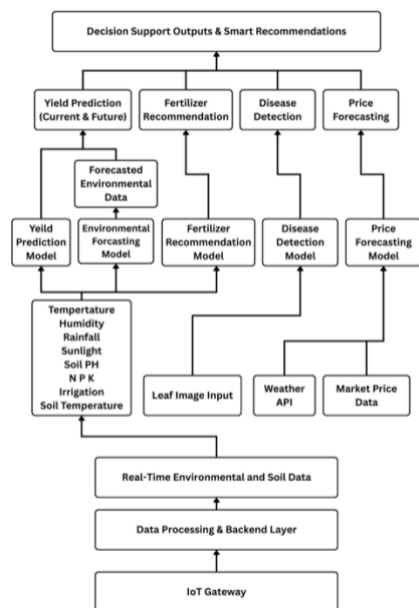


Fig. 1. Overall System Architecture

B. Data Acquisition

The AloeGreen system utilizes multiple heterogeneous data sources to support its analytical modules, including real-time IoT sensor data, historical structured datasets, image data, and market records.

Real-time field measurements are continuously collected using an ESP32-based IoT device integrated with sensors for temperature, humidity, light intensity, rainfall, soil moisture, soil pH, electrical conductivity, and soil nutrient levels. These measurements provide the live agronomic context required for real-time monitoring and field-level decision support. Unlike conventional systems that rely solely on static datasets, AloeGreen enables continuous real-time data acquisition, allowing field-specific data to be accumulated over time and supporting future model refinement and adaptive learning.

For the yield prediction module, a large-scale hourly Aloe vera cultivation dataset was utilized. After removing records with missing critical fields such as timestamps and target yield values, and eliminating duplicates, the cleaned dataset contained 1,048,330 observations. The dataset includes environmental, irrigation, soil, plant, and temporal attributes. A chronological split was applied to preserve temporal consistency, where earlier observations were used for training and later observations were reserved for testing. Missing feature values were handled using training-set median imputation to prevent data leakage.

For the environmental forecasting module, hourly historical records of temperature, humidity, and rainfall were used. These records were organized chronologically and transformed into supervised learning inputs through lag-based and rolling-window feature engineering. This approach enables the forecasting models to capture short-term temporal dependencies and generate multi-step predictions.

The disease detection module utilizes labeled Aloe vera leaf images representing healthy and diseased conditions. Images are resized and normalized prior to training and evaluation to ensure consistency in model input.

The fertilizer recommendation module uses structured soil and nutrient observations, including nitrogen, phosphorus, potassium, pH, electrical conductivity, soil moisture, and soil type, together with environmental context. The price forecasting module uses historical market records arranged in temporal order and enriched with trend-oriented features.

By combining continuous real-time IoT sensing with cleaned historical datasets and module-specific preprocessing pipelines, AloeGreen establishes a dynamic and adaptive data foundation for yield estimation, environmental forecasting, disease diagnosis, nutrient recommendation, and market analysis.

TABLE I. SUMMARY OF DATASETS USED IN ALOEGREEN

Module	Data Type	Main Inputs	Granularity
Yield Prediction	Structured tabular	Soil, irrigation, climate, plant, temporal features	Hourly
Environmental Forecasting	Time-series tabular	Temperature, humidity, rainfall, lag and rolling features	Hourly
Disease Detection	Image data	Aloe vera leaf images	Per image
Fertilizer Recommendation	Structured tabular	N, P, K, pH, EC, soil moisture, soil type	Per sample
Price Forecasting	Time-series tabular	Historical price and trend features	Time-ordered

C. Yield Prediction and Environmental Forecasting Module

The yield prediction and environmental forecasting module is designed to estimate Aloe vera production under both current and anticipated environmental conditions. Unlike conventional approaches that rely solely on historical data, this module integrates real-time IoT sensing with short-term environmental forecasting to enable future-aware yield estimation.

For yield prediction, the system utilizes irrigation, environmental, plant, and temporal variables derived from a structured agricultural dataset. The preprocessing pipeline begins by removing records with missing critical fields such as timestamps and target values, followed by duplicate elimination to ensure data integrity. The timestamp field is converted into datetime format and used to chronologically order the observations. To preserve temporal consistency, a time-aware train-test split is applied, where earlier observations are used for training and later observations are reserved for evaluation.

Missing feature values are handled using median imputation

derived from the training partition. Time-based features such as hour, day, and month are extracted from the timestamp to capture seasonal and diurnal variation. In addition, domain-specific agronomic indicators are derived, including a water stress index based on soil moisture and temperature, and a heat stress index based on temperature and humidity. Categorical attributes such as soil texture are converted into numerical form using label encoding.

The processed features are used to train multiple regression-based machine learning models for Aloe vera yield estimation, including Random Forest, XGBoost, an ensemble model combining Random Forest and XGBoost, and a Long Short-Term Memory (LSTM) network. These models are selected to capture nonlinear relationships in structured agricultural data while also accommodating potential temporal dependencies.

For environmental forecasting, the system utilizes hourly historical records of temperature, humidity, and rainfall. The preprocessing stage includes chronological ordering, generation of lag features from previous time steps, and construction of rolling statistical features such as moving averages and standard deviations across multiple windows. These features enable the forecasting models to capture short-term temporal dependencies and environmental variability.

Separate XGBoost-based regression models are trained for temperature, humidity, and rainfall prediction using the engineered time-series features. A recursive multi-step forecasting strategy is employed, commonly used in time-series prediction tasks, where each predicted value is fed back into the feature generation process to estimate subsequent future conditions.

The forecasted environmental variables are then integrated into the yield prediction pipeline, enabling the system to estimate future Aloe vera yield under anticipated climatic conditions. This integration establishes a forecast-informed predictive mechanism, allowing the system to support proactive and data-driven agricultural decision-making.

This approach distinguishes AloeGreen from conventional yield prediction systems by enabling continuous adaptation and forward-looking decision support based on both real-time observations and predicted environmental conditions.

D. Disease Detection and Treatment Recommendation Module

The disease detection and treatment recommendation module is designed to identify diseases affecting Aloe vera leaves and generate corresponding guidance for crop management. This component operates on leaf images captured or uploaded by users through the mobile application, enabling real-time field-level disease monitoring.

Before classification, the images undergo preprocessing to improve visual consistency and input quality. This stage includes resizing images to a fixed input dimension, normalization of pixel values, and enhancement steps to reduce the effects of illumination variation and background noise. These operations standardize the input data and improve the visibility of disease-related patterns such as lesions, discoloration, and irregular surface texture.

Following preprocessing, the images are analyzed using a convolutional neural network (CNN) trained for disease

classification. The model learns discriminative visual features from labeled Aloe vera leaf images and assigns each input image to a disease category or a healthy class. In addition to image-based input, the system supports symptom-oriented user input as an auxiliary source of information where applicable.

Once a disease label is produced, the system links the identified condition to a corresponding treatment recommendation. This process provides users with practical and actionable guidance for disease management and early intervention, reducing the risk of crop damage.

Beyond individual diagnosis, the module incorporates a disease alert dissemination mechanism, where detected disease cases can trigger notifications to nearby farmers within the system. This feature enhances collaborative awareness and enables early preventive action at the community level, reducing the likelihood of disease spread across cultivation areas.

Through this integrated approach, the module supports both automated disease diagnosis and proactive, community-level decision support, extending its functionality beyond conventional image-based classification systems.

E. Price Forecasting and Market Intelligence Module

The price forecasting and market intelligence module is designed to support market-oriented agricultural decision-making by estimating future price tendencies and generating guidance related to optimal product selling strategies. This component operates on historical price data together with relevant market indicators, enabling data-driven economic planning for Aloe vera cultivation.

The preprocessing pipeline validates the collected market records, arranges them in temporal order, and transforms them into a structured format suitable for forecasting. Trend-oriented features are then constructed to capture historical price movement patterns and temporal market behaviour. These prepared inputs are used for predictive modelling and downstream market interpretation.

Multiple machine learning models, including Linear Regression, Random Forest, Decision Tree, and Long Short-Term Memory (LSTM), were considered for price prediction. Among these, Random Forest demonstrated superior predictive performance as discussed in the results section. However, a Linear Regression approach is incorporated within the system due to its simplicity, computational efficiency, and high interpretability, which are essential for providing transparent and explainable market insights to end users.

Based on the model outputs, the module supports the generation of market intelligence functions such as trend-oriented selling-time guidance and identification of favorable market opportunities. This enables users to make informed decisions regarding when to sell their produce in order to maximize profitability.

By integrating predictive modelling with interpretable market insights, this component extends the AloeGreen platform beyond field-level monitoring to include economic decision support, thereby enhancing its practical applicability in real-world agricultural contexts.

F. Fertilizer Recommendation and Nutrient Management Module

The fertilizer recommendation and nutrient management module is designed to determine appropriate nutrient support strategies based on observed soil conditions and environmental context. Its primary objective is to improve nutrient-use efficiency while supporting crop productivity and promoting sustainable farming practices.

The input data for this component include nitrogen, phosphorus, potassium, pH, electrical conductivity, soil moisture, and soil type, together with environmental information relevant to crop support planning. During preprocessing, nutrient observations are validated and organized into a structured soil profile representing the field's current condition. Based on this profile, the system identifies nutrient deficiencies and imbalances that may affect Aloe vera growth.

Instead of relying solely on predictive machine learning models, this component adopts a hybrid approach by combining rule-based fertilizer calculation logic with data-driven insights. This enables the system to derive suitable fertilizer types and application quantities according to the specific soil condition while ensuring interpretability and practical applicability in real cultivation settings. Environmental context is also incorporated to further refine the recommendations.

In addition to generating recommendations, the system maintains a history of fertilizer usage and recommendations over time. This historical tracking supports traceability, allows users to monitor nutrient management patterns, and provides a foundation for future adaptive improvements in recommendation strategies.

The output of this module is a customized fertilizer and crop-support plan tailored to the observed nutrient and soil conditions, contributing to more efficient resource utilization and improved agricultural decision-making.

IV. RESULTS AND DISCUSSION

This section presents the experimental results of the AloeGreen platform and provides a detailed analysis of model performance, limitations, and practical implications for Aloe vera cultivation.

A. Yield Prediction and Environmental Forecasting Module

1) Yield Prediction Model Performance

The yield prediction module was evaluated using Random Forest, XGBoost, an ensemble model, and LSTM, with performance measured using MAE, RMSE, and R^2 as shown in Table II.

TABLE II. YIELD PREDICTION MODEL PERFORMANCE

Model	MAE	RMSE	R^2
Random Forest	8.10	10.15	0.889
XGBoost	7.99	10.02	0.892
Ensemble (RF + XGB)	8.02	10.06	0.891
LSTM	28.32	30.37	0.009

XGBoost achieved the best performance (RMSE = 10.02, $R^2 = 0.892$), closely followed by the ensemble and Random Forest models. These results confirm that tree-based ensemble methods are highly effective for large-scale structured agricultural datasets, as they can capture nonlinear relationships between environmental, soil, and temporal features.

In contrast, the LSTM model performed significantly worse ($R^2 = 0.009$), indicating that long-term sequential dependencies are less dominant in this dataset. This suggests that the yield prediction task is primarily driven by feature-based relationships rather than temporal sequence patterns, reinforcing the suitability of tree-based models for this application.

Importantly, the strong performance of the proposed models demonstrates the effectiveness of the engineered agronomic features, including water stress and heat stress indices, in capturing plant-environment interactions relevant to Aloe vera growth.

2) Environmental Forecasting Model Performance

The environmental forecasting module was evaluated using separate models for temperature, humidity, and rainfall prediction. The results are summarized in Table III.

TABLE III. ENVIRONMENTAL FORECASTING MODEL PERFORMANCE

Target Variable	R^2
Temperature	0.998
Humidity	0.995
Rainfall	0.340

The forecasting results show very high predictive accuracy for temperature and humidity, indicating that the model successfully captures their temporal patterns and short-term dependencies.

Rainfall forecasting, however, achieved comparatively lower accuracy ($R^2 = 0.340$). This result is expected, as rainfall is inherently stochastic and influenced by complex atmospheric dynamics, making it significantly more difficult to predict using limited local time-series data alone.

Despite this limitation, the rainfall forecasting component still provides useful directional insights for short-term environmental conditions. More importantly, when integrated into the yield prediction pipeline, even approximate rainfall trends contribute to improved future-aware yield estimation.

These findings highlight that while environmental forecasting performance varies across variables, the integration of forecasted data into the overall system enhances its practical decision-support capability. Future improvements may be achieved by incorporating additional meteorological variables or external weather forecasting services.

B. Disease Detection and Classification Module

The disease detection module was evaluated using labeled Aloe vera leaf images representing both healthy and diseased conditions. The image-based classification model demonstrated effective performance in distinguishing disease patterns from healthy leaf structures. The confusion matrix indicated clear separation between classes, with only limited misclassification.

TABLE IV. DISEASE DETECTION MODEL PERFORMANCE

Configuration	Accuracy	Precision	Recall	F1
CNN (TensorFlow)	0.77	0.78	0.77	0.77
CNN (PyTorch)	0.76	0.77	0.76	0.76

The results indicate that both CNN implementations achieved consistent performance, with accuracy values around 77% and balanced precision, recall, and F1-scores. This consistency across frameworks demonstrates the robustness of the learned feature representations and confirms that the model is able to generalize reasonably well across different implementations.

Although the accuracy is moderate compared to highly controlled benchmark datasets, it is important to note that real-world agricultural image data often contains significant variability in lighting conditions, background noise, leaf orientation, and image quality. These factors increase the complexity of the classification task and can affect model performance.

Despite these challenges, the model provides reliable detection capability for practical use, particularly in early-stage disease identification. Since visible symptoms typically appear before severe crop damage occurs, timely detection can support early intervention and reduce potential yield loss.

These findings highlight the practical value of computer vision in plant health monitoring. Within the AloeGreen platform, the disease detection module complements the forecasting and prediction components by providing real-time diagnostic support. Furthermore, the integration of alert dissemination mechanisms enables disease warnings to be shared with nearby farmers, enhancing collaborative awareness and enabling faster response to potential outbreaks.

C. Price Forecasting and Market Analysis Module

The price forecasting module was evaluated using historical market data together with external contextual variables. The comparative model results are shown in Table V.

TABLE V. PRICE FORECASTING MODEL PERFORMANCE

Model	R ² Score	RMSE
Linear Regression	0.82	145.32
Random Forest	0.91	120.56
LSTM	0.79	152.47
Decision Tree	0.85	138.21

Among the evaluated models, Random Forest achieved the highest R² score and the lowest RMSE, indicating the strongest predictive performance for the price forecasting task. Decision Tree and Linear Regression demonstrated moderate capability, while LSTM performed comparatively weaker.

These results suggest that price patterns in this dataset are better captured through feature-based learning rather than sequential temporal modeling. Tree-based models such as Random Forest are particularly effective in handling nonlinear relationships and interactions between multiple market-related variables, which are common in agricultural pricing scenarios.

The comparatively lower performance of LSTM indicates that long-term temporal dependencies may be less dominant in this dataset, or that the available data does not provide sufficient sequential patterns for deep learning models to fully exploit.

From a practical perspective, this module enables farmers to identify favorable selling periods and anticipate market trends. By integrating predictive insights with decision-support mechanisms, the price forecasting component enhances the economic dimension of the AloeGreen platform, supporting more informed and profit-oriented agricultural decision-making.

D. Fertilizer Recommendation and Nutrient Management Module

The fertilizer recommendation module was evaluated based on its ability to generate accurate nutrient-related outputs under varying soil conditions. Comparative results are presented in Table VI.

TABLE VI. FERTILIZER RECOMMENDATION MODEL PERFORMANCE

Model	MAE	RMSE
Linear Regression	0.061728	0.094674
Random Forest	0.060324	0.092321
LSTM	0.096262	0.118461

The fertilizer recommendation module demonstrated strong overall performance, with Random Forest achieving the lowest error values across both MAE and RMSE metrics. This indicates that tree-based models are well-suited for capturing nonlinear relationships between soil properties and nutrient requirements.

Linear Regression also showed competitive performance, suggesting that some nutrient relationships remain relatively linear. In contrast, the LSTM model exhibited higher error values, indicating that sequential dependencies are less relevant in this context and that structured feature-based modeling is more appropriate for nutrient prediction tasks.

Unlike purely predictive modules, this component adopts a hybrid approach by integrating rule-based fertilizer calculation logic with data-driven insights. This enables the system to generate context-aware and practically applicable fertilizer recommendations tailored to specific soil conditions.

In addition to generating recommendations, the system maintains a history of fertilizer usage and recommendations over time. This historical tracking supports traceability, allows users to monitor nutrient management patterns, and provides a foundation for future adaptive improvements in recommendation strategies.

E. Overall Discussion

Taken together, the results demonstrate that AloeGreen functions effectively as a multi-component decision-support system for Aloe vera cultivation. A consistent pattern across the evaluated modules is the strong performance of tree-based machine learning models for structured agricultural and market data. This indicates that ensemble methods such as Random Forest and XGBoost provide a robust balance between predictive accuracy, computational efficiency, and practical applicability in real-world farming environments.

A key strength of the proposed system lies in its integrated design, where multiple data modalities—including real-time IoT sensor data, historical tabular datasets, image data, and market records—are combined within a unified analytical framework. Each module addresses a distinct aspect of agricultural decision-making: yield estimation supports production planning, environmental forecasting enables future-aware insights, disease detection facilitates early diagnosis, price forecasting enhances market decision-making, and fertilizer recommendation improves nutrient management. The integration of these components significantly enhances the overall practical value of the platform compared to existing fragmented solutions.

Furthermore, the system introduces a forecast-informed yield prediction approach, where short-term environmental forecasts are incorporated into the yield estimation pipeline. This enables forward-looking decision support rather than relying solely on historical data. In addition, the use of domain-specific agronomic features, such as water stress and heat stress indices, improves the model’s ability to capture plant–environment interactions relevant to Aloe vera cultivation.

Beyond predictive capabilities, AloeGreen incorporates practical system-level functionalities, including disease alert dissemination to nearby farmers and fertilizer recommendation history tracking. These features extend the platform from a purely analytical tool to a collaborative and traceable decision-support system, improving real-world usability and impact.

Despite these strengths, certain limitations remain. In particular, environmental forecasting performance varies across variables, with rainfall prediction showing lower accuracy due to its inherently stochastic and complex nature. Additionally, the effectiveness of some modules depends on the availability and diversity of high-quality data. These limitations highlight opportunities for future improvements, such as incorporating additional meteorological data sources, expanding datasets across different regions, and exploring advanced modeling techniques.

Overall, the findings confirm that integrating real-time IoT sensing with multi-model machine learning analytics in a unified platform can significantly enhance agricultural decision-making. AloeGreen demonstrates strong potential as a scalable and practical smart agriculture solution for Aloe vera cultivation, while also providing a foundation for future adaptive and data-driven farming systems.

V. FUTURE WORK

Future work will focus on improving the scalability, accuracy, and practical applicability of the AloeGreen platform. Expanding the dataset to include diverse environmental conditions, cultivation practices, and geographic regions will enhance model generalization and robustness across different farming contexts.

The environmental forecasting component can be further improved, particularly for rainfall prediction, by incorporating additional meteorological variables and external weather forecasting services. The use of advanced time-series models and hybrid forecasting approaches may also help better capture the stochastic nature of rainfall patterns.

The disease detection module can be enhanced by utilizing larger and more diverse image datasets, including real-world

field images with varying lighting conditions and backgrounds. In addition, lightweight deep learning models can be explored to enable real-time or on-device inference, improving system responsiveness in field environments.

Integration of edge computing capabilities represents another important direction, allowing partial data processing and inference to be performed directly on IoT devices. This can reduce dependency on continuous internet connectivity and enable faster, low-latency decision-making in remote agricultural settings.

For the fertilizer recommendation module, future work may include the incorporation of adaptive learning mechanisms that utilize historical recommendation outcomes to continuously refine nutrient management strategies. This would support the development of a more personalized and farm-specific recommendation system over time.

Furthermore, the platform can be extended by integrating additional decision-support functionalities such as irrigation scheduling, pest monitoring, and crop health trend analysis. Enhancements to the mobile application, including improved user interface design and accessibility features, can further increase usability and adoption among farmers.

VI. CONCLUSION

This paper presented AloeGreen, an integrated AI–IoT-based smart agriculture system designed to support Aloe vera cultivation through a unified and data-driven decision-support framework. The proposed system combines yield prediction, environmental forecasting, disease detection, fertilizer recommendation, and price forecasting within a single architecture, enabling comprehensive support across the agricultural lifecycle.

A key contribution of this study is the development of a forecast-informed yield prediction approach, where short-term environmental forecasts are incorporated into the yield estimation pipeline to enable future-aware agricultural decision-making. In addition, the use of domain-specific agronomic features, such as water stress and heat stress indices, enhances the system’s ability to capture plant–environment interactions relevant to Aloe vera growth. The integration of real-time IoT sensing further enables continuous data collection, supporting adaptive model refinement and farm-specific intelligence over time.

Experimental results demonstrate that tree-based machine learning models achieve strong predictive performance for structured agricultural and market data, while the system effectively supports practical tasks such as disease diagnosis, nutrient management, and market timing. Although certain challenges remain, particularly in forecasting highly variable environmental factors such as rainfall, the overall system provides reliable and actionable insights for real-world farming conditions.

Beyond predictive analytics, AloeGreen incorporates practical features such as disease alert dissemination and fertilizer recommendation history tracking, extending its functionality to collaborative and traceable decision support. These capabilities enhance its usability and potential impact in real agricultural environments.

Overall, the findings indicate that integrating IoT-based sensing with intelligent analytics in a unified platform can significantly enhance agricultural productivity, reduce

uncertainty, and support sustainable farming practices. AloeGreen demonstrates strong potential as a scalable and practical smart agriculture solution, while also providing a foundation for future adaptive and data-driven farming systems.

REFERENCES

- [1] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artificial Intelligence in Agriculture*, vol. 2, pp. 1–12, 2019, doi: 10.1016/j.aiia.2019.05.004.
- [2] M. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 177, Art. no. 105709, 2020, doi: 10.1016/j.compag.2020.105709.
- [3] J. Šuljug, J. Spišić, K. Grgić, and D. Žagar, "A comparative study of machine learning models for predicting meteorological data in agricultural applications," *Electronics*, vol. 13, no. 16, Art. no. 3284, 2024, doi: 10.3390/electronics13163284.
- [4] K. Jhajharia, P. Mathur, S. Jain, and S. Nijhawan, "Crop yield prediction using machine learning and deep learning techniques," *Procedia Computer Science*, vol. 218, pp. 406–417, 2023, doi: 10.1016/j.procs.2023.01.023.
- [5] OpenWeather, "Weather API Documentation," 2024. [Online]. Available: <https://openweathermap.org/api>. [Accessed: Apr. 3, 2026].
- [6] National Oceanic and Atmospheric Administration (NOAA), "Climate Data Online," 2024. [Online]. Available: <https://www.ncdc.noaa.gov/cdo-web/>. [Accessed: Apr. 3, 2026].
- [7] N.-Q. Tran, A. Felipe, T. N. Ngoc, T. Huynh, Q. Tran, A. Tang, and T. Nguyen, "Predicting agricultural commodities prices with machine learning: A review of current research," arXiv:2310.18646, 2023.
- [8] Z. Chen, H. S. Goh, K. L. Sin, K. Lim, N. K. H. Chung, and X. Y. Liew, "Automated agriculture commodity price prediction system with machine learning techniques," *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 2, 2021.
- [9] R. K. Paul, M. Yeasin, P. Kumar, P. Kumar, M. Balasubramanian, H. S. Roy, A. K. Paul, and A. Gupta, "Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India," *PLOS ONE*, vol. 17, no. 1, Art. no. e0261230, 2022, doi: 10.1371/journal.pone.0261230.
- [10] H. M. B. P. Ranaweera, R. M. G. H. N. Rathnayake, and A. S. G. J. K. Ananda, "Crop price prediction using machine learning approaches: Reference to the Sri Lankan vegetable market," *Journal of Management Matters*, vol. 10, no. 1, pp. 19–34, 2023. [Online]. Available: <https://journals.sjp.ac.lk>
- [11] N. Qomariyah, S. D. Putra, D. A. Afifah, A. R. Supriyatna, and Z. Zuriati, "Applying random forest for optimal crop selection to enhance agricultural decision-making," in *Proc. Int. Conf. Advanced Engineering (ICAE)*, 2024, doi: 10.2991/978-94-6463-620-8_6.
- [12] H. Koormala, K. Kumar, and K. Reddy C, "Predictive modelling of crop yield using XGBoost: An advanced machine learning technique," in *Proc. IEEE Int. Conf. Smart Comput. Syst. Appl. (ICSCSA)*, 2025, pp. 1318–1325, doi: 10.1109/ICSCSA66339.2025.11170806.
- [13] Y. Perwej, H. Rizvi, F. Siddiqui, N. Akhtar, and S. Ahmad, "Crop yield forecasting through an attention-driven LSTM deep learning model," *JETIR*, vol. 12, pp. 891–898, Nov. 2025, doi: 10.6084/m9.jetir.JETIR2511510.
- [14] I. Attri, L. Awasthi, T. Sharma, and P. Rathee, "A review of deep learning techniques used in agriculture," *Ecol. Inform.*, vol. 77, p. 102217, Jul. 2023, doi: 10.1016/j.ecoinf.2023.102217.
- [15] L. Mintarya, J. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: A systematic literature review," *Procedia Comput. Sci.*, vol. 216, pp. 96–102, Jan. 2023, doi: 10.1016/j.procs.2022.12.115.
- [16] F. Delli, S. Nurdianti, A. Buono, and E. Santosa, "Agricultural price prediction models: A systematic literature review," in *Proc. Adv. Nat. Sci. Int. Conf. (ANSC)*, 2021, doi: 10.46254/AN11.20210532.
- [17] E. L. N. D. Madubhashini, "Predicting Vegetable Prices in Sri Lanka Using Machine Learning Techniques," M.Sc. thesis, Univ. Colombo School Comput., Colombo, Sri Lanka, 2021.
- [18] G. S. Nagaraja, A. B. Soppimath, T. Soumya, and A. Abhinith, "IoT based smart agriculture management system," in *Proc. 4th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. (CSITSS)*, India, Dec. 2019, pp. 1–5, doi: 10.1109/CSITSS47250.2019.9031025
- [19] B. Kashyap and R. Kumar, "Sensing methodologies in agriculture for soil moisture and nutrient monitoring," *IEEE Access*, vol. 9, pp. 8601–8618, Jan. 2021, doi: 10.1109/ACCESS.2021.3052478.
- [20] [20] Food and Agriculture Organization (FAO), "FAOSTAT," 2024. [Online]. Available: <https://www.fao.org/faostat>.
- [21] "S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big data in smart farming – A review," *Agricultural Systems*, vol. 153, pp. 69–80, 2017.
- [22] N. S. K. Kumar, P. R. K. Reddy, and M. V. R. Prasad, "Weather-based crop yield prediction using machine learning techniques," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 5, pp. 1234–1239, Mar. 2020.
- [23] S. R. Patil, P. P. Patil, and S. B. Patil, "Soil nutrient analysis and crop recommendation using machine learning," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 7, no. 3, pp. 456–461, 2021
- [24] A. Kumar, S. Kumar, and R. Kumar, "Fertilizer recommendation system using machine learning techniques," *Int. J. Eng. Res. Technol. (IJERT)*, vol. 9, no. 5, pp. 123–128, May 2020.
- [25] C. Musanase, A. Vodacek, D. Hanyurwimfura, A. Uwitonze, and I. Kabandana, "Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices," *Agriculture*, vol. 13, no. 11, p. 2141, 2023, doi: 10.3390/agriculture13112141.
- [26] N. Shelar, S. Shinde, S. Sawant, S. Dhumal, and K. Fakir, "Plant disease detection using CNN," in *Proc. International Conference on Electronics and Telecommunication Engineering*, Navi Mumbai, India, 2022.
- [27] K. Shivaprasad and A. Wadhawan, "Deep learning-based plant leaf disease detection," in *Proc. 7th Int. Conf. Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2023, pp. 360–365, doi: 10.1109/ICICCS56967.2023.10142857.
- [28] A. Dolatabadian, T. X. Neik, M. F. Danilevicz, S. R. Upadhyaya, J. Batley, and D. Edwards, "Image-based crop disease detection using machine learning," *Plant Pathology*, 2024, doi: 10.1111/ppa.14006.