SMART AGRICULTURE SUPPORT SYSTEM FOR ALOE VERA

Project ID: 2025-26J-166

Project Proposal Report

Megasooriya G.M.M.A.E.

BSc (Hons) in Information Technology Specializing in Information Technology

Department of Computer Science

Sri Lanka Institute of Information Technology
Sri Lanka

August 2025

SMART AGRICULTURE SUPPORT SYSTEM FOR ALOE VERA

Project ID: 2025-26J-166

Project Proposal Report

Megasooriya G.M.M.A.E. : IT22337962

Supervised by Ms. Jenny Krishara

Co-supervised by Dr. Dinuka Wijendra

BSc (Hons) in Information Technology Specializing in Information Technology

Department of Computer Science

Sri Lanka Institute of Information Technology
Sri Lanka

August 2025

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Megasooriya G.M.M.A.E	IT22337962	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

..........Jenny.......... ...2025/08/29.......

Signature of the Supervisor

Ms. Jenny Krishara

Date

....2025/08/29..........

Signature of the co-Supervisor

Dr. Dinuka Wijendra

Date

ACKNOWLEDGEMENT

I would like to extend my sincere gratitude to my supervisor, Ms. Jenny Krishara, and co-supervisor, Dr. Dinuka Wijendra, from the Faculty of Computing, for their valuable advice and assistance at the beginning of my research project. To increase my focus and the quality of my research their insightful feedback and constructive criticism have been invaluable.

Furthermore, I would like to acknowledge the contribution of my fellow team members for their support and collaboration.

Finally, I want to express my gratitude to everyone who volunteered to assist in this research. Sharing their experiences and viewpoints has been essential in assisting me in developing a deeper understanding of the topic.

ABSTRACT

Agriculture plays a vital role in ensuring food security, yet farmers face significant challenges in predicting yields due to uncertain weather conditions, soil variability, and lack of reliable decision-support tools. This research proposes the development of an IoT-enabled yield prediction and crop suitability component aimed at improving forecasting accuracy and assisting farmers with timely cultivation decisions. The system design integrates IoT devices for capturing real-time field data, including soil pH, moisture, temperature, and humidity, with historical datasets of environmental records and crop yields. Preprocessing techniques will be applied to clean and normalize the data, after which predictive models will be trained to estimate yields under varying conditions. When farmers use the mobile application, the current IoT readings will be processed by the model to forecast yield and determine whether a specific crop, such as Aloe vera, is suitable for cultivation during the given period. The anticipated outcome of the system is accurate yield forecasting and actionable recommendations. Pilot evaluations will measure prediction accuracy, system usability, and farmer satisfaction. The conclusions are expected to demonstrate that integrating IoT sensing with predictive analytics can reduce farming risks, support sustainable cultivation practices, and empower small-scale farmers through data-driven decision-making. Recommendations will emphasize further scaling of the model to additional crops and integration with broader agricultural advisory systems.

Keywords: IoT-based Agriculture, yield prediction, crop suitability, precision farming, recommendation

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI - Artificial Intelligence

CNN - Convolutional Neural Network

DL – Deep Learning

DNN – Deep Neural Network

IoT – Internet of Things

LSTM - Long Short-Term Memory

ML - Machine Learning

NLP - Natural Language Processing

UAV - Unmanned Aerial Vehicles

1. INTRODUCTION

Agriculture remains the backbone of food security and rural livelihoods, particularly in developing nations. In this context, crop yield prediction has emerged as a critical tool for optimizing farming practices, enabling farmers to make informed decisions about resource allocation, harvesting schedules, and market planning [4], [6]. Accurate forecasting of crop yields helps mitigate the risks associated with climate variability and resource scarcity, while also supporting national-level strategies for food supply and sustainability [16].

Much of the existing research in this domain has focused on staple crops such as rice, maize, and wheat [3], [18], [24], [28]. Machine learning and deep learning models applied to these crops have demonstrated significant improvements in predictive accuracy, often relying on large datasets sourced from Unmanned Aerial Vehicles (UAV) imagery, remote sensing, or satellite observations [25], [30]. For instance, [3] developed a deep neural network that outperformed traditional methods for maize yield prediction, while [24] showed how UAV-based monitoring could generate highly accurate crop yield maps. Although these studies provide strong evidence for the value of Artificial Intelligence (AI) in agriculture, their findings are largely restricted to staple crops, leaving non-staple but commercially valuable medicinal crops like Aloe vera underexplored.

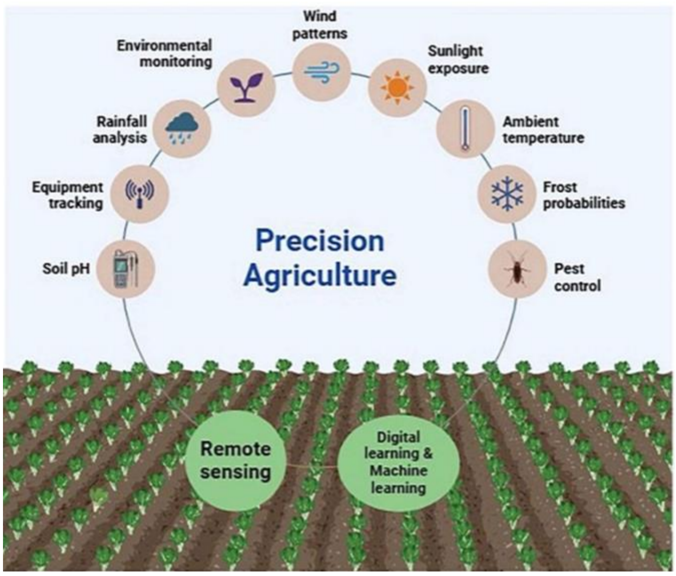

Figure 1.1: Precision Agriculture [6]

Aloe vera is widely cultivated for its medicinal, cosmetic, and nutritional applications, making it an economically important crop in many parts of Asia and beyond [1], [2]. However, most studies related to Aloe vera cultivation have emphasized agronomic factors such as soil type [1] and sunlight exposure [2] or explored applications in controlled greenhouse environments and disease detection [14], [15]. While these works highlight important influences on crop performance, they stop short of building predictive frameworks that can forecast yield outcomes. Moreover, the absence of publicly available datasets for Aloe vera poses a significant barrier to advancing data-driven agriculture for this crop.

Another limitation in the current literature is the lack of emphasis on farmer usability, especially in rural areas where Aloe vera is predominantly grown. Advanced AI systems often depend on high-resolution data sources and computationally expensive models [24], [25], [30], which may not be accessible or practical for smallholder farmers. Existing research has not sufficiently addressed interpretability and farmer trust, which are crucial for the adoption of predictive tools in low-resource contexts [23], [27].

The proposed study aims to address these challenges by developing a predictive framework for Aloe vera yield that integrates agronomic insights [1], [2] with machine learning and deep learning models [3], [18], [29]. Unlike prior studies focused on staple crops, this research emphasizes a non-staple medicinal crop, considers rural usability constraints, and incorporates interpretability features to ensure predictions are understandable and actionable for farmers. By filling these gaps, the work contributes to extending smart agriculture practices beyond cereals, enabling data-driven decision support for Aloe vera cultivation and enhancing the livelihoods of rural farming communities.

1.1. Background & Literature Survey

1.1.1. Background

Agriculture has always been more than just an occupation, it is a way of life, deeply connected to the wellbeing of families and the stability of entire communities. Yet, farming today looks very different from how it did in the past. Farmers everywhere, and especially in countries like Sri Lanka, face mounting challenges. Changing weather patterns make it difficult to plan ahead, unexpected crop failures put livelihoods at risk, and the rising cost of inputs often leaves small farmers struggling to make ends meet. At the same time, global markets continue to fluctuate, leaving farmers uncertain about the value of their produce even after a successful harvest.

Traditionally, most of these decisions when to plant, how to manage the soil, what to apply, and when to sell have relied on the farmer's own knowledge, experience, and community advice. While this wisdom is invaluable, it can sometimes fall short in the face of today's rapid environmental and economic changes. This is where modern technology can play a transformative role. By combining real-time monitoring tools with advanced data analysis, farmers can be given timely, reliable insights that reduce uncertainty and support better decision-making.

Recent innovations in IoT devices, artificial intelligence, and predictive analytics are paving the way for farming to become more precise and adaptive. These technologies allow for continuous monitoring of environmental conditions, intelligent analysis of risks and opportunities, and personalized guidance tailored to each farmer's unique context. Instead of reacting to problems after they occur, farmers can anticipate challenges, improve efficiency, and plan ahead with confidence.

In developing regions, where access to expert knowledge and resources is often limited, such systems can be particularly valuable. A farmer-friendly decision support platform that integrates data from the field with wider market and environmental trends has the potential to make agriculture more sustainable, resilient, and profitable. By presenting information in a simple and accessible way through mobile applications, even small-scale farmers can benefit from the power of technology without needing specialized training.

This research aims to contribute to this transformation by building an integrated support system that not only helps farmers protect their crops and optimize their resources but also empowers them with insights to make smarter choices for the future. In doing so, it seeks to bridge the gap between traditional farming wisdom and modern technological advancements, ensuring that farmers are better equipped to thrive in an increasingly uncertain world.

1.1.2. Literature Review

Crop yield prediction has become one of the most widely studied topics in precision agriculture due to its direct connection with food security and sustainable farming practices. Researchers have investigated a variety of factors from environmental conditions and soil health to hybrid deep learning models in order to improve prediction accuracy [4], [5]. Recent studies emphasize that while classical approaches such as regression and ensemble methods remain valuable, deep learning has shown stronger potential in extracting complex nonlinear patterns hidden within agricultural data [3], [18], [22]. In addition, specific attention has been given to medicinal crops such as Aloe vera, where optimized environmental management and modern prediction models can significantly enhance both yield and quality [1], [2], [15].

Early work on Aloe vera demonstrated that soil type strongly influences its growth and gel production. [1] compared seven different soil types and concluded that calcareous and acid soils yielded the best performance, particularly in gel weight and leaf biomass. Similarly, [2] examined the impact of sunlight exposure and found that plants receiving 5–7 hours of daily sunlight exhibited significantly higher biomass than those with lower exposure. Later studies expanded the scope by considering greenhouse-integrated photovoltaics, where [15] showed that photovoltaic-integrated systems could maintain adequate growth while providing sustainable energy sources. Collectively, these works highlight how environmental and soil management remain crucial foundations for Aloe vera yield, before even introducing machine learning methods.

In parallel, disease detection and crop health monitoring have emerged as complementary avenues for improving Aloe vera productivity. [14] developed an Edge

AI-powered architecture specifically for Aloe vera disease detection, demonstrating that CNN models deployed on lightweight devices can provide real-time monitoring and reduce leaf losses due to infections. This aligns with broader smart agriculture frameworks, where IoT and AI are synergized to provide predictive and prescriptive insights for both disease prevention and yield improvement [8], [12].

The use of machine learning for agricultural yield forecasting has expanded rapidly in the last decade. Systematic reviews such as [4] and [16] catalogued a wide range of ML techniques applied to crops worldwide, including regression trees, SVM, random forests, and gradient boosting. These works highlight that ensemble learning often provides a robust baseline, but hybrid frameworks that integrate multiple feature sources achieve superior performance. Similarly, [6] provided a dedicated review on precision agriculture, noting that the combination of remote sensing data, big data analytics, and ML has created opportunities to not only predict yield but also optimize resources like fertilizer and irrigation.

Among specific ML algorithms, random forests and boosting methods continue to dominate due to their interpretability and high performance on tabular agricultural datasets [9], [10]. [9] demonstrated that Random Forest achieved high predictive accuracy for yield forecasting in Indian crop datasets, outperforming linear regression and naïve models. [11] extended these experiments with a comparative study, concluding that ensemble methods such as XGBoost provided higher R^2 values and reduced error rates compared to simpler models. Similarly, [10] discussed how combining regression and ensemble algorithms within smart agriculture pipelines enhanced resource efficiency while maintaining accurate predictions.

Deep learning has further transformed yield prediction research. [3] proposed a deep neural network trained on both genotype and environmental data for maize, which significantly outperformed classical models, reducing RMSE and achieving a validation correlation coefficient above 0.81. [18] further advanced this by introducing hybrid deep learning architectures that integrated Long Short Term Memory (LSTM)s with Convolutional Neural Networks (CNNs), enabling spatiotemporal analysis of agricultural datasets. [26] and [28] showed that spatiotemporal deep learning networks could capture seasonal patterns and climate variations, achieving superior accuracy in rice and maize yield prediction compared to shallow learning methods. [29] introduced

CROPUP, a deep learning recommendation system that integrated CNN and XGBoost to provide real-time suggestions to farmers, reporting accuracy levels above 99%. These studies collectively demonstrate the significant shift toward deep learning (DL) models in agriculture, owing to their ability to uncover complex nonlinearities that classical models struggle with.

Remote sensing has also been tightly integrated into yield prediction pipelines. [30] reported that vegetation indices such as NDVI provided strong predictors for maize yield when combined with CNN-based models. [24] demonstrated that UAV imagery, when processed through deep learning, can generate automated yield maps with higher spatial resolution than traditional survey methods. [25] proposed a deep Gaussian process integrating satellite data and ground-based climate data, demonstrating superior predictive capabilities for large-scale monitoring. These advances in remote sensing are particularly relevant for Aloe vera, which often grows in semi-arid and resource-constrained environments where climate and soil heterogeneity strongly affect yield outcomes [1], [2], [13].

IoT-based and hybrid AI approaches have gained traction in recent years. [8] developed a soil-moisture and nutrient-sensing IoT system, which fed into ML models to predict crop yield with higher precision than using climate data alone. Similarly, [12] and [13] explored hybrid deep learning frameworks where CNN-based feature extraction was combined with IoT-collected soil and environmental data. Their findings suggest that multi-source integration consistently leads to improved predictive accuracy. [20] also noted that integrating IoT with ML enhances real-time adaptability, making these systems highly applicable for resource-limited farming regions.

Another emerging dimension is explainable AI (XAI) and model interpretability. [27] applied Bayesian neural networks to yield prediction under extreme weather conditions, providing uncertainty estimates along with predictions. This is essential for farmer adoption, as black-box predictions alone may not be sufficient for decision-making. Similarly, [23] highlighted the importance of extreme climate events, showing how models must be sensitive to rare but impactful weather fluctuations. These works underline that yield prediction models must balance accuracy with transparency, especially in regions like South Asia where adoption by smallholder farmers depends on trust and interpretability.

Beyond global staple crops, Aloe vera has begun to attract attention in prediction-focused research due to its medicinal and economic importance. The combination of controlled environment parameters (soil type, sunlight, water management) [1], [2], IoT-based monitoring [8], and ML/DL models [10], [12], [18] provide a strong foundation for designing Aloe vera yield prediction frameworks. With advances in edge AI [14] and greenhouse-integrated energy systems [15], Aloe vera presents itself as a case study for applying next-generation smart agriculture practices to non-staple but high-value crops.

In summary, existing literature demonstrates that crop yield prediction has evolved from simple regression-based approaches [4], [9] to advanced hybrid deep learning frameworks integrating IoT, UAV imagery, and remote sensing [18], [24], [30]. Aloe vera studies emphasize the critical role of soil, sunlight, and disease management [1], [2], [14], [15]. When combined with ML/DL models that have proven effective in major crops [3], [18], [22], Aloe vera yield prediction systems can achieve both high accuracy and practical applicability. Future directions will likely focus on integrating explainable AI, low-cost IoT devices, and transfer learning models for localized adaptation, ensuring that predictive frameworks benefit both large-scale and smallholder farmers [6], [20], [27].

1.2. Research Gap

The reviewed studies collectively emphasize the importance of accurate crop yield prediction for enhancing food security and resource management. Different approaches, ranging from classical regression and ensemble models to advanced deep learning and IoT-driven architectures, have been applied across diverse crops. However, several research gaps become evident when analyzing these works in the specific context of Aloe vera yield prediction.

Research A focuses on soil and environmental conditions for Aloe vera cultivation. [1], [2] demonstrated that soil type and sunlight exposure significantly influence Aloe vera growth and gel yield. While these studies provide important agronomic insights, they rely on traditional experimentation and do not employ predictive modeling. Thus, the challenge remains in scaling these findings into intelligent frameworks that can automatically forecast yield under varying soil and climate conditions.

Research B addresses technological integration through greenhouse-based and energy-efficient Aloe vera systems. [15] examined greenhouse photovoltaics and their impact on Aloe vera growth, whereas [14] proposed Edge AI for Aloe vera disease detection. These efforts reveal opportunities for integrating smart technologies, but their scope is either infrastructure optimization or disease diagnosis. The actual prediction of Aloe vera yield using ML/DL is largely unaddressed.

Research C represents broad systematic reviews on yield forecasting with machine learning. [4], [16], and [6] all highlighted the effectiveness of ensemble models and deep learning for crop prediction in general. However, none of these reviews explicitly include Aloe vera as a case study, nor do they address non-staple medicinal crops. The over-representation of major cereals in datasets leaves a gap for crops like Aloe vera that have high commercial but less-explored yield forecasting potential.

Research D emphasizes IoT-driven prediction. [8] integrated soil moisture and nutrient sensors with ML algorithms to enhance yield prediction, while [12] combined IoT-collected features with CNN models for hybrid prediction. These studies highlight the strength of sensor-based approaches, but they are applied primarily to cereals and pulses. For Aloe vera, where soil-water relationships and nutrient sensitivity are unique [1], [2], IoT-based approaches remain underexplored.

Research E advances deep learning for agricultural yield prediction. [3] proposed DNNs for maize yield prediction, and [18] introduced spatiotemporal hybrid models integrating CNNs and LSTMs. Similar methods by [28] and [29] confirmed that DL-based pipelines outperform classical ML. Despite their strong performance, these models rely on datasets with robust availability (e.g., maize, rice, wheat). Aloe vera lacks large, publicly available yield datasets, creating a gap for DL application.

Research F focuses on remote sensing and UAV-based yield prediction. [30] reported vegetation indices as strong predictors of maize yield, and [24] demonstrated UAV imagery integrated with ML for automated yield mapping. [25] extended this to deep Gaussian processes, showing state-of-the-art predictive capabilities. However, these frameworks require high-resolution aerial and satellite datasets that are not readily available for Aloe vera plantations, particularly in smallholder contexts.

Research G reflects on model interpretability and explainability. [27] employed Bayesian neural networks to incorporate uncertainty in yield prediction, while [23] examined climate extremes in model design. These advances underscore the importance of trustworthy AI in agriculture. Yet, for Aloe vera yield forecasting, no work has explicitly addressed explainability, an important factor for farmer adoption in developing countries.

Taken together, the reviewed studies reveal three main research gaps. First, Aloe vera has been studied primarily through agronomic experiments [1], [2], [15] but not through predictive modeling frameworks. Second, while ML and DL methods demonstrate strong performance in cereals and major crops [3], [18], [22], their application to Aloe vera remains absent due to dataset unavailability. Third, IoT-based and hybrid AI systems [8], [12], [20] have shown promise in other contexts, but their potential to capture Aloe vera's unique environmental sensitivities is yet to be realized.

Therefore, this research aims to bridge these gaps by designing a predictive framework for Aloe vera yield using machine learning and deep learning approaches, supported by IoT-driven environmental data. By integrating agronomic insights [1], [2], AI-based predictive modeling [3], [18], [29], and considerations for interpretability [23], [27], the proposed study will advance Aloe vera yield prediction and provide a novel extension of smart agriculture to non-staple medicinal crops.

Table 1.2: Research gap table

Research Gap	**Research A [1]**	**Research B [2]**	**Research C [14]**	**Research D [15]**	**Research E [3]**	**Research F [24]**	**Proposed System**
Specific focus on Aloe vera	✓	✓	✓	✓	✗	✗	✓
General crop yield prediction	✗	✗	✗	✗	✓	✓	✓
Consideration of environmental factors (soil, sunlight, climate)	✓	✓	✗	✓	✓	✓	✓
Emphasis on non-staple/medicinal crops	✓	✓	✓	✓	✗	✗	✓
Practical field-level application (data collection, experiments)	✓	✓	✗	✓	✗	✓	✓
Inclusion of farmer-oriented usability (trust, interpretability)	✗	✗	✓	✗	✗	✗	✓

1.3. Research Problem

Most existing research in crop yield prediction has focused on staple crops such as maize, rice, and wheat [3], [18], [24], [28]. Deep learning and machine learning models applied to these crops have achieved high accuracy and robust performance, often utilizing large datasets from UAV imagery, remote sensing, or national repositories [25], [30]. However, medicinal, and non-staple crops like Aloe vera have received far less attention. The few studies related to Aloe vera cultivation, such as those on soil type [1] and sunlight exposure [2], or greenhouse and disease detection systems [14], [15], do not include predictive yield modeling. As a result, there is a significant lack of dedicated datasets for Aloe vera yield forecasting, limiting the potential for data-driven agricultural decision-making in this domain.

Another limitation in the current literature is that predictive studies rarely consider the needs of farmers in rural or resource-constrained settings. Most frameworks rely on high-resolution UAV imagery [24], [25] or cloud-based processing [22], [30], which are often inaccessible to smallholder farmers. While IoT-based systems [8], [12] and advanced ML reviews [4], [6], [16] emphasize technical innovation, they provide little attention to usability, interpretability, or deployment in rural contexts where Aloe vera farming is common. This is a critical barrier, as farmer adoption of predictive tools depends not only on model accuracy but also on practical accessibility and trust [23], [27].

Therefore, the research problem lies in the absence of Aloe vera-specific predictive datasets, the overemphasis on staple crops in yield prediction studies, and the lack of farmer-oriented, interpretable systems designed for rural agricultural contexts. Addressing this gap requires developing a predictive framework that integrates agronomic factors specific to Aloe vera [1], [2], leverages machine learning and deep learning methods [3], [18], [29], and emphasizes usability and adaptability for farmers in rural regions [23], [27].

2. RESEARCH OBJECTIVES

2.1. Research Question

How to enhance Aloe vera farming efficiency by developing an IoT-powered yield prediction system using advanced Machine Learning and Deep Learning models trained on soil and environmental parameters?

2.2. Main Objective

The main objective of this research is to develop a decision support system that assists farmers in making data-driven agricultural decisions by integrating IoT-based monitoring, predictive analysis, and personalized recommendations.

2.3. Sub Objectives

a. Development of Aloe vera yield prediction models.

Develop and train machine learning and deep learning models specifically designed to predict Aloe vera yield using soil and climatic variables (pH, rainfall, sunlight, temperature, humidity).

b. Integration of IoT Sensors for real-time data collection.

Deploy IoT-based soil and environmental sensors to collect continuous real-time data from Aloe vera cultivation fields. Parameters such as soil pH, soil moisture, rainfall, and temperature will be transmitted to the predictive system for model input.

c. Optimization of prediction accuracy through Hybrid Modeling.

Enhance prediction accuracy by implementing hybrid deep learning models (CNN-LSTM with Empirical Mode Decomposition), parameter fine-tuning, and cross-validation. This ensures the models capture both temporal (time-series) and spatial (environmental) relationships effectively.

d. Resource-efficient model deployment for field use.

Adapt models for deployment in resource-constrained environments by reducing computational complexity (e.g., pruning, lightweight architectures). This enables predictions to be accessible on IoT edge devices or farmer dashboards without compromising performance.

e. Validation of the proposed system through case studies.

Validate the proposed system using Aloe vera-specific field data and case studies. Compare predicted yields with actual yields under varying soil and environmental conditions to ensure the practical applicability of the system in real farming environments.

3. METHODOLOGY

3.1. Overall System Methodology

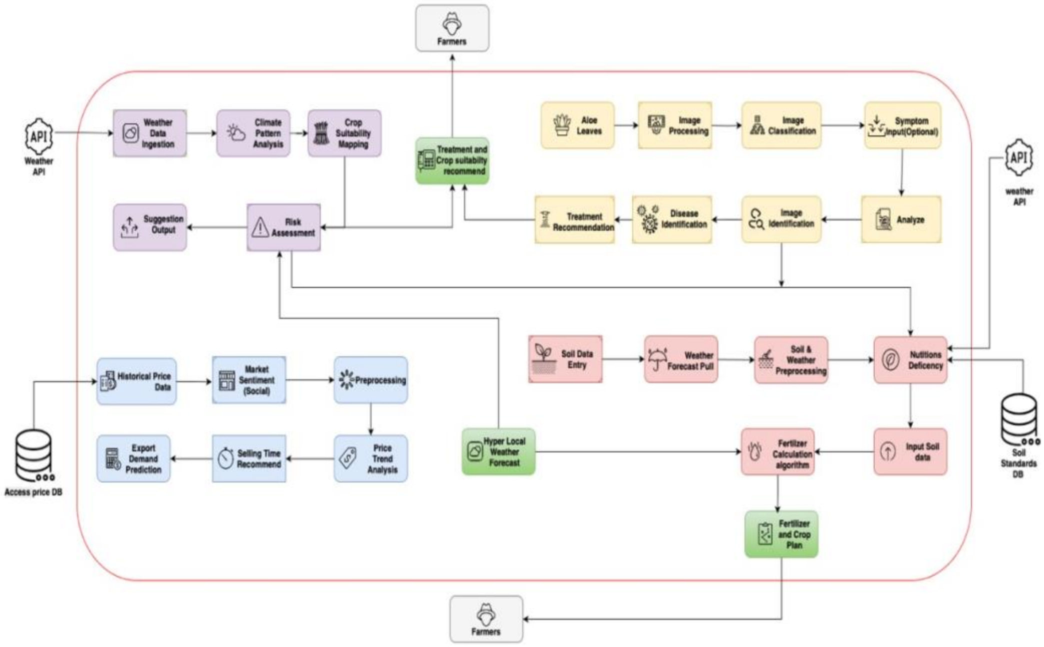

Figure 3.1: Overall System Diagram

This illustrates the overall high-level diagram of the proposed Aloe Vera disease detection component. Initially, the farmer accesses the mobile application. The farmer is provided with an option to either capture a new leaf image or upload an existing image of the Aloe Vera plant using a cross-platform mobile application developed with React Native. The captured or uploaded image is preprocessed and then converted into a Base64 stream before being sent to the Flask server via the mobile app. The server applies machine learning and image processing techniques to analyze the leaf for visible disease symptoms. In addition to image analysis, the farmer can also input symptoms through text or voice in Sinhala, Tamil, or English. The text input is preprocessed and passed through an NLP model to extract structured symptom information.

3.2. Component Methodology

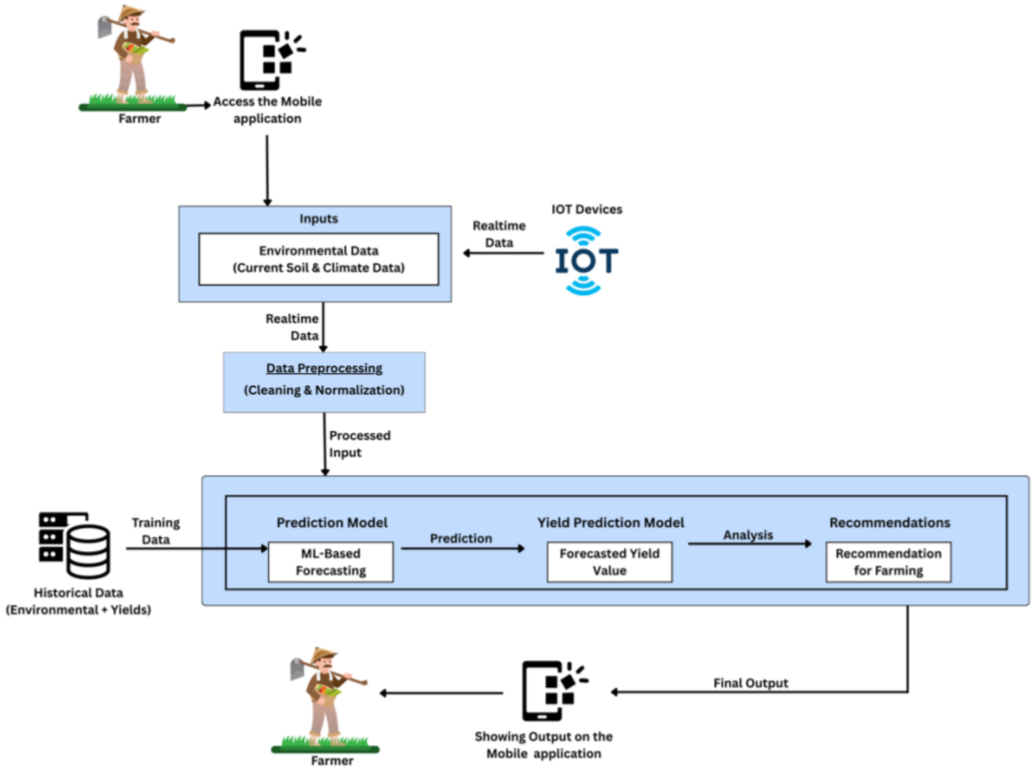

Figure 3.2: Component Diagram

The proposed component, as illustrated in Figure 3.2, will be developed following a design, development, and evaluation approach, integrating IoT-based environmental sensing, data preprocessing, machine learning-driven yield prediction, and a farmer-facing mobile interface into a unified framework. Since training an agricultural prediction model requires both historical and current data, this research will utilize a combination of past environmental records and corresponding crop yield data to build a reliable training dataset. The IoT devices deployed in the field will capture real-time environmental parameters such as soil pH, moisture, temperature, and humidity, ensuring that predictions are grounded in the current conditions of the farmer's land.

The collected sensor data will undergo preprocessing to remove noise, handle missing values, and normalize measurements for consistency with the historical dataset. Once processed, the data will be passed into the trained prediction model to forecast the expected crop yield. This approach allows the system to function not only as a yield estimation tool but also as a forecasting mechanism, enabling farmers to anticipate potential outcomes under current and near-future environmental conditions.

To make the system actionable, the predicted yield will be accompanied by recommendation outputs that provide farmers with crop-specific guidance on cultivation decisions. For example, the system may indicate whether it is suitable to grow Aloe vera during a particular period, based on the prevailing environmental and soil conditions. These recommendations will help farmers decide on the most viable crops for cultivation in their fields, reducing risks and optimizing productivity. All results will be delivered through a farmer-friendly mobile application, ensuring accessibility and ease of use.

The component will be evaluated through pilot testing with local farmers, measuring prediction accuracy, the quality of recommendation outputs, and overall system usability. In addition to quantitative performance metrics, qualitative feedback from farmers will be collected to assess the perceived usefulness and effectiveness of the system in supporting data-driven agricultural decision-making.

3.2.1. Requirement Gathering

Meetings and discussions with agriculture experts will be conducted to:

- Identify the most influential environmental and soil factors affecting Aloe vera yield.
- Understand practical constraints of Aloe vera farming in Sri Lanka and other regions.
- Gather requirements for system usability, such as real-time monitoring, prediction accuracy, and interpretability of results.

3.2.2. Data Gathering

- Historical datasets on Aloe vera yields will be sourced from agricultural research institutions and published studies.

- Field-level IoT sensor data will be collected, including soil pH, soil moisture, rainfall, temperature, sunlight exposure, and humidity.

3.3. Research Constraints

3.3.1. Data availability and quality

- Access to comprehensive historical datasets that include both environmental parameters and actual crop yields may be limited, particularly in Sri Lanka.
- IoT sensor data may contain noise, missing values, or calibration errors, which can affect model training and prediction accuracy.

3.3.2. IOT device limitations

- Deployment of IoT devices in real-world farming environments may face challenges such as hardware malfunctions, battery limitations, and unreliable network connectivity in rural areas.
- Sensor coverage might not fully capture micro-variations across large or uneven fields, potentially limiting accuracy.

3.3.3. Model generalization.

- Models trained on limited or region-specific data may not generalize well to different crops, soil types, or climates.
- Seasonal variations and sudden weather anomalies (e.g., droughts, floods) may reduce the accuracy of yield forecasts.

3.3.4. Infrastructure and resource constraints

- Limited computational resources in rural settings may restrict real-time data processing and on-device inference.
- Internet connectivity issues may hinder real-time communication between IoT devices, the backend server, and the mobile application.

3.3.5. Time constraints

- The research timeline limits the extent of model fine-tuning, testing across multiple greenhouse conditions, and exploration of alternative preprocessing methods.
- The time-intensive process of dataset creation and manual annotation further constrains experimentation.

4. PROJECT REQUIREMENTS

4.1.Functional Requirements

1. IoT Data Acquisition

- The system should capture real-time environmental parameters (e.g., soil pH, moisture, temperature, humidity, rainfall) from field-deployed IoT sensors.
- It should ingest external weather data (current/forecast) to complement on-field readings.

2. Data Input and Management

- Store historical datasets and time stamped IoT readings in a secure database for training, inference, and traceability.

3. Yield Prediction & Forecasting

- The system should generate a yield estimate using current IoT readings and trained historical patterns.
- It should support seasonal/short-term forecasting for the active crop cycle.

4. Crop Suitability Recommendation

- The system should determine whether Aloe vera is suitable to grow in the current period based on prevailing conditions.
- It should present a clear decision (e.g., “Suitable” / “Not suitable”) with a brief reason (e.g., “low soil pH” or “optimal temperature range”).

5. User Interface

- The system should provide a farmer-friendly dashboard to view predictions, suitability, and reasons in simple language with visuals.
- It should allow users to trigger data capture and on-demand prediction from the app.

6. Notifications & Alerts
    - The system should notify farmers when environmental conditions change enough to affect suitability or yield (e.g., rainfall spike, pH drift).
    - It should suggest when to recheck conditions or rerun predictions.

4.2. Non-Functional Requirements

1. Performance

    - The system should generate yield predictions and crop suitability recommendations within a few seconds after IoT data is submitted.
    - It should handle continuous IoT data streams without significant delays.

2. Security

    - The system should ensure that datasets and prediction results are stored securely.
      Only authorized users should have access to view, update, or export the stored data.

3. Usability

    - The application should provide a simple and farmer-friendly interface requiring minimal training.
    - Prediction outputs and quality reports should be displayed in easy-to-understand tables and charts.

4. Reliability

    - The system must provide consistent and accurate predictions based on the input data.
    - It should include error-handling mechanisms for missing, incomplete, or invalid data entries.

5. Maintainability

    - The system should be designed in a modular way to allow regular updates and improvements to prediction models.

- Dataset updates and model retraining should be possible without major downtime.

6. Efficiency

- The system should process input data quickly and return prediction results within seconds.
- It should be optimized to use minimal computational resources, ensuring smooth performance even on low-spec devices.

4.3.Technical Requirements

1. Data Storage and Management

Available Technologies

- MySQL: Widely used relational database, strong consistency, and reliability. But scalability for very large datasets can be limited without optimization.
- PostgreSQL: Advanced relational database supports complex queries, indexing, and analytics. But heavier configuration compared to MySQL.
- MongoDB: NoSQL, flexible schema for unstructured/semi-structured data. But not ideal for structured tabular yield datasets.

2. Machine Learning Model for Yield Prediction

Available Technologies

- Linear/Polynomial Regression: Simple, interpretable, fast for small datasets. But limited in capturing non-linear relationships in agricultural yield.
- Random Forest / Gradient Boosting (XGBoost, LightGBM, CatBoost): Strong performance on structured data, handles missing values, interpretable with feature importance. But it requires extensive parameter tuning and may struggle with very complex feature interactions.
- Deep Learning Models (DNNs): Powerful in capturing complex, non-linear relationships across multiple features, scalable to large datasets, and capable of modeling interactions that traditional models may miss. But risk of

overfitting on small datasets requires larger training time and higher computational resources.

3. Dashboard and User Interface

Available Technologies

- Frontend: React.js, Angular, Vue.js. React is most popular, component-based, reusable, and lightweight.
- Backend: Flask (Python), Django (Python), Spring Boot (Java). Flask provides simple integration with ML models. Django is heavier but provides built-in admin tools.
- Database: PostgreSQL vs MySQL vs MongoDB

Selected Technology:
Stack:

- Frontend: React.js (for modular and user-friendly dashboards).
- Backend: Flask (Python) for seamless ML model integration.
- Database: PostgreSQL for storing tabular agricultural data, reports, and predictions.

4.4. Expected Test Cases

Table 4.4: Test Cases

Test Case ID	Test Case Description	Precondition	Expected Outcome
TC001	User registration	User opens the application	User account is successfully created and stored in the DB
TC002	Collect environmental data from IoT devices	IoT devices are properly installed and connected to the mobile app	Real-time soil pH, temperature, humidity, and moisture data are successfully captured
TC003	Predict yield based on current IoT and trained historical data	Historical training data is available and IoT sensors provide current values	Forecasted yield value is displayed
TC004	Generate recommendations for user based on prediction results	Yield prediction is completed	Actionable recommendations (fertilizer, irrigation, crop selection) are provided
TC005	Display output in user's mobile app	Prediction and recommendation modules have completed execution	Forecasted yield and recommendations are displayed clearly in the app dashboard

5. COMMERCIALIZATION

Our proposed smart agriculture support system for Aloe vera is intended to be commercialized as a mobile application to address the specific needs of the Sri Lankan agricultural sector. The target audience includes Aloe vera farmers, agricultural researchers, and stakeholders in the cosmetic and pharmaceutical industries.

The system will be offered in a freemium model with two versions. The Community version will be freely available and will include essential features such as basic disease detection, simple fertilizer recommendations, and general weather alerts. This version aims to establish a broad user base and demonstrate the system's value. The Premium version will be available through a subscription and will offer advanced functionalities. This includes real-time fertilizer management with AI-driven alerts, predictive analytics for market price trends, and more comprehensive yield forecasting.

Our system is designed with a user-friendly interface that does not require advanced technical knowledge or have age restrictions, making it accessible to a wide range of farmers. Our development team, composed of technically proficient graduates with industry experience, is well-equipped to build and maintain this robust platform.

The required funding is estimated at $20 per month, which will cover the operational costs for development and maintenance. Our marketing strategy will focus on targeted outreach within the farming community through agricultural workshops, partnerships with government bodies like the Department of Agriculture, and advertising via local media and social media platforms to ensure maximum visibility and adoption. Revenue will be generated primarily through subscriptions to the premium version, with potential future income from partnerships with agricultural input suppliers and data-driven insights for industry partners.

6. GANTT CHART AND WORK BREAKDOWN STRUCTURE

6.1. Gantt Chart

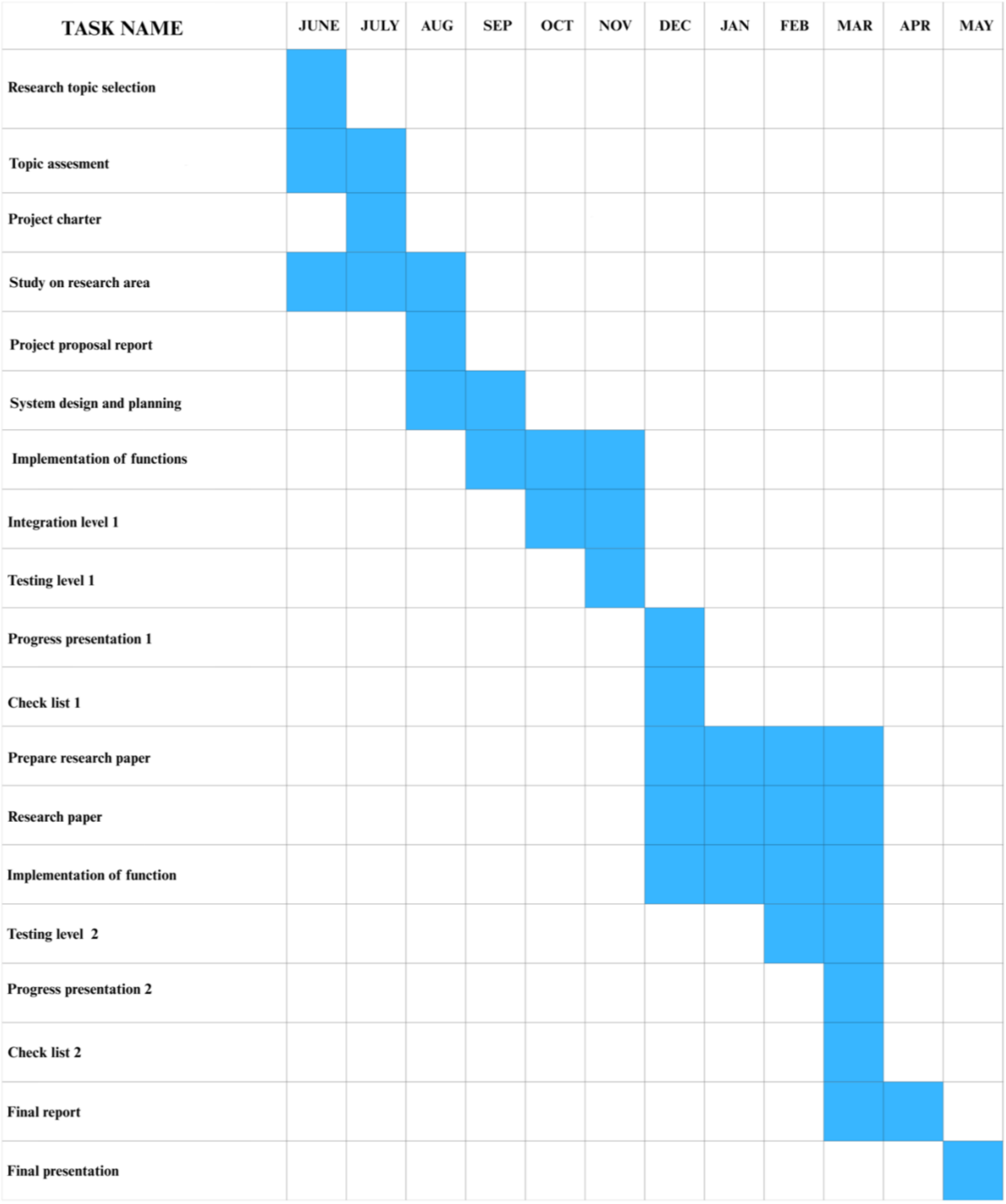

Figure 6.1: Gantt Chart

6.2. Work Breakdown Structure

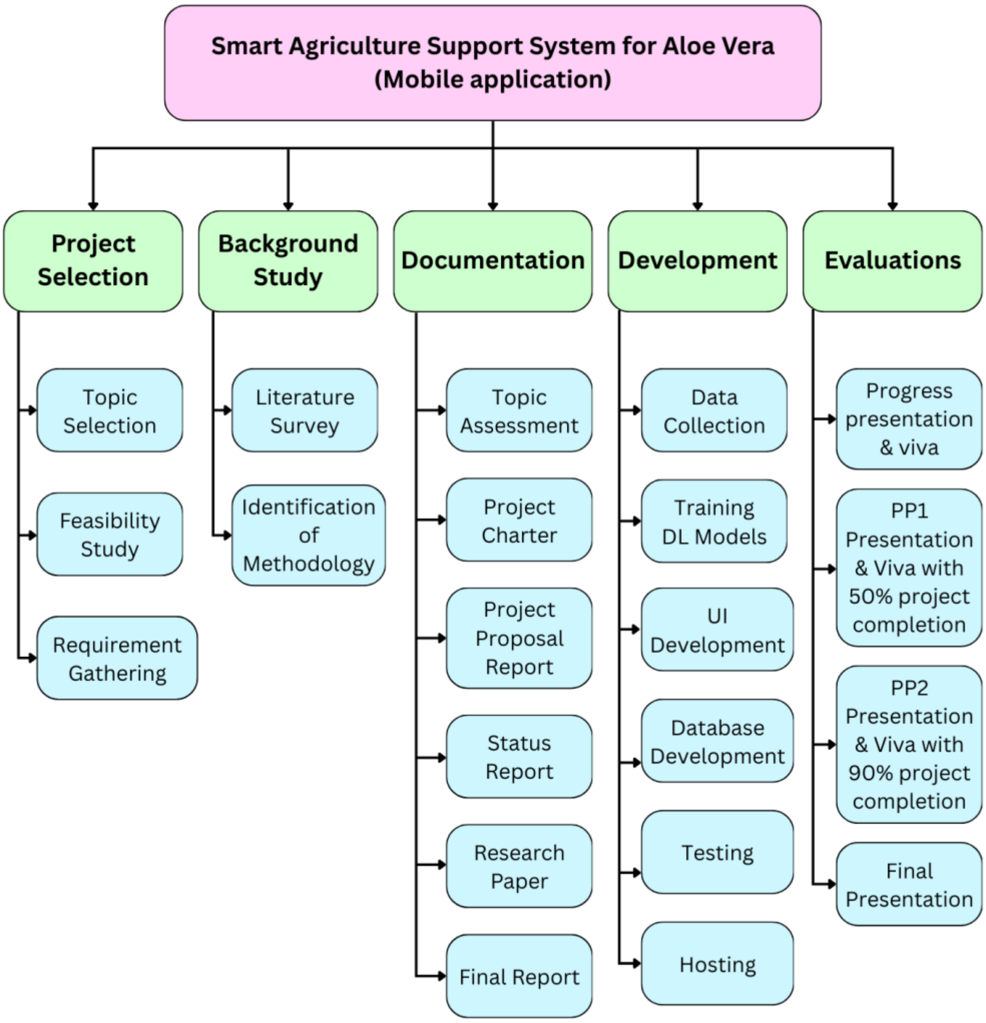

Figure 6.2: Work Breakdown Structure

REFERENCES

[1] T. Chowdhury, M. A. Rahman, K. Nahar, M. A. H. Chowdhury, and M. S. I. Khan, “Growth and yield performance of *Aloe vera* grown in different soil types of Bangladesh,” *J. Bangladesh Agric. Univ.*, vol. 16, no. 3, pp. 448-456, 2018, doi: 10.3329/jbau.v16i3.39416.

[2] G. Wagan, A. H. Soomro, and F. Channa, “Influence of sunlight on growth and yield of *Aloe vera* plant,” *Am. J. Biosci. Bioeng.*, vol. 10, no. 4, pp. 58-65, 2022, doi: 10.11648/j.bio.20221004.12.

[3] A. E. Khaki and L. Wang, “Crop yield prediction using deep neural networks,” *Front. Plant Sci.*, vol. 10, art. 621, pp. 1–14, 2019, doi: 10.3389/fpls.2019.00621.

[4] M. Van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” *Comput. Electron. Agric.*, vol. 177, art. 105709, pp. 1-18, 2020, doi: 10.1016/j.compag.2020.105709.

[5] K. Jha, A. Doshi, P. Patel, and M. Shah, “A comprehensive review on automation in agriculture using artificial intelligence,” *Artif. Intell. Agric.*, vol. 2, pp. 1-12, 2019, doi: 10.1016/j.aiia.2019.05.004.

[6] S. Saha, O. D. Kucher, A. O. Utkina, and N. Y. Rebouh, “Precision agriculture for improving crop yield predictions: A literature review,” *Front. Agron.*, vol. 7, art. 1566201, pp. 1-14, 2025, doi: 10.3389/fagro.2025.1566201.

[7] S. Addu, S. Sheelam, S. Mekala, N. Sulthana, L. Mekala, and Z. Alsalami, “Assessing environmental impact: Machine learning for crop yield prediction,” in *Proc. ICSMEE*, Hyderabad, India, 2024, pp. 1-10, doi: 10.1051/e3sconf/202452903008.

[8] K. S. Babu, V. Kumar, and P. Reddy, “Enhancing crop yield prediction using IoT-based soil moisture and nutrient sensors,” in *Proc. ICIAITES*, Tamil Nadu, India, 2025, pp. 1-7.

[9] M. Ramesh, P. Priya, and U. Muthaiah, “Predicting yield of crops using machine learning algorithm,” in *Proc. ICAICCIT*, Chennai, India, 2024, pp. 1-6.

[10] R. Sharma, S. Jain, and P. Kumar, “Machine learning approaches applied in smart agriculture for the prediction of agricultural yields,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 3, pp. 140-150, 2024, doi: 10.14569/IJACSA.2024.0150318.

[11] A. Patel and V. Singh, “Machine learning approaches applied in agriculture yield forecasting: An experimental study,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 6, pp. 101-110, 2024, doi: 10.14569/IJACSA.2024.0150612.

[12] A. Sharma, R. Gupta, and S. Verma, “Synergizing smart agriculture with hybrid deep learning: Predicting crop yields using IoT,” *BIO Web Conf.*, vol. 45, art. 05009, pp. 1-6, 2024, doi: 10.1051/bioconf/20244505009.

[13] A. Gupta and K. Patel, “Hybrid deep learning with empirical mode decomposition for yield prediction,” *BIO Web Conf.*, vol. 46, art. 02007, pp. 1-7, 2024, doi: 10.1051/bioconf/20244602007.

[14] S. K. R. Kumar and R. Manikandan, “Edge AI-powered system architecture for *Aloe vera* plant disease detection,” *SSRG Int. J. Electron. Commun. Eng.*, vol. 12, no. 1, pp. 45-52, 2025, doi: 10.14445/23488549/IJECE-V12I1P106.

[15] N. Kavga, K. Tsafarakis, and M. Kokkinakis, “The impact of greenhouse integrated photovoltaics on *Aloe vera*cultivation,” *Not. Bot. Horti Agrobot. Cluj-Napoca*, vol. 52, no. 4, pp. 145-156, 2024, doi: 10.15835/nbha52413214.

[16] V. Mirza, A. Khan, and P. Singh, “Crop yield forecasting using machine learning techniques - a systematic literature review,” *Eur. J. Agron.*, vol. 112, art. 125933, pp. 1-15, 2023, doi: 10.1016/j.eja.2023.125933.

[17] M. Tripathi, R. Gupta, and A. Singh, “Crop yield prediction using systematic review process based machine learning algorithm,” in *Proc. ICAICCIT*, Chennai, India, 2024, pp. 1-7.

[18] R. Govindaraj, N. Narayanasamy, and P. Vijay, “Crop yield prediction using advanced deep learning techniques,” in *Proc. ICICCS*, Coimbatore, India, 2025, pp. 1-8, doi: 10.1109/ICICCS57786.2025.10412345.

[19] A. Haval and M. Rahman, “Synergizing smart agriculture with hybrid deep learning -predicting crop yields using IoT,” *BIO Web Conf.*, vol. 44, art. 03005, pp. 1-6, 2024.

[20] A. Imade, S. Ounacer, and M. Elghoumari, “Machine learning approaches applied in smart agriculture for the prediction of agricultural yields,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 10, pp. 832-840, 2024, doi: 10.14569/IJACSA.2024.0151088.

[21] Z. Alsalami, G. Mohammed, and T. Srinivas, “Enhancing Crop Yield Prediction Using IoT-Based Soil Moisture and Nutrient Sensors,” SHS Web Conf., vol. 216, art. 01029, 2025, doi: 10.1051/shsconf/202521601029.

[22] L. Nevavuori, A. Narra, and T. Linna, “Crop yield prediction with deep convolutional neural networks,” *Comput. Electron. Agric.*, vol. 163, art. 104859, pp. 1-9, 2019, doi: 10.1016/j.compag.2019.104859.

[23] D. Lobell, G. Hammer, and P. Chenu, “The critical role of extreme heat for maize production in the United States,” *Nat. Clim. Change*, vol. 3, pp. 497-501, 2013, doi: 10.1038/nclimate1832.

[24] P. Nevavuori, N. Narra, P. Linna, and T. Lipping, “Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models,” Remote Sens., vol. 12, no. 23, art. 4000, 2020, doi: 10.3390/rs12234000.

[25] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, “Deep Gaussian process for crop yield prediction based on remote sensing data,” in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, USA, 2017, pp. 4559-4566.

[26] F. Liu, X. Jiang, and Z. Wu, "Attention Mechanism-Combined LSTM for Grain Yield Prediction in China Using Multi-Source Satellite Imagery," Sustainability, vol. 15, no. 12, art. 9210, 2023, doi: 10.3390/su15129210.

[27] C. Zhang and C. Diao, "A Phenology-Guided Bayesian-CNN (PB-CNN) Framework for Soybean Yield Estimation and Uncertainty Analysis," ISPRS J. Photogramm. Remote Sens., vol. 205, pp. 50–73, 2023, doi: 10.1016/j.isprsjprs.2023.09.025.

[28] M. S. Mia, R. Tanabe, L. N. Habibi, N. Hashimoto, K. Homma, M. Maki, T. Matsui, and T. S. T. Tanaka, "Multimodal Deep Learning for Rice Yield Prediction Using UAV-Based Multispectral Imagery and Weather Data," Remote Sens., vol. 15, no. 10, art. 2511, 2023, doi: 10.3390/rs15102511.

[29] M. Sobhana, S. C. Ch., D. N. V. S. L. S. Indira, and K. K. Kumar, "CROPUP – A Crop Yield Prediction and Recommendation System with Geographical Data using DNN and XGBoost," Int. J. Recent Innov. Trends Comput. Commun., vol. 10, no. 11, pp. 53–62, 2022, doi: 10.17762/ijritcc.v10i11.5780.

[30] M. F. Danilevicz, P. E. Bayer, F. Boussaid, M. Bennamoun, and D. Edwards, "Maize Yield Prediction at an Early Developmental Stage Using Multispectral Images and Genotype Data for Preliminary Hybrid Selection," Remote Sens., vol. 13, no. 19, art. 3976, 2021, doi: 10.3390/rs13193976.

APPENDICES

Appendix – A: AleoGreen Logo

Figure 8.1: AleoGreen Logo

Figure 8.2: Home UI

Figure 8.3: Welcome UI

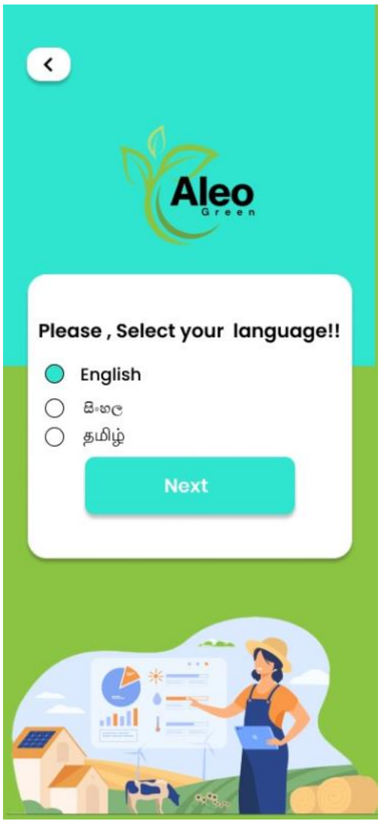

Figure 8.4: Language Select UI

Appendix– B: Field Visit Photographs

Figure 8.5: Aloe vera plantation overview

Figure 8.6: environmental management of the field.

Figure 8.7: Close-up of an Aloe vera plant.

Appendix– C: Plagiarism score using Turnitin