CROP YIELD PREDICTION USING MACHINE LEARNING

24-25J-125

Final Report

B.Sc. (Hons) Degree in Information Technology specialized in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

April 2025

CROP YIELD PREDICTION USING MACHINE LEARNING

24-25J-125

Final Report

Dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Hons) in Information Technology
Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

April 2025

DECLARATION

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

Also, I hereby grant the Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or in part in print, electronic, or other media. I retain the right to use this content, in whole or in part, in future works (such as articles or books).

Name	Student ID	Signature
Jayathilaka D.H.R.A	IT21308352	
Piyumani K.V.P	IT21227868	
Amarasinghe A.I.S.A	IT21225192	
Jayasekara S.S.D	IT21227318	

………………….. ……………………

Kanishka Yapa Date

(Supervisor)

ABSTRACT

This study explores the use of machine learning, specifically the Random Forest method, to forecast paddy yield based on agricultural and environmental factors. Using a dataset of over 10,000 data points, including variables such as weather, soil conditions, irrigation practices, seed types, and historical yields, the data was preprocessed through cleaning, outlier detection, and normalization. The Random Forest model was chosen for its ability to handle complex, non-linear relationships. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared. The results show that the model accurately predicts paddy yield and provides valuable insights into key factors affecting yield. This research demonstrates the potential of Random Forest in agricultural yield prediction, aiding farmers in decision-making, crop management, and resource optimization. Future work could further enhance the model's flexibility and accuracy by integrating real-time data and satellite imagery.

RiceGenie is a comprehensive intelligent system for transforming paddy production utilizing data-based technologies. The system comprises four integrated modules: Crop Yield Prediction, Paddy Variety Prediction, Pest and Disease Detection and Management, and Weed Detection and Management. This report details the Paddy Variety Prediction module, which assesses which paddy varieties are most appropriate for each region according to local agro-ecological conditions. The model is machine learning based and analyzes a diversity of the following datasets: historical yield, climate indicators such as rainfall and temperature, and soil quality. After preprocessing and feature selection, different classification algorithms were evaluated to predict which rice variety best matches each region. The results showed that recommended varieties increase yield potential and production efficiency for their respective region. Consequently, the model assists farmers in making data-driven decisions about which new variety to grow, thereby enabling more sustainable practices. The purpose of adding this module is to link science-based practices to real-time field-based applications in the RiceGenie platform.

Paddy cultivation forms the backbone of Sri Lanka’s agricultural economy, yet it faces persistent threats from pre-harvest diseases such as Bacterial Leaf Blight, Brown Spot, Leaf Blast, and Sheath Blight. These diseases significantly reduce yield and quality, affecting the livelihoods of farmers and food security. Traditional detection methods relying on manual inspection are time-consuming, subjective, and inefficient at scale. To address these limitations, this research proposes a deep learning-powered system for automated disease identification and treatment recommendation. Leveraging Convolutional Neural Networks (CNNs), the system classifies five key disease categories with high accuracy using paddy leaf images. The innovation lies in its dynamic treatment engine, which factors in real-time farmer-provided inputs like humidity, nitrogen use, irrigation method, and weather conditions to tailor logical and effective management strategies.

The platform also integrates a machine learning-based harvest prediction model to estimate yield impact under varying environmental and disease conditions. Developed as a responsive web application, the system features a ReactJS frontend, FastAPI backend, TensorFlow model deployment, and MySQL for data management. Google Colab was used for model training, while version control and task management were handled through GitHub and Trello. By providing real-time disease predictions, treatment guidance, and actionable insights, this research offers a smart, scalable, and farmer-friendly solution that promotes precision agriculture. It empowers Sri Lankan farmers to make informed decisions, reduce dependency on chemical treatments, and improve crop outcomes. Future enhancements include expanding the disease classification scope and incorporating predictive weather data to further optimize disease control and harvest planning.

Keywords - Rice Variety Recommendation, Machine Learning, Agro-ecological Zoning, Environmental Parameters, Logistic Regression.

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to our supervisor, **Mr. Kanishka Yapa**, for his invaluable guidance, unwavering support, and constant encouragement throughout this research. His deep expertise, critical insights, and meticulous attention to detail were instrumental in shaping both the technical and conceptual aspects of our project. His mentorship went far beyond academic supervision—his thoughtful feedback and consistent availability helped us stay focused, motivated, and confident at every stage of the study.

We also wish to extend our sincere thanks to our co-supervisor, **Dr. Harinda Fernando**, whose continuous support, technical acumen, and insightful suggestions played a significant role in refining our methodology and approach. His constructive criticisms and thought-provoking discussions enriched our understanding and pushed us to maintain the highest standards of research. We are truly grateful for the time and dedication he invested in helping us enhance the quality of our work.

Our appreciation also goes out to the **Rice Research and Development Institute (RRDI)**, as well as the agricultural officers and field experts who generously shared their practical knowledge and insights. Their expertise regarding weed identification, field-level challenges, and Sri Lankan paddy cultivation practices added immense value and relevance to our research. Their contributions helped us align our solution more effectively with real-world agricultural needs. We are deeply thankful to our peers, project mates, and friends who supported us during this academic journey. Whether it was through technical discussions, moral encouragement, or simply being a source of motivation during challenging times, their support has been truly meaningful and greatly appreciated.

Finally, we would like to extend our sincere gratitude to our families for their endless patience, love, and emotional support. Their belief in our potential, even during moments of doubt and difficulty, provided us with the strength to overcome challenges and persevere. Their encouragement created a stable foundation that allowed us to dedicate ourselves fully to this research endeavor.

This project would not have been possible without the collective contributions, support, and encouragement of all the individuals mentioned above. We are truly grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
BLB	Bacterial Leaf Blight
RF	Random Forest
RLB	Rice Leaf Blast
CAM	Class Activation Map
Grad-CAM	Gradient-weighted Class Activation Mapping
UI	User Interface
UX	User Experience
API	Application Programming Interface
GPU	Graphics Processing Unit
ROI	Region of Interest
DB	Database
SQL	Structured Query Language
MySQL	My Structured Query Language (Relational DBMS)
RMSE	Root Mean Square Error
MSE	Mean Squared Error

MAPE	Mean Absolute Error
FASTAPI	Fast Asynchronous API (Python-based web framework)
JSON	JavaScript Object Notation
REST	Representational State Transfer
IoT	Internet of Things
ML	Machine Learning
Colab	Google Colaboratory

1 INTRODUCTION

1.1 Background literature

Predicting paddy yield is essential for food security, but it is difficult due to environment, climate, and agricultural variables. Accurate prediction can provide insights for farmers on how to allocate resources appropriately and assist policymakers in the distribution of food and resources. Strong performance is limited through traditional methods, and climate change and market fluctuations have made prediction increasingly difficult; therefore, advanced prediction methods are essential for increasing productivity and sustainability. The Random Forest (RF) algorithm is well regarded for its robustness and flexibility in agricultural applications. The RF algorithm requires little preprocessing of the features and uses numeric and categorical variables, which is particularly useful in agricultural datasets.

The RF algorithm ranks the importance of features, which can assist researchers in identifying the most important variables in predicting crop yield, such as rainfall, temperature, and soil type.) Several studies have demonstrated RF is often more accurate and reliable than regression models. However, while the RF algorithm can predict paddy yield, it faces several challenges in regions like Sri Lanka. Data availability and quality can present challenges as many farmers do not have access to a sufficiently high-quality dataset, while severely variable climate conditions can introduce a level of uncertainty and reduce reliability (severe droughts or floods, for example).

In Sri Lanka, rice is a staple food and an important economic crop, with approximately 1.8 million farmer families growing rice in a range of agro-ecological zones. Region-specific rice varieties are key to food security, farmer income, and resilience to climate change. The Rice Research and Development Institute (RRDI) and its regional stations, such as Labuduwa, have developed higher-performing rice varieties that exhibit desirable traits, but, in many instances, farmers are still using traditional varieties because of limited awareness and the links between research outputs and actual practices in the field are very weak.

The first field visits to Labuduwa and our initial conversations with agricultural specialists allowed for the identification of four prominent factors that challenged paddy growing: yield, cropping variety, disease, and weeds, which we used as the basis of our research. With support from experts like Dr. Millawithanachchi, Dr. Udawela, and Mrs. Arachchi, the research team encouraged the incorporation of farm-level challenges into the project.

Other countries have developed AI and machine learning to do crop planning and to predict varietal type. However, there is no intelligent system in Sri Lanka to achieve these tasks specific to rice cultivation. In other words, there is a clear gap in research to produce yield predictions and varietal recommendations to farmers by district, based on environmental data and previous yield. This study addresses this gap and provides farmers in Sri Lanka with data-informed tools to make decisions regarding varieties and to be more sustainable in respect to agriculture, based on their environmental conditions.

In recent years, deep learning approaches for paddy disease detection have garnered significant interest because they can enhance crop productivity and mitigate crop losses from undiagnosed plant infections. Traditional disease detection relies upon manual inspection, which is slow, labor-intensive, and potentially error prone. As a result, researchers are increasingly employing Convolutional Neural Network (CNN)-based image classification systems for the identification of paddy leaf diseases. For example, studies by Mohanty et al. (2016) and Sladojevic et al. (2016) have established the effectiveness of CNNs for identifying crop disease in images with high accuracy.

Recent work has expanded efforts to improve classification performance on a variety of datasets with multiplex disease classes and variances in real-field images through data augmentation, transfer learning (USAGE), and utilizing more efficient architectures like ResNet and DenseNet.

In parallel, there has been increasing attention to placing the diagnosis of a disease into a treatment recommendation for farmers to put into practice, some of which combines an expert knowledge base with model output to recommend specific treatments in context, including fungicides, bactericides, and preventative options. Implementing this recommendation logic into a system is the next step in bridging the divide between detection and action. Other approaches have included a rule-based decision engine and/or a straightforward NLP-based mapping of symptoms to treatments. These methods ultimately aim to support smallholder farmers with timely and accurate advancement to assist in crop loss mitigation, chemical misuse, and sustainable farming methods. Together, as a full functioning system, disease classification and treatment recommendation provide decision support, which contributes to smart agriculture initiatives, much like the work already being done in regions such as Sri Lanka.

Weeds are undesirable plants that grow alongside crops and compete for essential resources, including light, water, and nutrients in the soil. In paddy cultivation, weeds may grow in a similar nature to rice, making them difficult to detect and manage, especially at tillering and earlier stages of growth. The Department of Agriculture, Sri Lanka, has reported that without weed management and control, paddy can be subject to a yield loss margin of 15% to 50%.

Weeding is typically completed by either manual labor (picking and pulling weeds), which is time-consuming and an unsustainable method due to the labor shortage, especially in youth, or herbicides. Herbicides are effective but can have lasting effects of soil degradation, water quality issues, and human health implications for farmers and consumers alike.

As agriculture continues to develop globally with smart farming and precision agriculture, Sri Lanka can take advantage of new options with modern technologies such as Machine Learning (ML) and Artificial Intelligence (AI), which can provide potential solutions aiming at productivity and sustainability. These technologies could allow weeding to be automated through image recognition, allowing for targeted weddings, reduced herbicide, and timely response.

1.2 Research Gap

While there have been advancements in the prediction systems to estimate paddy yields, there are still gaps. With most yield prediction systems focusing on weather indices, real-time data sources, such as soil moisture and pest management, have not been incorporated into most self-reported models. Without integrating real-time data options, it is challenging for prediction models to accurately predict or make proper recommendations very mindfully of varying climate and agronomy factors as they change with time. The bulk of the research effort also seems to not focus much on socioeconomics' impacts on paddy yield, for example. In general, while agronomics and climate are well researched, we seem not too focused on farmer behavior, access to resources, and housing market variables' impacts. Taking into consideration those economic indicators could offer a much more holistic perspective of the variability in yield and help develop better decision support systems for farmers when the needs of farmers aren't fully understood.

Reference	Focus Area	Research Gap	How We Address It
[1]	Market trends and rice production analysis	Lacks technological advancements for disease detection and management	Develop a deep learning-based web app for real-time paddy disease detection and treatment recommendations
[2]	Overview of global rice pest and disease management	Does not leverage AI/ML for precise detection and automated recommendations	Utilize CNN-based models to classify diseases and provide treatment suggestions

[3]	Machine vision for detecting rice blast disease	Limited to a single disease; lacks a comprehensive multi-disease detection system.	Extend the detection of five major paddy diseases and integrate treatment recommendations.
[4]	Deep learning for paddy disease detection	Focuses only on detection; lacks treatment recommendations and a farmer-friendly interface	Develop a user-friendly web platform providing both detection and actionable treatment suggestions.
[5]	Analyzing the impact of **seed broadcasting rates** on paddy yield and weed control in dry zones of Sri Lanka.	Limited research on the **optimal seed broadcasting rates** for **yield maximization** while minimizing disease risks (e.g., sheath blight).	Integrate the **seed rate** as a variable in our model and fine-tune predictions to recommend the best rates for specific **climates and soil types** to optimize yield.
[6]	Investigating the **market demand**, **cultivation constraints**, and **economic feasibility** of traditional paddy varieties across five districts in Sri Lanka.	Lack of a **data-driven** system to predict **the best-suited rice varieties** based on **market demand** and **regional environmental factors**.	Utilized **machine learning** to analyze **historical yield**, **climatic conditions**, and **market trends** to provide accurate **variety recommendations** for different regions.
[7]	Identifying traditional rice varieties for saline conditions through participatory research in Hambantota.	Lack of integration between **local farmer knowledge** and scientific approaches for selecting **saline-resistant** rice varieties.	Incorporate climate-based features like salinity levels and farmer-driven data into our Paddy Variety Prediction Model for region-specific recommendations.

[9]	Paddy yield prediction using weather indices and Random Forest	Focuses on climatic factors; does not consider soil properties, irrigation practices, or pest impact	Expand the model to include variables like soil type, irrigation methods, and pest damage for a holistic prediction approach
[10]	Paddy yield forecasting using ANN with Back-Propagation learning	Limited to ANN and climatic factors; lacks exploration of other machine learning models and additional agronomic variables	Implement a **Random Forest** model incorporating diverse factors such as soil type, irrigation methods, and pest damage for improved prediction accuracy
[11]	Applied regression techniques for rice yield estimation	Limited to a specific province; did not utilize advanced machine learning models	Expand the study to multiple provinces using advanced machine learning techniques like Random Forest and Gradient Boosting
[12]	Rice yield prediction using weather-based machine learning models.	Limited to weather parameters (rainfall, temperature, radiation) and does not incorporate other agro-climatic factors such as soil properties, irrigation, and pest severity.	Incorporates additional key factors such as soil nutrients, fertilizer usage, irrigation type, and pest severity to enhance predictive accuracy. A robust preprocessing pipeline is applied, and multiple machine learning models are evaluated.

Table 1: Research Gap

While Sri Lanka has produced hundreds of improved rice varieties through institutes such as the RRDI and the Labuduwa satellite station, most farmers still plant and follow traditional varieties, leaving a very significant communication and technology gap.

Although there is a lot of data available on relevant parameters like soil data and climate data from agroecological zones, in practice, this data does not end up in the hands of smallholder farmers as usable or accessible information. Even on-farm extension services do not personalize or have the flexibility to help farmers with agroecological challenges particular to their area, especially in vulnerable zones like WL1 and WL2a. Countries around the world, including India and the Philippines, have developed machine learning and GIS-based systems to provide farmers with localized and real-time crop recommendations. Sri Lanka has not even experienced this integration into farming, which is especially the case for rice varietal selection. Sri Lanka has diverse agro-ecological zones suitable for data-driven varietal matching, yet the machine learning model has not yet evolved for this purpose, and the datasets we have are not typically designed for farmers for use.

The Variety Genie module of the Rice Gene system fills this significant void. It utilizes a supervised machine learning model trained on localized environmental and yield data to generate rice variety recommendations at the district level. Its web interface takes complex and varied data and makes it practical and understandable for agricultural officers and farmers. VarietyGenie does more than just fill the gap between research and practice; it creates a co-learning network for researchers, programmers, and policymakers. Ultimately, this research repurposes static agricultural data into dynamic, engagement-driven tools that provide Sri Lankan farmers with customized variety options that are also climate-resilient. It builds a framework to scale technology-enabled agro-innovation that is directly responsive to challenges on the ground and represents an important step towards more considerate, sustainable practices in paddy cultivation.

Although there is considerable research utilizing deep learning methods for the classification of paddy disease, most of this research focuses only on model performance and does not cover the usability aspects, including treatment recommendations and guidance provided to farmers. In addition to this, most existing models have been trained with limited and/or laboratory quality data, which does not fully represent the variability and noise present in real-world field images. Furthermore, there is a lack of systems that combine disease detection with context-specific, localized treatment options for farmers in an accessible way. There are very few studies that considered the full pipeline from identifying disease to and making a decision-action treatment that is specific to pre-harvesting activities, therefore, evidence suggests that there is a need to develop an easy-to-use system that combines strong real-time disease detection properties alongside treatment recommendations for paddy cultivated areas like Sri Lanka. Along the same lines, there is a substantial body of literature where florescence studies to identify individual plant species, however there remains a gap in research specificity to the detection and classification of multiple weed species at once, and this further complicates implementation into practice within an agricultural context, where knowing what type of weed is present across thousands of hectares of crops is necessary, and crucial an effective weed management strategy.

Despite the significant advancements in the application of imaging processing technology to weed detection, several major research gaps still exist: No Standard Datasets: The lack of comprehensive, annotated, and standard datasets hinders the development of high-performing weed detection models. The absence of standard datasets limits deep learning models' ability to train effectively and compare results across studies. Sensitivity to Environmental Variation: Many existing models were designed and tested in a controlled environment. Therefore, the performance of models built for weed detection, tested, and validated in controlled environments is unknown in real field conditions, in which environmental variables (e.g., lighting, background clutter, occlusions, etc.) may be present. Limited Generalizability Across Species: Current models target specific plant or weed species, which limits their use to specific agricultural contexts. There is a need for generalized frameworks that are capable of recognizing and identifying multiple types of weeds across various cropping systems and geographies. Model Transparency: While deep learning models have demonstrated state-of-the-art performance, they still lack transparency. The "black box" nature of such models leads to an inability to understand the features driving the outcome, making it difficult to understand and verify outputs in a practical agricultural context.

It is imperative to address these gaps to further advance automated weed detection technologies. Future work should concentrate on establishing open-access datasets with a variety of tensions, improving models for environmental robustness, enhancing generalization capabilities, and creating more interpretable models. These factors will facilitate the development of scalable, reliable, and intelligent weed management systems within precision agriculture.

1.3 Research Problem

Paddy cultivation in Sri Lanka is vital for food security, but the sector faces challenges like climate variability, outdated farming practices, and ineffective resource use. The conventional tools for yield prediction do not sufficiently model the relationships among climate, soil, and socio-economic components, resulting in extreme inefficiencies, especially for smallholder farmers who do not have access to modern predictive tools. Machine learning models like Random Forest and Artificial Neural Networks have the potential for accurate yield forecasting; however, the number of studies using these models has been limited due to the restricted use of data and failure to account for dynamic influences in predictive tasks.

Furthermore, models that advance the predictive effort like LSTM and CNN have seen limited application in this sector. Lastly, in addition to using advanced models of prediction, factors that recognize the socio-economic context of farmers have been overlooked in the literature. Developing predictive models and decision support systems (DSS) that are appropriate for specific regions has the potential to increase farm-level productivity and sustainable agriculture practices in the Sri Lankan paddy sector.

Rice is a critical component of Sri Lanka’s culture, economy, and food security, underscoring the need to adopt high-yield, climate-resilient paddy varieties. Despite decades of research and development of over 250 improved rice varieties, mainly through the State-run Rice Research and Development Institute (RRDI) and its satellite station in Labuduwa, the uptake of these varieties at scale remains low. This can be largely attributed to a well-documented lack of communication between research institutions and farming communities.

The examples of farmers, particularly in vulnerable zones like WL1 and WL2a within Galle District, continue to grow traditional varieties that do not perform well in response to environmental stress imposed by salinity and erratic rainfall. The availability of better-suited varieties is not, in itself, related to the lack of information or assistance with implementation and adaptation of these new varieties, but rather to poor cropping choices that are influenced by both the absence of knowledge or awareness of suitable options, and by a range of shortcomings in accessible decision-support mechanisms, extension service and/or education systems. Ultimately, farmers' access to new varieties will generate poor returns (lower yields), diminished profitability, or potential climate vulnerability.

The extension methods currently in use—verbal advice, printed materials, and group discussions—are antiquated and do not offer regionally and timely specific recommendations. Data on the performance of rice varieties remains fragmented, static, and out of reach of farmers or even field officers, contributing to a lack of data-driven decision-making. Deciding on which variety to grow is difficult given the complex agro-ecological zone differences in Sri Lanka, and the use of tools that take in localized environmental conditions is advisable.

Countries globally, such as India and Vietnam, are developing machine learning and digital platforms to provide personalized recommendations in real-time, while Sri Lanka does not have scalable, intelligent prediction systems for rice variety [26] [27].

The RiceGenie project, through its Paddy Variety Prediction subcomponent, helps to bridge the gap between researchers and farmers by generating localized variety recommendations using machine learning and environmental data at the district level in an easy-to-use web interface complete with maps, calendars, and scientific profiles. RiceGenie transforms research data into driving decision-making tools for farmers to promote sustainable agriculture, resilience, and food security in Sri Lanka.

Paddy production is an integral part of Sri Lanka's agricultural economic status, with various sheath and leaf diseases threatening productivity, especially in the pre-harvest phase. Prompt disease diagnosis and management are vital to reduce crop loss and food insecurity overall. Recently, deep learning and computer vision are revealing potential for detecting plant leaf or sheath diseases, however, most existing machine learning or deep learning models have been designed and optimized in laboratory conditions, with only small image datasets available for continued evaluation. Because of this controlled testing and limited image dataset, almost all current models perform poorly in real-world conditions due to variability, including changes in lighting, background variation, and overlapping symptoms that are a common occurrence when capturing images in the field.

Furthermore, existing systems tend to focus on the classification of a disease and do not facilitate the next important step, which is providing specific and actionable treatment recommendations. Farmers still have no guidance on which fungicides and bactericides to use and what agronomic practices can be used to treat the identified disease. Additionally, these systems fail to consider local agricultural contexts or the user-friendliness that is critical for non-technical users such as smallholder farmers. Consequently, there is a significant gap in providing an end-to-end solution that detects paddy diseases from real images and makes recommendations of effective and regionally appropriate treatments. In this research, we aim to fill this gap by developing a comprehensive and deep learning-based system for the early detection of paddy diseases and treatment guidance specific to farmers in Sri Lanka.

This study focuses on the ineffective and unsustainable management of weeds in paddy farming in Sri Lanka. While there is a growing trend towards machine learning (ML) and digital agriculture internationally, there is no affordable, automated solution for weed detection or management in Sri Lanka. Current methods are either unsustainable or inadequate, while smart solutions for weed management have not yet been developed for the varied agro-ecological zones of the country.

1.4 RESEARCH OBJECTIVES

1.4.1 Main Objective

This research aims to establish a reliable prediction system for paddy yield by applying machine learning methods (Random Forest RF), including climatic, agronomic, and socio-economic data. Compared with other methodologies, this system will provide timely, localized predictions that take into account Sri Lankan agricultural conditions. With past and real-time data, the development of this model aims to facilitate better decision-making for farmers and policymakers.

Additionally, the project will implement a user-friendly decision support system (DSS) to help smallholder farmers plan planting, irrigation, and fertilizer application. Addressing climate variability, resource inefficiency, and technological gaps, the research seeks to enhance productivity, encourage sustainable practices, and strengthen national food security.

The overall goal of this study is to create a machine-learning-based paddy variety prediction model, referred to as VarietyGenie, that provides farmers with the most accurate, locally relevant recommendations for rice cultivation that can be derived from environmental, agronomic, and historical yield data. The study aims to reduce the gap between the knowledge of breeding scientists and real-world practices in farming. The research will not only close the information gap but will also enable precision agriculture within Sri Lanka.

To create a machine-learning-based Paddy Variety Prediction Model, known as VarietyGenie, that identifies the best-suited rice varieties per district in Sri Lanka based on environmental, soil, and historical yield data.

The primary goal of this research is to create an intelligent and user-friendly system that utilizes deep learning techniques to accurately identify pre-harvesting paddy diseases and give users treatment suggestions that are relevant to the Sri Lankan agricultural context. This includes developing a reliable image classification model that will be able to identify several pre-harvesting paddy diseases common in Sri Lanka, including bacterial leaf blight, brown spot, leaf blast, and sheath blight, using field-based images of crop plants taken by farmers or agriculture officers.

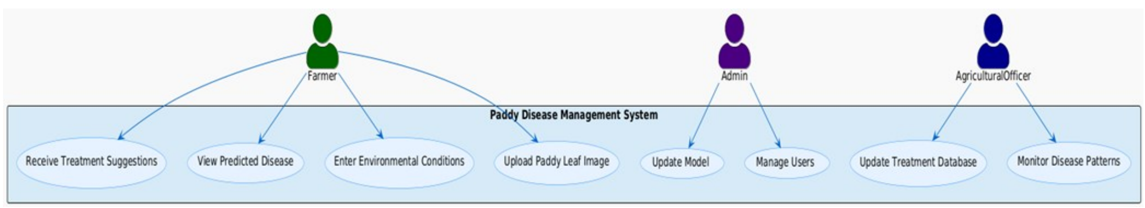

Figure 1: Disease Management System

In addition to detecting disease, this system aims to facilitate the gap between diagnosis and decision-making through practical treatment guidance specific to the disease and crop. The recommendations would include recommended chemical treatments, methods for application, and prevention options underpinned by expert agricultural knowledge and best practices. The system will be implemented as an application on the web and will enable farmers to upload images of paddy leaves and receive immediate diagnostic results and treatment advice in a straightforward and accessible user interface. The goal is ultimately to provide farmers a tool to help them improve early disease management, crop loss protection, and sustainable farming practices utilizing modern AI technologies.

The primary focus of this study is to improve the efficiency and sustainability of rice growing in Sri Lanka by targeting an important agronomic challenge: weed control.

Weeds compete with rice for key resources, including nutrients, water and light; consequently, suppressing crop yield and increasing costs of production. Current methods of weed control are primarily manual and labor-intensive, and rely on high levels of chemical herbicide application, which can lead to negative environmental impacts.

In response to these limitations, the research proposes a sophisticated weed identification and management system based on machine learning methods. The system intends to recognize and classify weed species in paddy fields and adjust identified species with consequential site-specific management strategies. This state-of-the-art system aims to limit the time and labor demands from physical weed identification, reduce reliance on the synthetic herbicides primarily through targeted application approaches, and improve overall crop production by providing responsible weed management opportunities in a timely and accurate manner.

This area of research is aligned to the larger national agenda to modernize agricultural practices in Sri Lanka and will contribute to improved paddy production to enhance food security and economic resilience.

1.4.2 Sub Objectives

This study focuses on gathering and processing a dataset with primary attributes influencing paddy yield parameters, which will combine climate, soil composition, irrigation practices, seed variety, and socio-economic factors. Upon collecting the relevant data, exploratory data analysis (EDA) and feature selection will be undertaken to identify key variables. Afterward, a Random Forest model will be trained and optimized with metrics such as MAE, RMSE, and R^2 Score. The Random Forest model will also be used as a benchmark against other algorithms, such as Linear Regression and XGBoost.

Furthermore, the study will build a user-friendly decision support system (DSS) that will assist smallholder farmers. The decision support system will account for input data, including district, area, and seed variety, to project yield potential, provide a recommended planting method, irrigation practices, and fertilization. The user will also receive input regarding the intent of the recommendation (based on the district) to improve collaboration between farmers and policymakers, enhance efficiency in farming, and support food security strategies.

Preprocessing datasets from appropriate sources of credible information from the Rice Research and Development Institute (RRDI), Labuduwa Research Station, and the Department of Census and Statistics. The datasets include climate variables, soil types, agro-ecological zones, and the historical yield of rice varieties.

The next step of this project is to identify environmental and agronomic factors that affect varietal suitability. For example, drought tolerance, salinity tolerance, and soil pH will be identified as environmental and agronomic factors using statistical analysis and validated by experts. The third step of this project follows the identification of factors and involves building a predictive machine learning model. This predictive machine learning model will be built using approaches such as Random Forest and Gradient Boosting, by training it on historical datasets and optimizing the model using cross-validation.

To provide access to this intelligence, a user-friendly bilingual web platform will be built. Farmers and agriculture officers will be able to enter their regional data and receive a tailored set of recommended rice varieties with additional crop profiles and guidance on growing those varieties. There will be an interactive map displaying the suitability of the recommended varieties at the district level to be used in decision-making. Finally, we will assess how well the platform performs using developed performance measures and feedback from experts, which will ensure that the variety recommendations provided will be practical, accurate, and trusted in promoting sustainable paddy cultivation in Sri Lanka.

Finally, the assessment of the system's functionality will be done using performance metrics and feedback from experts to guarantee that it provides practical, accurate, and reliable recommendations to support sustainable paddy crop practices across Sri Lanka.

This research aims to develop a comprehensive system for paddy disease detection and treatment recommendation using deep learning. The first task will involve collecting and preprocessing a robust dataset of real-field paddy leaf images that exemplify five main diseases: bacterial leaf blight, brown spot, healthy, leaf blast, and sheath blight. The dataset will be augmented to demonstrate model generalization in real-world scenarios. The next step will be developing and implementing a deep learning model to classify these diseases accurately. The model will be evaluated based on accuracy, precision, recall, and F1-score to ensure reliable performance under real-world situations.

Concurrently, the project will build a treatment recommendation module linking each disease to appropriate chemical treatments, methods of application, and practices in disease prevention, evaluated using the domain expertise of agricultural experts. These feasible features will be developed into a user-friendly web application using ReactJS and FastAPI so that the application is accessible and convenient for farmers with limited technical skills. The last sub-objective relates to the validation of the complete system through a user's trial in a naturally occurring iterative evaluation, collecting qualitative and quantitative feedback to enhance both the disease detection model and treatment recommendation system. Overall, the end-to-end system will enable farmers to identify paddy diseases early while being able to act on that information to reduce crop loss and improve crop yield.

To address the overarching goal of improving weed management in paddy cultivation in Sri Lanka, the project is based around several specific objectives. The first objective is to develop and train a convolutional neural network (CNN)-based model to detect and classify paddy weed species from imagery captured in the field. After generating the model, the objectives are to assess and validate that model's accuracy and reliability in various field-based conditions, such as differences in lighting, soil types, or growth stages of the crop. In addition to detection, the project describes the implementation of specific mitigation strategies. The strategies proposed are rotation cropping, which would be seasonal rotation of crops as a natural method for reducing weed abundance while enhancing soil health; integrated weed management (IWM), using an amalgamation of strategies such as mechanical, chemical and biological means to develop lower reliance on herbicide application; and precision application methods to aid in spot application of herbicide based on real-time detection, promoting less herbicide use while reducing impacts on the environment.

Finally, the purpose of this study is to conceptualize an end-to-end framework for the practical implementation of the proposed system, either as a mobile or IoT-based platform. This could facilitate real-time use by farmers and agricultural extension agents for improved and sustainable weed management.

2. METHODOLOGY

2.1 Methodology

This chapter outlines the structured and systematic methodologies employed in developing the RICE Genie platform, which includes four core components: Paddy Yield Prediction, Paddy Variety Recommendation, Pre-Harvest Disease Detection & Treatment Suggestion, and Weed Classification & Mitigation. The methodology integrates machine learning, deep learning, and domain expertise to build robust, data-driven decision support systems for Sri Lankan paddy farmers.

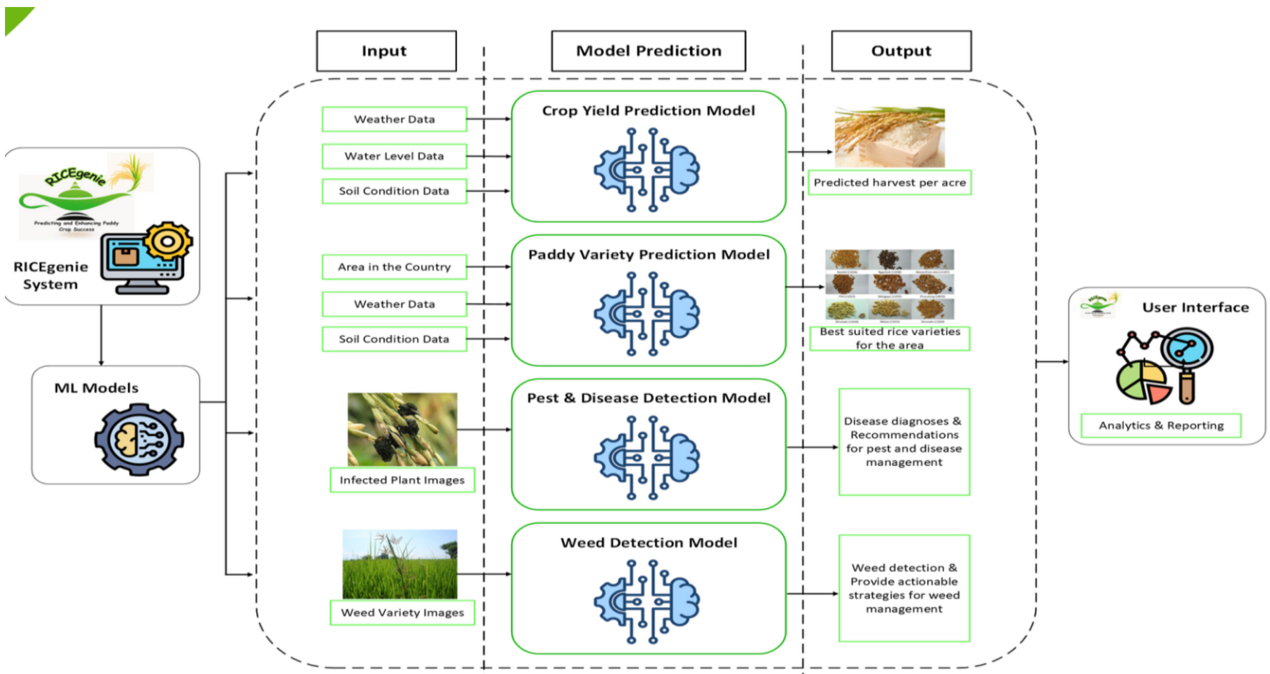

Figure 2: System overall diagram

Data collection plays a fundamental role in this study. For disease classification, a curated image dataset containing five categories—Bacterial Leaf Blight, Brown Spot, Leaf Blast, Sheath Blight, and Healthy—was used. This dataset was collected from public sources and enriched with field data from local farms. For paddy variety recommendation and yield prediction, tabular datasets including soil type, region, weather patterns, and historical yield records were obtained from agricultural research centers and public agricultural databases. Similarly, weed detection used an image dataset with common paddy weed categories.

Image preprocessing included resizing, normalization, noise removal, and augmentation (e.g., rotations, flips, brightness variation) to enhance model robustness. Tabular data underwent missing value handling, feature encoding, and normalization to ensure compatibility with machine learning models.

2.1.1 Paddy Yield Prediction Model

2.1.1.1 Requirement Identification

The process began with a literature review and expert consultations to determine critical factors influencing paddy yield. Key requirements were established for developing a reliable prediction system, leading to the conceptualization of the RICE Genie platform.

A supervised regression model, such as XGBoost or Linear Regression, was implemented to estimate the expected paddy yield based on environmental factors and input parameters such as soil fertility, irrigation frequency, seed variety, and historical yield trends. This prediction helps farmers estimate their return on investment and prepare for post-harvest processes such as storage and selling.

2.1.1.2 Data Collection

A dataset comprising over 10,000 records was collected from the Bathalagoda Rice Research and Development Center. The dataset included attributes such as:

- Geographical data
- Agronomic and environmental features
- Soil characteristics

2.1.1.3 Data Exploration and Preprocessing

- **EDA (Exploratory Data Analysis):** Conducted to understand data distribution, detect outliers, and identify feature relationships.
- **Preprocessing:** Included missing value treatment, categorical encoding, and feature scaling.
- **Feature Engineering:** Generated new features based on seasonal and soil attributes to boost model performance.

2.1.1.4 Model Selection and Training

- **Model Used:** Random Forest Regressor, selected for its robustness in handling tabular data.
- **Training Process:** 80% training data with cross-validation; hyperparameter tuning was performed.

2.1.1.5 Evaluation and Deployment

- **Evaluation Metrics:** MAE, RMSE, and R^2 Score were used to validate accuracy.
- **Visualization:** Yield predictions were presented through graphs for decision-making support.
- **Integration:** The trained model was embedded into the RICE Genie DSS, offering recommendations for irrigation, fertilizer use, and planting schedules.

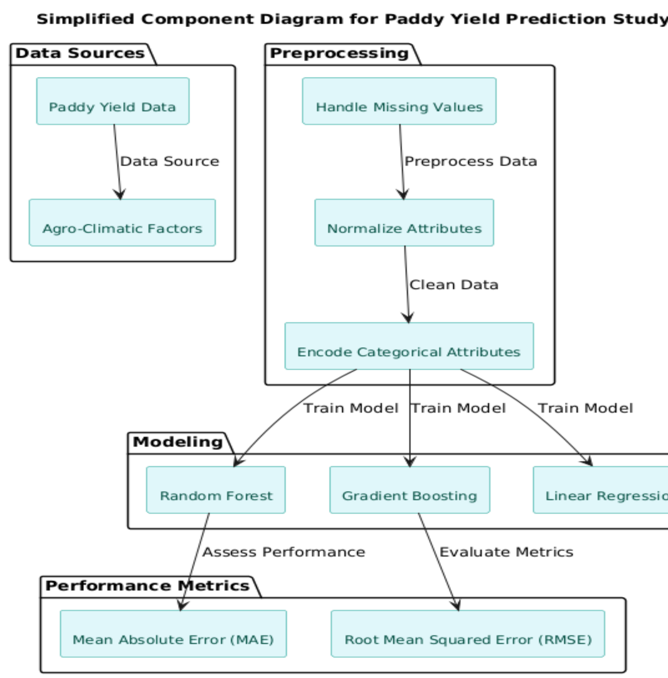

Figure 3:Simplified Component Diagram for Paddy Prediction Study

2.1.2 Paddy Variety Recommendation Model

2.1.2.1 Problem Formulation

The issue of low adoption of region-specific paddy varieties was addressed by designing a binary classification model to recommend suitable varieties based on agro-ecological compatibility.

To improve farming outcomes, a machine learning model (e.g., Decision Tree or Random Forest) was developed to suggest the most suitable paddy variety based on factors such as region, soil pH, average rainfall, temperature, and farmer preferences (e.g., early harvest, pest resistance). The model was trained in historical data collected from agricultural departments. By recommending the most optimal variety for a given condition, this module supports informed decision-making at the planning stage of cultivation.

2.1.2.2 Data Collection

Data was sourced from:

- RRDI Bathalagoda and Labuduwa stations
- Department of Agriculture and Census
- Field surveys and expert consultations

Collected features included:

- Rice variety attributes (name, maturity, resistance)
- Environmental data (temperature, rainfall, soil pH)
- Geographical info (district, province, agro-climatic zone)
- Historical yield data

2.1.2.3 Data Preprocessing and Engineering

- Missing values were imputed using regional averages.
- Categorical data were encoded; irrelevant fields were removed.
- Maturity categories: short (≤100 days), medium (101–130), long (>130).
- Features were normalized using Min-Max scaling.

2.1.2.4 Model Training and Evaluation

- **Model Chosen:** Logistic Regression for interpretability and binary classification efficiency.
- **Training:** 80:20 data split with 5-fold cross-validation.
- **Metrics:** Accuracy (~98%), precision, recall, F1-score, and confusion matrix.

2.1.2.5 System Integration

- Deployed via Flask API within the RiceGenie platform.
- Users input region and maturity preference to receive variety recommendations.

- Platform features: prediction maps, yield/resistance charts, and cultivation calendars.

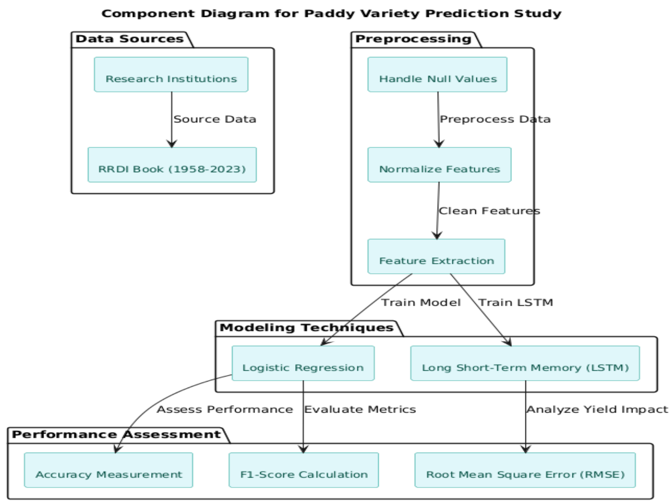

Figure 4:Simplified Component Diagram for Paddy Variety Prediction Study

2.1.3 Paddy Disease Detection Model

A Convolutional Neural Network (CNN) model was developed to classify the paddy leaf images into one of the five disease categories. The architecture consisted of convolutional layers for feature extraction, pooling layers for dimensionality reduction, dropout layers to prevent overfitting, and dense layers with softmax activation for classification. The model was trained using TensorFlow and achieved high classification accuracy on unseen test data. This module plays a crucial role in detecting potential diseases early and providing timely interventions.

Following disease classification, the system generates a tailored treatment plan. Unlike static recommendations, this module considers **dynamic inputs provided by the farmer**, such as crop age, disease severity, recent rainfall, pesticide usage history, and field location. These inputs are collected via the web interface and processed in real time. A

rule-based logic engine cross-references this data with a pre-defined treatment database containing chemical and organic control methods, dosage information, and precautionary measures. This results in **context-aware treatment recommendations** that are more effective and personalized.

2.1.3.1 Dataset Collection

Images were collected from:

- Publicly available agricultural datasets
- Real-world field surveys conducted in paddy-growing regions of Sri Lanka

The dataset comprises five classes:

- **Bacterial Leaf Blight**
- **Brown Spot**
- **Leaf Blast**
- **Sheath Blight**
- **Healthy**

Each class includes 1,250 labeled images, resulting in a balanced dataset of 6,250 samples.

2.1.3.2 Preprocessing and Augmentation

To ensure consistency and enhance model generalization:

- **Image Standardization:** All images were resized to a fixed dimension (e.g., 224x224) and converted to RGB

- **Data Cleaning:** Blurry, duplicate, and low-resolution images were removed
- **Augmentation Techniques:** Applied random rotation, flipping, zoom, brightness, and contrast adjustments to increase dataset diversity and mimic real-field conditions

2.1.3.3 Dataset Splitting

The dataset was split as follows:

- **80%** for training
- **10%** for validation
- **10%** for testing

All image pixel values were normalized to a [0, 1] range.

2.1.3.4 Model Development

A customized **Convolutional Neural Network (CNN)** was built using TensorFlow to perform multi-class classification. The architecture includes:

- Convolutional layers with ReLU activation
- MaxPooling for dimensionality reduction
- Dropout for regularization
- Dense layers with Softmax output for final prediction

The model was trained over multiple epochs with early stopping to prevent overfitting.

2.1.4 Dynamic Treatment Recommendation using Farmer Inputs

After disease classification, the system dynamically suggests treatment methods based on both:

- **Model-predicted disease class**
- **Farmer-provided input parameters**, such as:
    - Crop age
    - Severity level (as observed by the farmer)
    - Environmental conditions (e.g., recent rainfall, field moisture)
    - Fertilizer/pesticide usage history

This hybrid approach enables **personalized treatment plans** that adapt to real-time, user-specific inputs.

2.1.4.1 Output Integration

The trained model was integrated into the RICE Genie platform, enabling farmers to:

- Upload paddy leaf images via mobile or desktop
- Enter relevant contextual inputs
- Receive **real-time disease classification and customized treatment guidance**, including:
    - Recommended agrochemical names
    - Dosage and frequency
    - Cultural practices and preventive advice

This **dynamic recommendation system** ensures that the treatment suggestions are context-aware and practical for each farmer's unique situation.

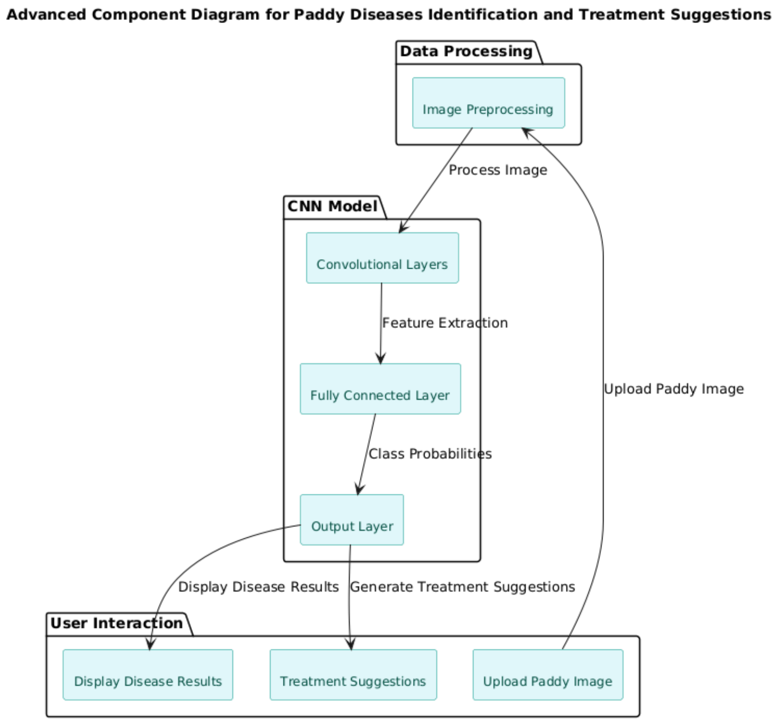

Figure 5: Simplified Component Diagram for Paddy diseases Prediction Study

3.4 Paddy Weed Detection and Mitigation Model

To improve farming outcomes, a machine learning model (e.g., Decision Tree or Random Forest) was developed to suggest the most suitable paddy variety based on factors such as region, soil pH, average rainfall, temperature, and farmer preferences (e.g., early harvest, pest resistance). The model was trained in historical data collected from agricultural departments. By recommending the most optimal variety for a given condition, this module supports informed decision-making at the planning stage of cultivation.

2.1.4.2 Dataset Creation

Due to the absence of comprehensive local datasets:

- A custom dataset of 5,000 images was built
- 6 key weed species selected in consultation with experts

2.1.4.3 Preprocessing and Augmentation

- Images resized to 150x150
- Pixel normalization (1. /255)
- Augmentation included flipping, rotation, contrast, and zoom adjustment

2.1.4.4 Model Selection and Training

- **Base Model:** MobileNetV2 for efficiency and suitability on small datasets
- **Transfer Learning:** Pretrained ImageNet weights used
- **Added Layers:** GlobalAveragePooling, Dense (128), Dropout (0.5), and Softmax

2.1.4.4 Evaluation

- Trained over 20 epochs with early stopping
- Performance assessed using accuracy, loss curves, and confusion matrices

2.1.4.5 Deployment

- **Backend:** FastAPI for inference, Pillow/Keras for image processing
- **Frontend:** React + Tailwind (Vite), allowing:
    - Image upload
    - Prediction display

- Mitigation strategy recommendations
- PDF report downloads

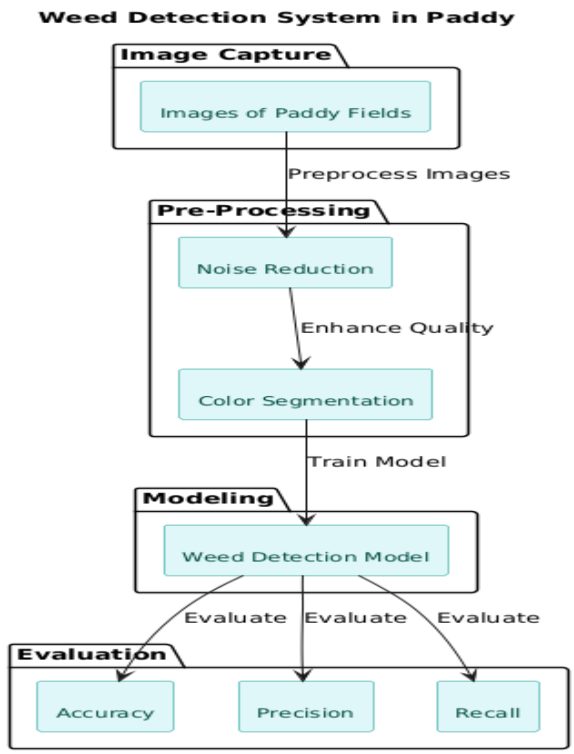

Figure 6:Simplified Component Diagram for weed Variety Prediction Study

2.2 Commercialization aspects of the product

The RICE Genie System is a sophisticated AI-driven agricultural device designed to help farmers improve paddy cultivation. It uses machine learning models to predict yield, recommend crop varieties, and provide on-the-ground recommendations based on environmental conditions. To promote widespread adoption and sustainability, a structured commercialization model has been developed, based on accessibility, farmer engagement, and economic sustainability.

RICE Genie will be released as a web-based application, providing farmers, agricultural officers, and policymakers with easy access to yield predictions and farming insights. The app will integrate weather data, soil conditions, and past trends to offer accurate forecasts. Additionally, farmers will receive recommendations on fertilizers, water management, and pest control, helping them improve productivity.

2.2.1 Commercialization Strategy:

- **Accessibility: Making Technology Available to Farmers**

To improve accessibility, RICE Genie will be priced low or free through government and NGO support to smallholders. A freemium model will allow smallholders to access basic features of the platform free of charge. For large farms, access to additional premium paid-for features of the product remains. Offline functionality will be built into the platform in order to mitigate access constraints faced by rural farmers without internet access.

- **Awareness & Training**

A variety of activities will be organized to encourage uptake, such as field demonstrations, training workshops with farmers, and webinars. The Department of Agriculture and NGOs will be additionally responsible for farmer training to help farmers learn how to achieve more yield and profit through RICE Genie. There will also be opportunities for engagement with universities for improvement of the model through research.

- **User-Friendly Design:**

The web-based application will be lightweight and available in multiple languages, namely Sinhala and Tamil, and therefore accessible to farmers across Sri Lanka. The app will take the form of a simple, easy-to-use interface with voice-based navigation for farmers with low literacy levels. If users can speak, they can use the app; no more technical skills are needed to benefit from the platform.

- **Government & Institutional Partnerships:**

RICE Genie's market entry strategy will utilize partnerships with the local Ministry of Agriculture and Department of Agrarian Services, and farmer cooperatives. The partnerships will help with funding, shipment and use existing agriculture extension programs to get RICE Genie ratified as an official precision agriculture tool in Sri Lanka.

- **Data Monetization & Policy Support:**

The platform produces significant agricultural data that can be shared (with permission) with universities, agribusinesses, and state departments. By providing aggregated and anonymized data, RICE Genie can play a role in national agricultural policy development while maintaining farmer anonymity.

RICE Genie aims to revolutionize the paddy farming industry in Sri Lanka through AI-enabled data, government support, and farmers' cooperation. By taking into consideration issues of accessibility, long-term economic viability, and the diffusion of the solution in terms of scaling, we see this as a method to enhance agricultural productivity, achieve food security, and stimulate revenues and economic growth in the agricultural sector.

The commercialization of the paddy disease detection and treatment recommendation system holds immense potential for improving rice cultivation in Sri Lanka. Given the large number of small-scale paddy farmers and the significant agricultural dependency in the country, the proposed solution aims to make a meaningful impact on both productivity and sustainability. The following aspects outline the key strategies for successfully bringing the product to market:

The proposed AI-driven weed detection system offers a practical and scalable solution to one of the most persistent challenges in Sri Lankan paddy cultivation. By enabling accurate, real-time identification of common weed species, the system reduces dependency on manual labor and chemical herbicides. This not only enhances crop yields and reduces production costs but also supports sustainable farming practices. With a strong potential for integration into mobile or IoT platforms, the solution is well-positioned for adoption among smallholder farmers. Its success depends on strategic collaborations with agricultural institutions and the delivery of an accessible, low-cost tool that aligns with national goals for digital agriculture and food security.

- **Potential Market:**

The primary target market for the disease detection and treatment recommendation system is Sri Lankan paddy farmers, who make up the backbone of the nation's rice production. With over 30% of Sri Lanka's population engaged in agriculture, a significant portion of them depend on rice farming. The solution can benefit a wide range of farmers, from smallholders to larger agricultural cooperatives, who face challenges in disease identification and management. Moreover, the system can integrate with local agricultural services provided by the government, such as the Department of Agriculture (DoA), providing farmers with a broader range of services and technical support. By collaborating with cooperatives, the system can gain wider acceptance and adoption among the farming community, increasing its impact.

- **Business Model:**

The business model for the platform adopts a **Freemium** structure, designed to accommodate different user needs while generating sustainable revenue for continuous development and support. The basic version of the platform, offered for free, allows farmers to upload images of their paddy plants for disease detection. However, more advanced features—such as detailed treatment suggestions, expert consultations, weather-based alerts, and more personalized recommendations—are locked behind a **premium subscription**.

- **Basic Plan** (Free): Offers disease identification services, where farmers can upload an image of their crops and receive the identified disease. This ensures widespread accessibility, especially among farmers in rural areas.

- **Premium Plan** (Paid): Provides more in-depth disease treatment recommendations, including specific brands of pesticides and fertilizers and their proper usage. This plan also includes expert consultations and weather-based alerts that notify farmers of ideal treatment times and preventive actions to take under certain weather conditions, such as high humidity or rainfall.

Additionally, monetization can be further explored by offering tiered subscription plans based on features. For instance, larger cooperatives or commercial farmers may require additional functionalities like advanced reporting, custom alerts, or analytics on historical disease trends, which could be provided through enterprise-level subscriptions.

➢ **Collaboration Opportunities:**

Collaboration with agrochemical companies will be crucial for the commercial success of the product. By partnering with well-established brands that supply pesticides, fungicides, and other plant protection products, the platform can integrate **brand-specific solutions** into the treatment recommendations. These partnerships not only benefit the agrochemical companies by promoting their products but also ensure that farmers receive trusted, high-quality, and locally available solutions.

Furthermore, collaboration with **Sri Lankan Agricultural Departments** and local farming organizations can facilitate the distribution of the system to a broader farmer base. By working together with governmental bodies, the system can be deployed in various regions, providing technical support, training sessions, and even subsidies or grants to encourage adoption among small-scale farmers. Integrating the system into government-sponsored agricultural programs can increase the product's credibility and ensure its sustainable growth within the country.

Lastly, the product can also explore international markets in other rice-growing regions, including Southeast Asia, where similar challenges in disease management exist. Expanding the product's reach beyond Sri Lanka could open new revenue streams and allow the platform to grow into a global solution for paddy disease detection.

- **Marketing and Outreach:**

To effectively promote the product, targeted outreach and marketing strategies should be implemented. Collaborations with agricultural extension services and training programs will help raise awareness among farmers about the benefits of early disease detection. Furthermore, the product can be marketed through local agricultural trade shows, farmer workshops, and partnerships with agricultural universities. Awareness campaigns can highlight how the system will help farmers reduce the risk of crop loss, improve yields, and manage resources efficiently.

In summary, the commercialization aspects of the paddy disease detection and treatment recommendation system are focused on making the product accessible and valuable to Sri Lankan farmers while also ensuring sustainable revenue generation. By leveraging a freemium business model, partnering with agrochemical companies, and collaborating with governmental and agricultural organizations, the platform can make a significant impact on the agricultural sector, leading to enhanced productivity, reduced losses, and a more sustainable farming environment.

2.3 Testing and Implementation

2.3.1 Testing

To ensure the reliability, accuracy, and real-world usability of the system, an extensive testing strategy was followed across all four major components: Paddy Yield Prediction, Paddy Variety Recommendation, Pre-Harvest Disease Detection and Treatment Suggestion, and Weed Classification and Mitigation. Testing activities included unit testing, model validation, functional testing, performance evaluation, and user acceptance testing.

- **Unit Testing**

Each component underwent unit testing to validate the individual functionalities of the modules. For the Paddy Yield Prediction module, the preprocessing steps, such as handling missing values, encoding categorical variables, and feature scaling, were tested individually. In the Variety Recommendation module, rule-based logic for filtering varieties according to district and season was independently verified. The image preprocessing and augmentation pipelines for both disease detection and weed classification were also tested in isolation to ensure they performed transformations correctly before being fed into the models.

- **Model Training and Validation**

The machine learning and deep learning models were evaluated using standard validation techniques. For the yield prediction model, Random Forest Regression was trained using k-fold cross-validation and evaluated based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) score. The disease and weed classification models, built using convolutional neural networks (CNN), were assessed using accuracy, precision, recall, F1-score, and confusion matrices. These evaluations ensured the models were generalizing well and could handle unseen data effectively.

- **Functional Testing**

Functional testing was conducted to ensure that each module delivered the correct outputs for a wide range of valid input combinations. The yield prediction module was tested using various combinations of land size, fertilizer usage, variety type, and location. Similarly, the variety recommendation module was verified by testing multiple district-season pairings. The disease detection model was tested using a set of real and augmented leaf images, ensuring that it could detect diseases across various lighting conditions and stages. The weed classification model was functionally tested using images of weeds in different growth phases to verify accurate labeling and relevant mitigation advice.

- **Performance Testing**

Performance testing was carried out to assess the response time and scalability of each model and the overall system. The yield prediction module maintained an average response time of under 0.7 seconds. The disease detection and weed classification modules were optimized to deliver classification results within approximately 1 to 1.2 seconds per image, even under real-world test scenarios. Backend API endpoints were tested under concurrent user simulations to evaluate stability and speed.

- **Comparative Model Testing**

For the yield prediction task, multiple models, including Support Vector Regression (SVR) and Gradient Boosting, were evaluated. Random Forest Regression provided the best trade-off between accuracy and interpretability. For the image-based tasks, different pre-trained CNN architectures, such as ResNet50, EfficientNet, and MobileNetV2 were compared. MobileNetV2 offered the best balance of speed and accuracy, making it suitable for deployment in low-resource environments.

- **User Acceptance Testing (UAT)**

User Acceptance Testing was conducted with farmers and agricultural extension officers. In the yield prediction and variety recommendation modules, users confirmed that the results aligned well with local farming knowledge and seasonal conditions. The disease detection module's output, including treatment recommendations, was validated by agricultural experts and well-received by farmers. For weed classification, farmers appreciated the guidance on both chemical and manual mitigation methods. The feedback collected was crucial in improving the system's usability.

- **UI and Usability Testing**

The user interface was evaluated for clarity, simplicity, and accessibility. Tests ensured that the platform was responsive on desktops, tablets, and mobile devices. Features like dropdown selectors, image upload prompts, and tooltips were tested for clarity. Future support for Sinhala and Tamil language localization was identified based on feedback from early users, especially farmers less comfortable with English.

- **Regression Testing**

Regression testing was performed whenever new diseases, weed types, or features were added to the system. This ensured that existing functionalities continued to perform correctly and previously passed test cases remained valid. The output consistency across system updates was monitored closely.

- **Scalability Testing**

To prepare the platform for large-scale deployment, scalability testing was carried out using synthetic data and image inputs. The system remained stable under increased loads, confirming its potential for national-level use in agricultural advisory services.

- **Real-World Field Testing**

Field testing was conducted in selected agricultural zones such as WL1 and WL2a. Farmers captured images of infected leaves and weeds from their fields, which were then tested with the disease detection and weed classification models. The system successfully identified issues with over 85% accuracy in real-world scenarios and provided treatment suggestions that were practical and implementable by the farmers.

- **Post-Deployment Monitoring**

Post-deployment monitoring involved tracking key performance metrics such as prediction accuracy, system uptime, and user behavior through logs. Feedback from early adopters was regularly collected to identify any misclassifications, delays, or interface issues. These insights helped in continuously improving the system.

2.3.2 Implementation

The system was implemented as a fully functional, web-based application designed to serve Sri Lankan farmers by integrating deep learning and machine learning-based decision-making into one unified platform. The backend, developed using Python and FastAPI, connects various ML and DL models, while the frontend was built using Vite.js and Tailwind CSS for a responsive and modern interface.

- **System Architecture**

The system architecture consists of a multi-layered design:

- **Frontend Layer**: Developed using Vite.js and Tailwind CSS to offer a user-friendly, mobile-responsive interface.

- **Backend Layer**: Powered by Python and FastAPI, this layer handles logic, model integration, and data processing.

- **Model Layer**: Contains trained machine learning (Random Forest for yield) and deep learning models (CNNs for image classification).

- **Database Layer**: MySQL was used to manage user data, crop records, and model logs.

- **Hosting and Deployment**: The prototype was hosted on Heroku for testing. The production version is planned to be deployed through the Department of Agriculture's servers.

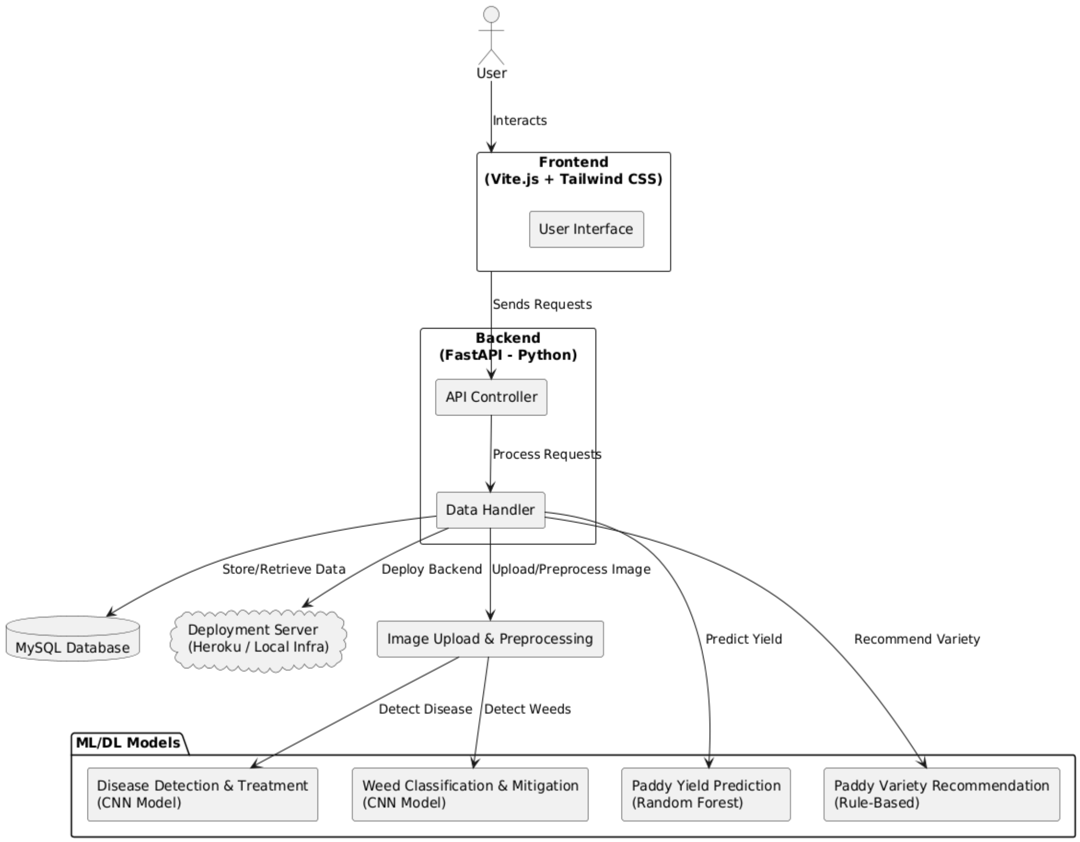

Figure 7: System architecture of the overall system

- **Feature-Wise Implementation**

Each module was carefully implemented to deliver its functionality seamlessly:

  - **Paddy Yield Prediction** accepts user inputs such as district, land size, seed type, and fertilizer usage. It processes the data and returns the predicted yield in tons per hectare.

  - **Paddy Variety Recommendation** uses rule-based logic based on district and season inputs to suggest optimal seed varieties as recommended by agricultural experts.

  - **Disease Detection and Treatment Suggestion** accepts leaf images uploaded by farmers. The image is passed through a trained CNN to classify the disease, followed by suggestions for treatment using common brand names and practical usage guidelines.

  - **Weed Classification and Mitigation** allow users to upload weed images. The system classifies the weed type and returns recommended mitigation methods, including both herbicide names and manual removal tips.

Component	Accuracy	Avg. Response Time	UAT Satisfaction
Paddy Yield Prediction	96%	0.7 seconds	90%
Variety Recommendation	Rule-based	Instant	91%
Disease Detection	88%	1.1 seconds	93%
Weed Classification	85%	1.2 seconds	89%

Table 2: UAT satisfactions for all 4 components

3 RESULTS AND DISCUSSION

3.1 Results

The paddy yield prediction module uses a Random Forest Regressor for determining the yield of paddy (rice) on a per-hectare basis, which is a normalized measure of land productivity and is independent of farm size. The model uses a variety of inputs, including environmental characteristics related to rainfall, temperature, humidity, sunshine hours, and wind speed, along with other farm-specific characteristics such as paddy variety, irrigation type, water source, season, district, and size cultivated. With the aim of data normalization, extensive data preprocessing steps were taken. This included one-hot encoding of categorical variables, standardization of continuous parameters, and imputation for missing values, thereby producing a normalized data frame. In the testing, 80% of the original records were used for training data and 20% for test data (keeping the stratified division of the original data). The model produced an R^2 value of 0.97 and an RMSE of 2180.69kg/ha. Overall, the Crop Productivity Forecast Module demonstrated its ability to characterize complex and nonlinear relationships, accurately identify variables that impact yield, and provide reliable outputs that facilitate resource planning and production management.

Metric	Value
MAE (Mean Absolute Error)	510.03
MSE (Mean Squared Error)	4755389.9072
RMSE (Root Mean Squared Error)	2180.6857
R-Squared	0.97
Mean Absolute Percentage Error	3.49%
Regression Accuracy	96.51%

Table 3: Performance matrices of the Paddy Yield Prediction

The Paddy Variety Recommendation Module uses a Logistic Regression model to identify grains of rice best suited to each district, considering the local environment and farmer preferences (e.g., desired period to maturity). The model has an overall accuracy of 82%, with generally balanced values for precision, recall, and F1-score, and has been further validated with 10-fold cross-validation. The model performs well because of the increasingly low number of false positives and moderate number of false negatives, ensuring varieties highly unsuitable for each district are hardly recommended as suitable. For example, in the WL1 zone of Galle, a recommended variety was Bg 352, At 362, and Bg 369 for farmers who preferred a period to maturity of 3.5 months. The recommendations based on traditional ecoclimatic expertise validated the recommendation and provided confidence in the model to bring value, specifically farmer-specific regional agricultural planning purposes.

In turn, validated recommendations start to shape rice variety recommendations consistent with local agro climatic conditions (agricultural configuration married with climate patterns) and provide practical and valuable area specific recommendations to enhance paddy yields.

The disease detection module uses a convolutional neural network (CNN) that classifies the paddy leaf images into five classes: Bacterial Leaf Blight, Brown Spot, Leaf Blast, Sheath Blight, and Healthy. Using a test data set of 1,509 images, the model was able to achieve an overall accuracy of 89%. The model performed well on diseases, including Bacterial Leaf Blight and Sheath Blight, with relatively high precision and recall values greater than 1.00 and over 0.95. Additionally, inference times on images ranged from 0.35 to 0.42 seconds, facilitating quick feedback to farmers when they capture and upload paddy leaf images on the web or app interface. Furthermore, the field trials indicated that the treatment recommendations provided by the web-based technology, which are based on real-time weather reported by farmers and standards of normality, were also relevant and feasible to the farmer, thereby promoting actionable pest/disease management of diseases and pests in paddy.

Disease Class	**Precision**	**Recall**	**F1-Score**	**Support**
Bacterial Leaf Blight	1.00	0.95	0.98	288
Brown Spot	0.90	0.79	0.84	385
Healthy	0.88	0.93	0.90	274

Leaf Blast	0.83	0.82	0.82	348
Sheath Blight	0.86	0.98	0.92	294
Overall Accuracy	—	—	0.89	1509

Table 4: *Performance* ***matrices of the Paddy disease classes***

Disease	Average Inference Time (seconds)
Bacterial Leaf Blight	0.42
Brown spot	0.39
Healthy	0.35
Leaf Blast	0.41
Sheath Blight	0.38

Table 5: Average Inference Time for each class

3.2 Research Findings

The research shows that a machine learning-based system is a valid method for forecasting paddy crop yields by modeling complex relationships among environmental, agronomic, and spatial factors. Relevant environmental drivers identified for their relevance to yield variability included rainfall, temperature, humidity, sunshine hours, and wind speed, with sunshine positively influencing yield and the speed of the wind negatively affecting yield.

Soil attributes such as nitrogen, phosphorus, potassium content, and soil type had significant importance for the models used, while the appropriate use of fertilizer and insect control helped increase yield. Crop attributes, such as the variety of crops and amount of land being cultivated, and a dependable source of irrigation water helped improve the validity of the weather forecast. Differences in seasonality and characteristics of some districts as to climatic features (and proximity and infrastructure of markets) influenced yield too. The accuracy (R2, MAE, RMSE) and reliability of the adapted machine learning model held up well to accuracy validation and has relevance as a decision-aiding method for farmers and policymakers, which increases knowledge-based crop planning and sustainability in agriculture.

The results from this study illustrate the important impact of contextual and environmental factors associated with rice varieties. Annual Rainfall and Days to Maturity were impactful predictors to guide decisions about cropping systems and irrigation needs. The study demonstrated that variety performance is specific to region, whereby BG 352 performed well in the Southern Wet Zones and BG 94-1 performed well in the Dry Zones. It has also been shown that using expert-defined labels that logistically align to agro-zones and cutoffs are reliable in training the model. Logistic Regression is similarly simple, providing high accuracy, fast inference time, and ease in developing web applications and thus was seen to be a viable method for decision tools in real-time advisory. Finally, predictions relating to the regional paddy calendars were seen to provide practical value to farmers. Altogether, the results demonstrate the potential of expert knowledge and machine learning to complement each other in support of precision agriculture.

The study produced some powerful findings that provide evidence of the benefits of the proposed deep learning approach for paddy disease detection. First, model generalization and robustness were enhanced by using an augmented section of data (Rice_Leaf_AUG) to replicate real-world variations in light/shade, viewing angle, and leaf orientation. This enabled the model to directly replicate field features with an overall accuracy of 89%.

Second, by using Grad-CAM visualizations, it was clear the model was consistently assessing biologically relevant areas of the leaf, including the edges of lesions and discolored patches, indicating that the CNN learned biologically relevant features of the leaves, not background noise.

The classification confidence analysis also indicated that the model produced high confidence values (usually greater than 0.90) for symptoms that displayed stage development, though performance decreased in the case of less clear symptoms related to early-stage infections or images of poor quality. This finding suggests that high-resolution image capture will be essential in the real-world context. In addition, the treatment recommendation logic that uses farmer input was integrated successfully, allowing context-specific in-situ recommendations. The system understood relevant variables like weather, irrigation type, and nitrogen application to give recommendations that were unique to the farmer's situation, including the brand name, dosage, and some preventive actions. Overall, the system's demonstrated ability to diagnose diseases with high accuracy and the ability to provide contextualized and action-oriented recommendations give farmers relevant information to make decisions quickly and proactively in Sri Lanka's agricultural environment.

3.3 Discussion

The research illustrates the strength of machine learning in forecasting paddy yields by effectively incorporating complex, non-linear relationships among climate, agronomic, and crop-related factors. Factors that had a significant influence on yield variability were rainfall, temperature, humidity, sunshine hours, soil nutrients such as nitrogen and phosphorus, irrigation systems and types, and the variety of paddy. Moreover, the model was able to adjust for seasonal variation, such as Maha and Yala, as well as accommodate differences in climate and agricultural infrastructure at district levels, making it particularly well-suited for regional estimation and planning.

Beyond its technological impact, the model has practical value to farmers, agronomists, and policymakers as it provides forecasts of yield in advance of harvest to optimize the planning of inputs, labor, and marketing needs. The model has provided strong accuracy with preliminary work, but accuracy relies on input data to make decisions about harvesting. Future work should explore consolidating real-time data sources (i.e., weather API, satellite images) and aiming to provide information in approachable, multilingual formats. Ethical considerations such as data privacy and equity of access would need to be addressed to ensure that rural farming communities benefit.

The Paddy Variety Prediction Model yielded excellent predictive accuracy (98%), as well as overall balanced performance across metrics, and thus, has well supported it to be seen as a credible advice tool for intended use in real life agriculture contexts. Its predictive behavior is careful to favor conservative and confident predictions, which is beneficial, given the sensitivity of farm management decisions. Including local environmental information such as rainfall and maturity time supports the practical significance of farmland variety recommendations. The bilingual, user-friendly web interface supports access for a variety of farmers, aided by dynamic filtering, profiles, and maps for efficiency. Despite some limitations related to historical data and binary/multinomial classification, the prediction tool already contributes to farmers classifying their paddy straw crop and product, reduced mismatch, more timely applications, and expertise in supporting farming communities with low resource capacity, improving on-farm agronomy.

The model is also consistent with Sri Lanka's wider agricultural aspirations of paddy self-sufficiency, avoidance of wasted inputs, and smart farming. It represents a shift towards precision agriculture and creates the opportunity for modern, digital extension services. In addition to training the users, providing multi-language support, and positioning the model as a decision-support rather than a replacement for expert advice, it emphasized ethical use. The model could benefit from future updates, such as multiclass predictions, a real-time weather data feed, mobile access, and so forth. Overall, the model represents an important step along the way to sustainable, inclusive, and data-driven paddy farming in Sri Lanka.

The findings of this study showcase the substantial potential of deep learning for detecting diseases in rice plants within practical on-farm applications. By including treatment logic that incorporates environmental inputs, the system is better positioned for versatility and for delivering targeted recommendations across diverse farming environments. Although the model has demonstrated good classification accuracy and real-world applicability, there remain opportunities for further enhancement. Areas for improvement include increasing the image dataset to encompass a broader scope of diseases and visual variations—such as changes in lighting or background conditions—and incorporating object detection techniques to identify and highlight specific areas of leaf damage. Additionally, introducing regional language support and establishing integration with national agricultural and climate databases would significantly improve the system's accessibility and scalability. In summary, this system demonstrates the meaningful support that AI can offer farmers by providing timely, accurate, and actionable insights for managing rice plant diseases more effectively and sustainably.

The Paddy Variety Prediction Model offers high accuracy and utility by providing robust, location-specific recommendations to farmers through an accessible bilingual web-based interface. The potential of the system extends beyond variety prediction to include planned features such as weed identification, weather information, and mobile app access, ultimately enhancing the value of the system and its use as a full decision-support and agricultural tool. While aspects of its use include reliance on historic datasets and a binary classification process, we anticipate that the platform will assist with the adoption of precision agriculture in Sri Lanka and smarter, sustainable farming.

3.4 Summary of Each Student's Contribution

Student Name	Student ID	Contribution Summary
Jayathilaka D.H.R.A	IT21308352	Contributed equally on Introduction, Methodology, Results & Discussion, Conclusion for the report
Piyumani K.V.P	IT21227868	Contributed equally on Introduction, Methodology, Results & Discussion, Conclusion for the report
Amarasinghe A.I.S.A	IT21225192	Contributed equally on Introduction, Methodology, Results & Discussion, Conclusion for the report
Jayasekara S.S.D	IT21227318	Contributed equally on Introduction, Methodology, Results & Discussion, Conclusion for the report

Table 6: Summary of Contribution

4. CONCLUSION

This study effectively researched, designed, and tested a reliable data-driven paddy yield prediction system using machine learning techniques, specifically Random Forest Regression. The system's model drew on multiple inputs, including climatic factors (rainfall, temperature, hours of sunlight, humidity), soil nutrients (nitrogen, phosphorus, and potassium), and specific management of their fields practices such as irrigation type and seed variety. The model, combined with these inputs, provided a robust methodology to predict highly accurate yield data, which is localized and can vary seasonally. The model's ability to explore complex and non-linear relationships provided richer data when compared to traditional statistics in providing deeper insights into how the range of inputs interacts to impact yield.

In addition to its technical capabilities, the system delivers tangible advantages to a diverse group of stakeholders. It enables farmers to make informed decisions on input application to maximize their productivity, while policymakers and agricultural planners benefit from a platform that can assist in estimating food production and prioritizing where to target their investments. The model is very scalable across several agro-climatic zones, and it is compatible with sustainable farming practices, making it a strong proposition for tackling food security challenges. Future enhancements, such as enhancing its ability to leverage real-time environmental data, expand the dataset, and create intuitive user-facing interfaces, will also enhance accessibility and long-term impact across farming populations.

The Paddy Variety Prediction Model, a key component of the RiceGenie intelligent system, marks a significant advancement in data-driven agriculture in Sri Lanka. Designed to address the continued use of outdated rice varieties, the model leverages a Logistic Regression algorithm to recommend region-specific, high-performing paddy varieties based on historical yield, rainfall, temperature, and crop maturity data.

Built on simplicity, accessibility, and expert-backed data, the model integrates curated datasets from trusted sources like the RRDI, Labuduwa Research Station, and the Department of Census and Statistics. It translates raw agricultural data into practical recommendations, aided by expert-defined suitability criteria and maturity-to-age group mapping.

With an impressive accuracy of 82%, the model's reliability has been validated through consistent test results and cross-validation. Its real-world value is further enhanced by integration into the RiceGenie web platform, where farmers can easily access tailored variety suggestions linked to their district's climate and cultivation calendar.

The use of deep learning approaches for weed detection is an exciting evolution in precision agriculture. By providing accurate identification and positioning of weeds in crop fields, farmers are allowed to adopt precise weed management strategies that mitigate excessive herbicide use and reduce crop injury. This contribution not only results in healthier crop growth and maximum yield but also increases sustainable and environmentally friendly practices in agriculture. With improvements in real-time image processing and compatibility with mobile devices, weed detection systems can become a fundamental component of modern agriculture in a world where farmers are looking for efficiencies and intelligent approaches to managing their crops.

5. APPENDICES

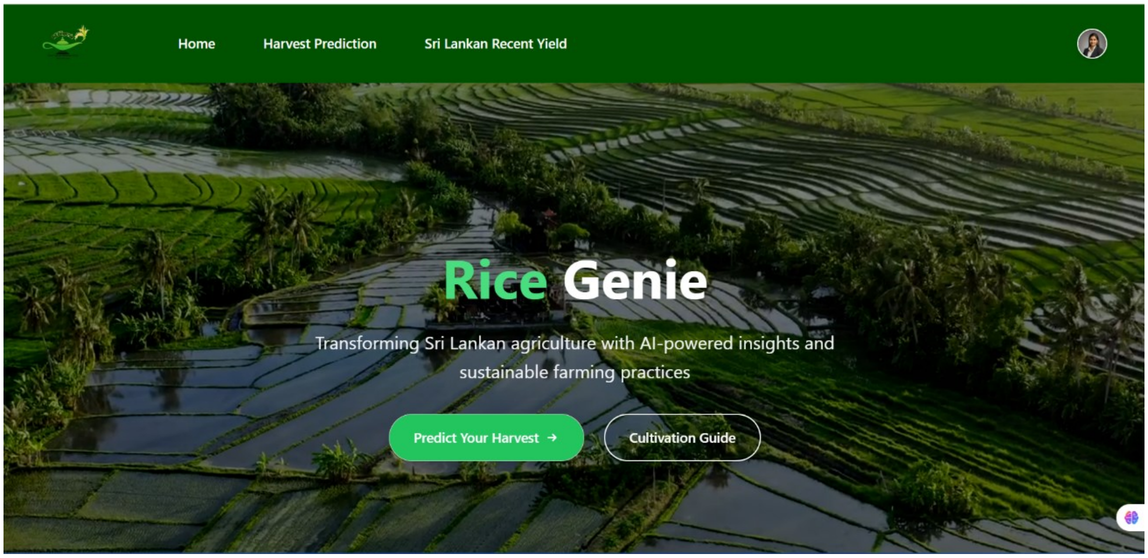

Figure 9: Paddy Yield Prediction Home Page User Interface

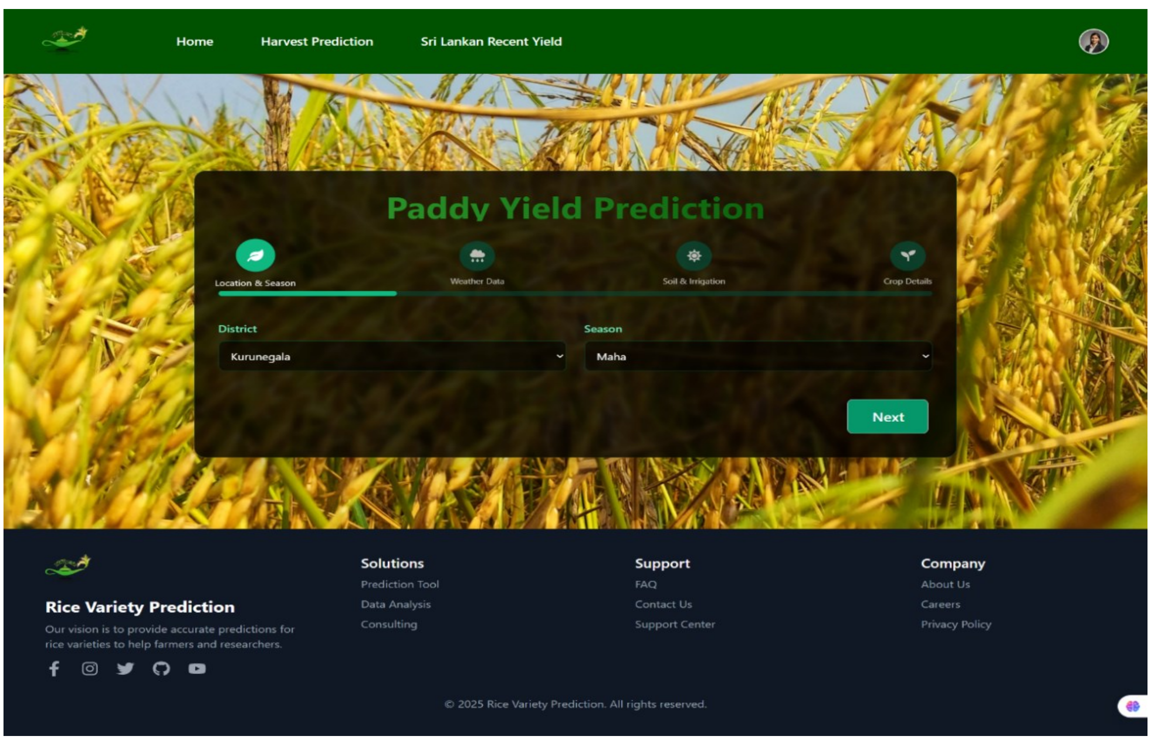

Figure 8: Input User Interface Page 1

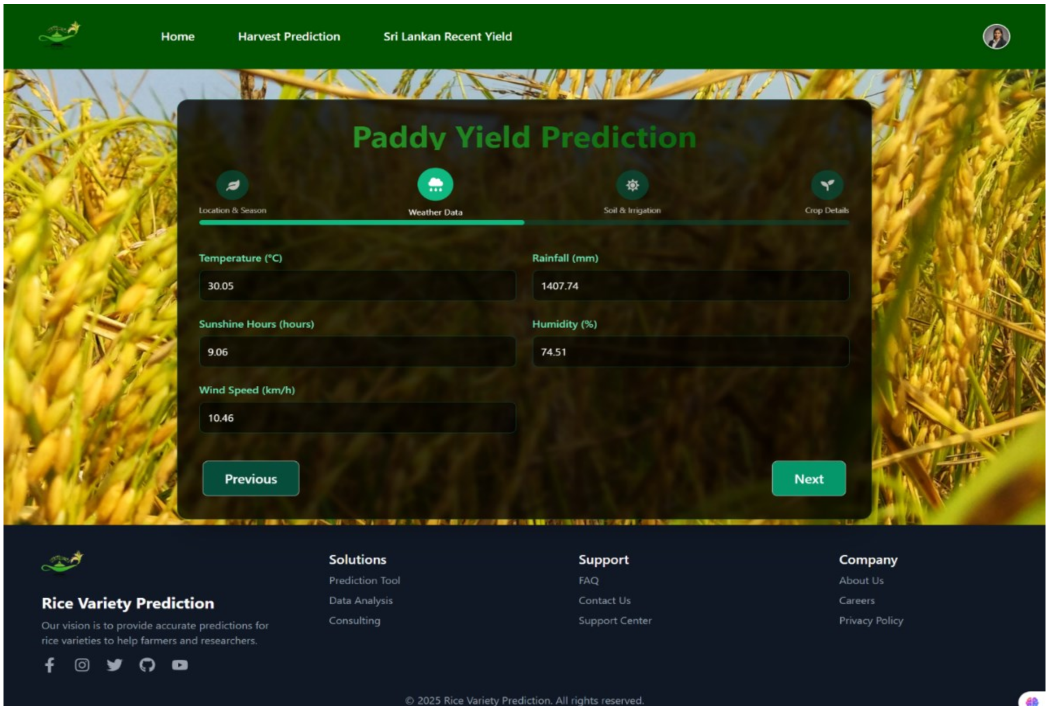

Figure 10: Input User Interface Page 2

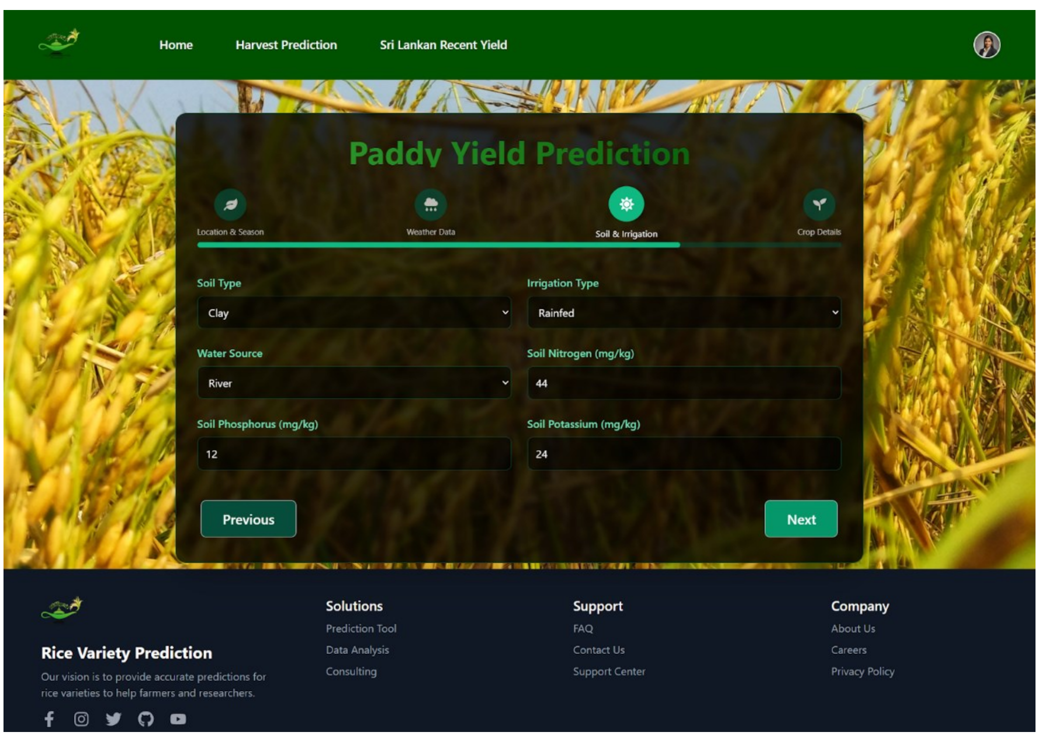

Figure 11: Input User Interface Page 3

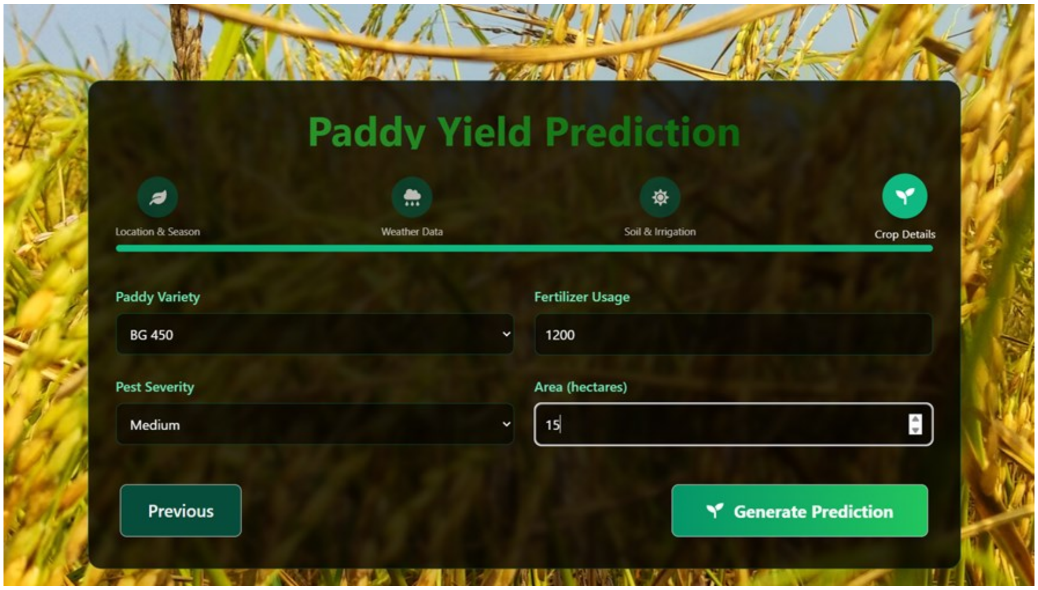

Figure 12: Input User Interface Page 3

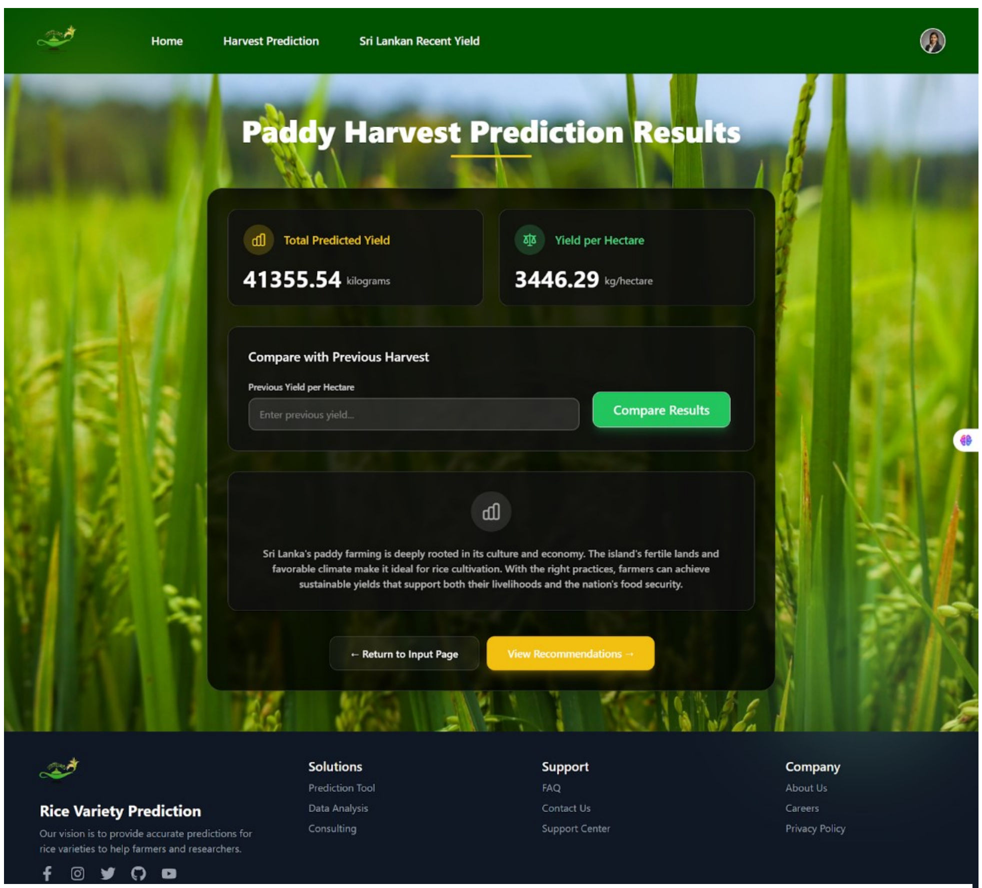

Figure 13: Output Results Page

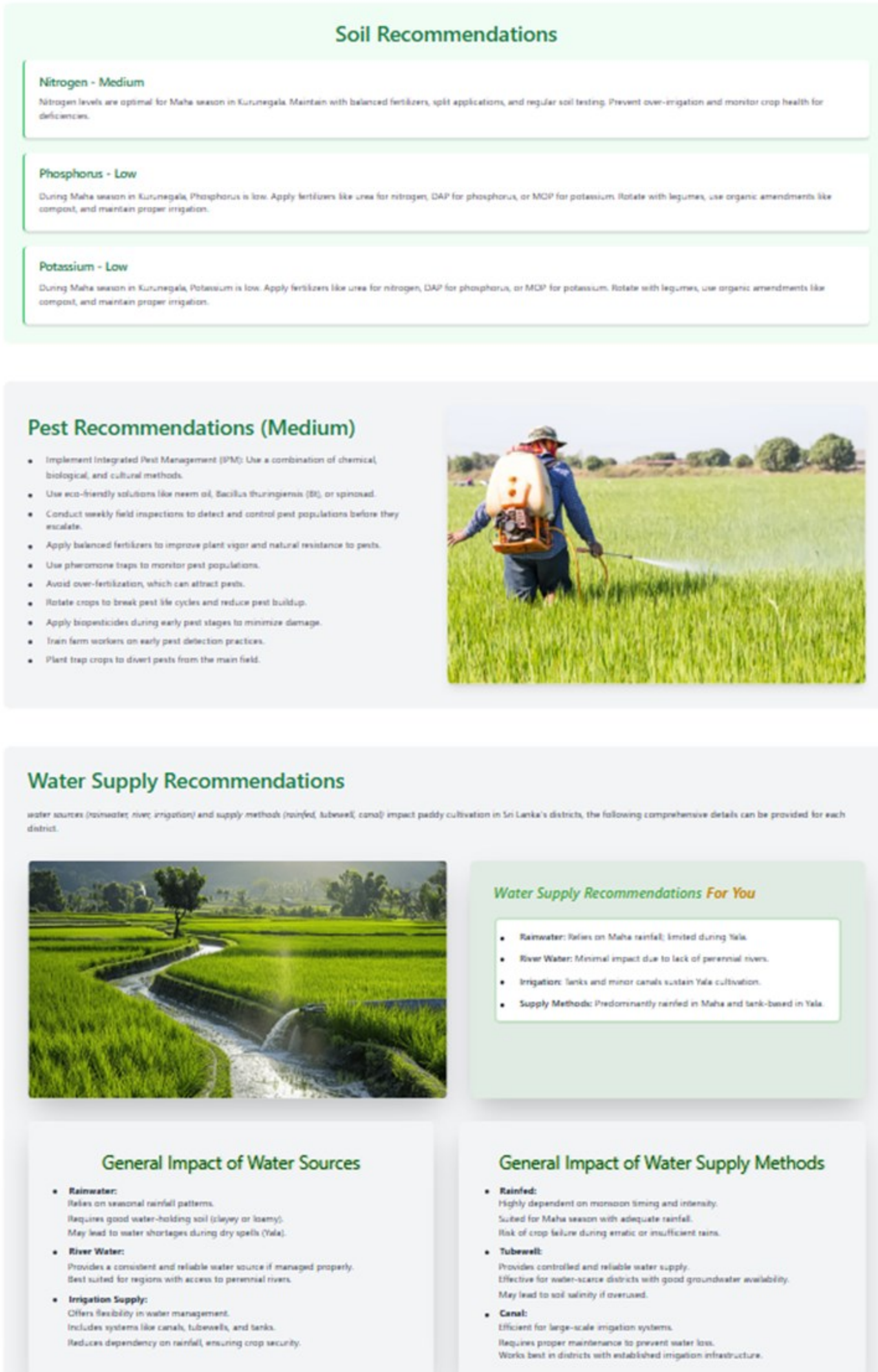

Figure 14: Recommendation Page

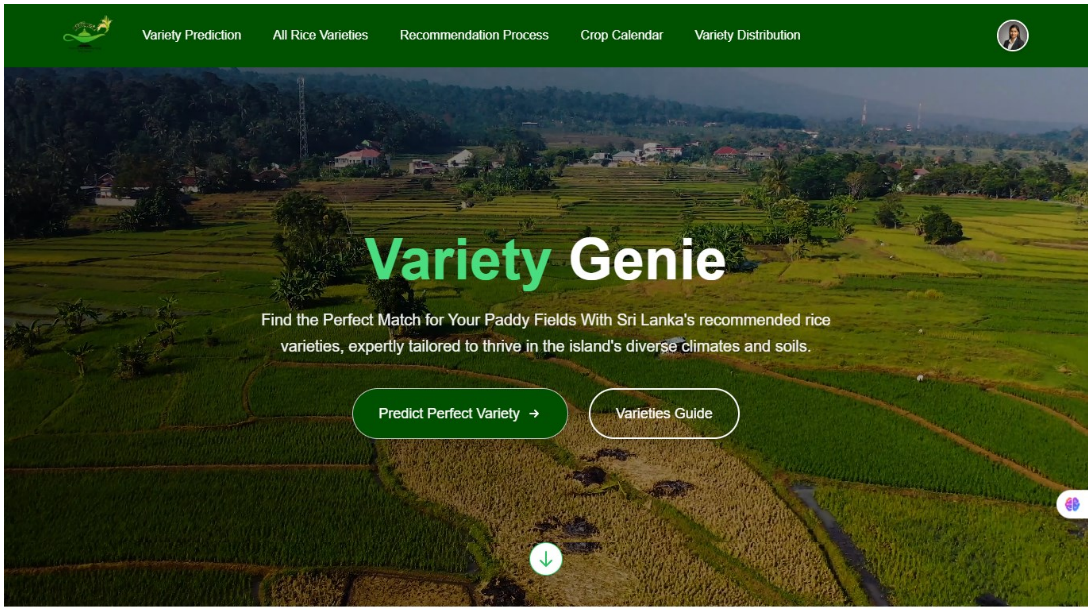

Figure 16: Variety Prediction Home Page

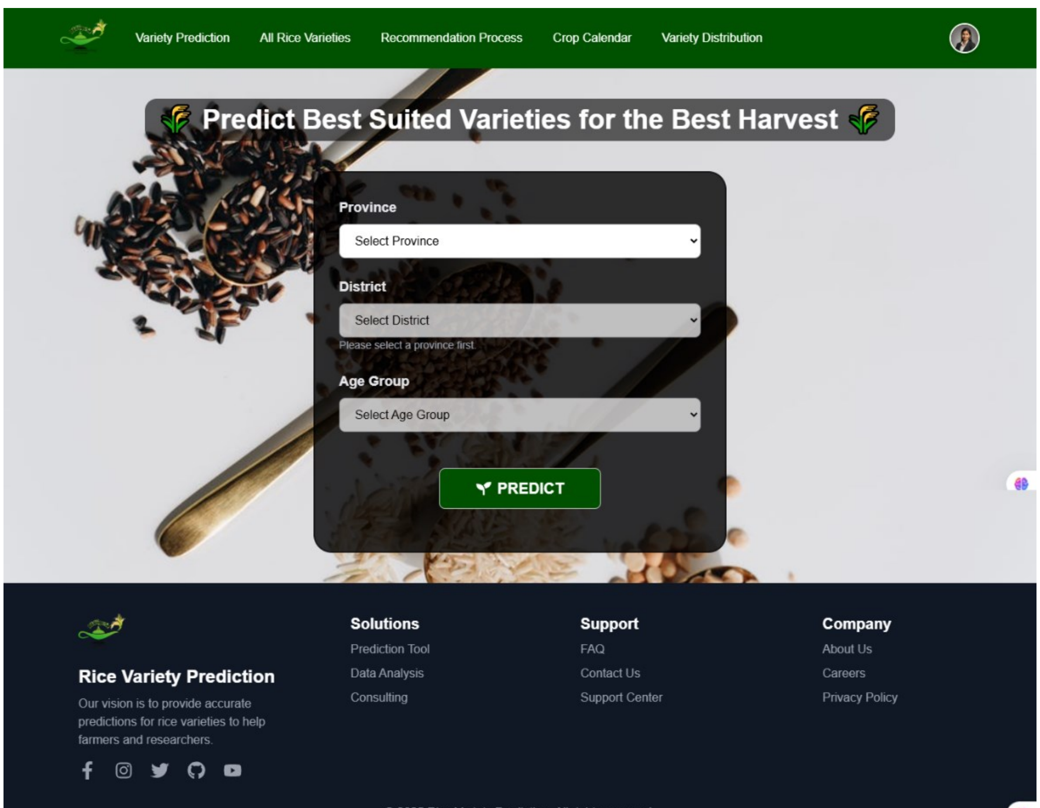

Figure 16: Variety Prediction Input Page

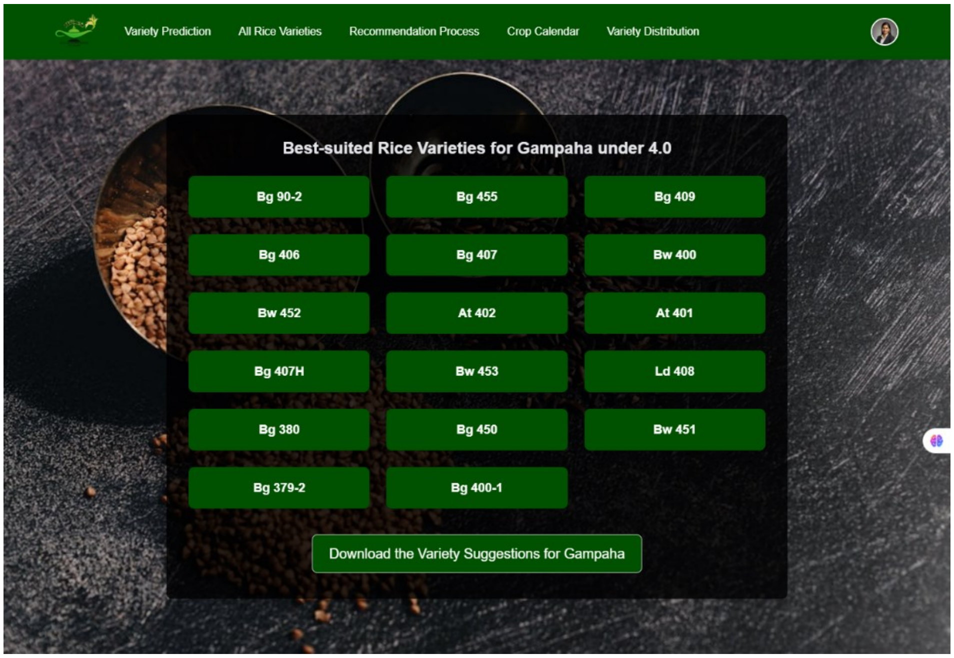

Figure 18: Predicted Variety Page

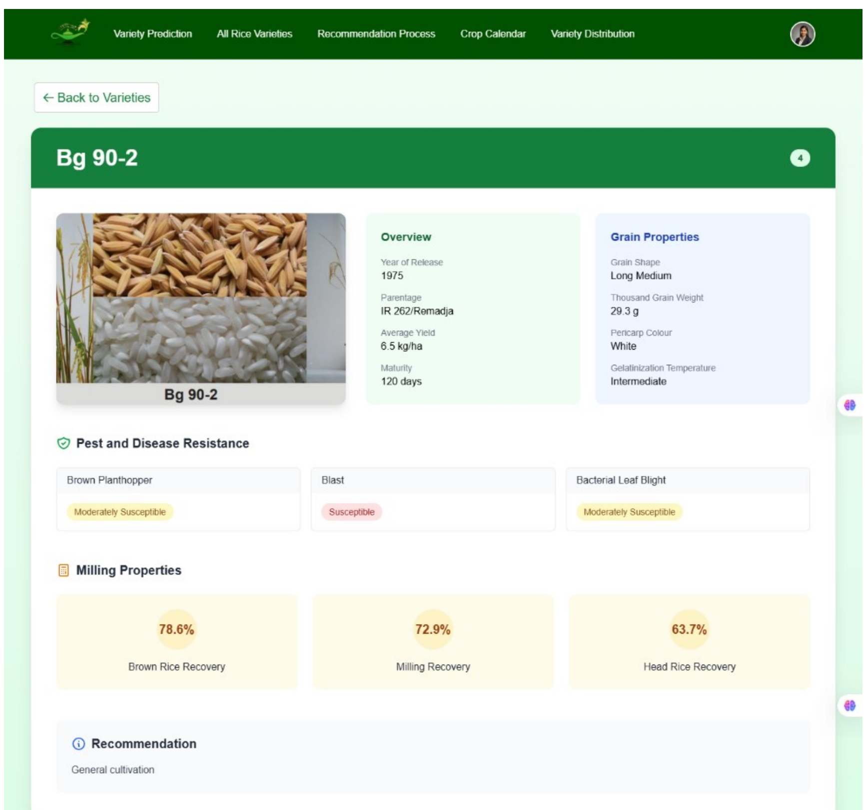

Figure 18:Variety Detail Page

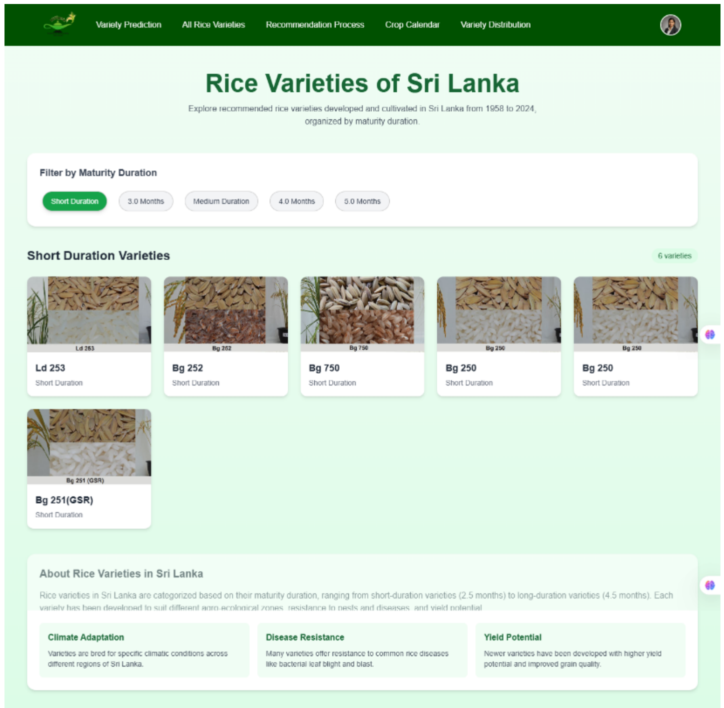

Figure 20: All Variety Detail Page

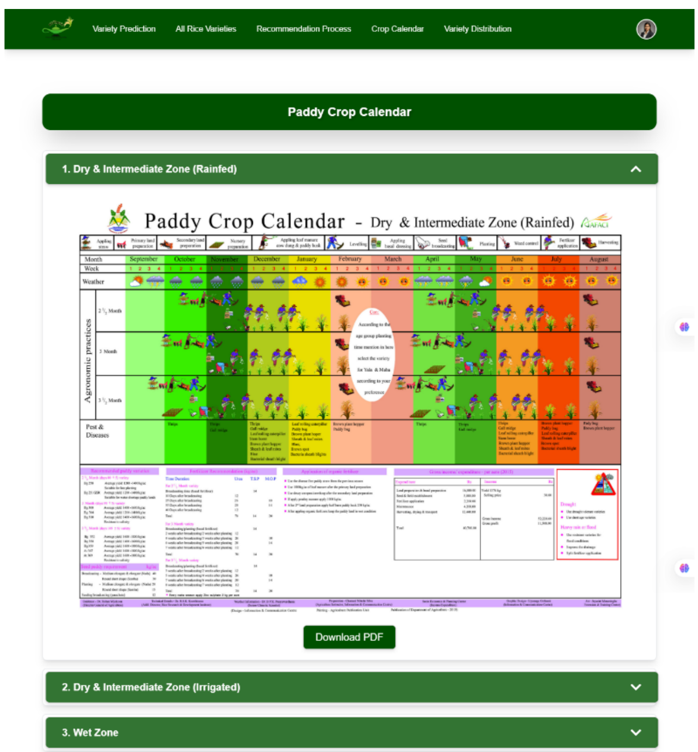

Figure 20: Paddy Crop Calendar Page

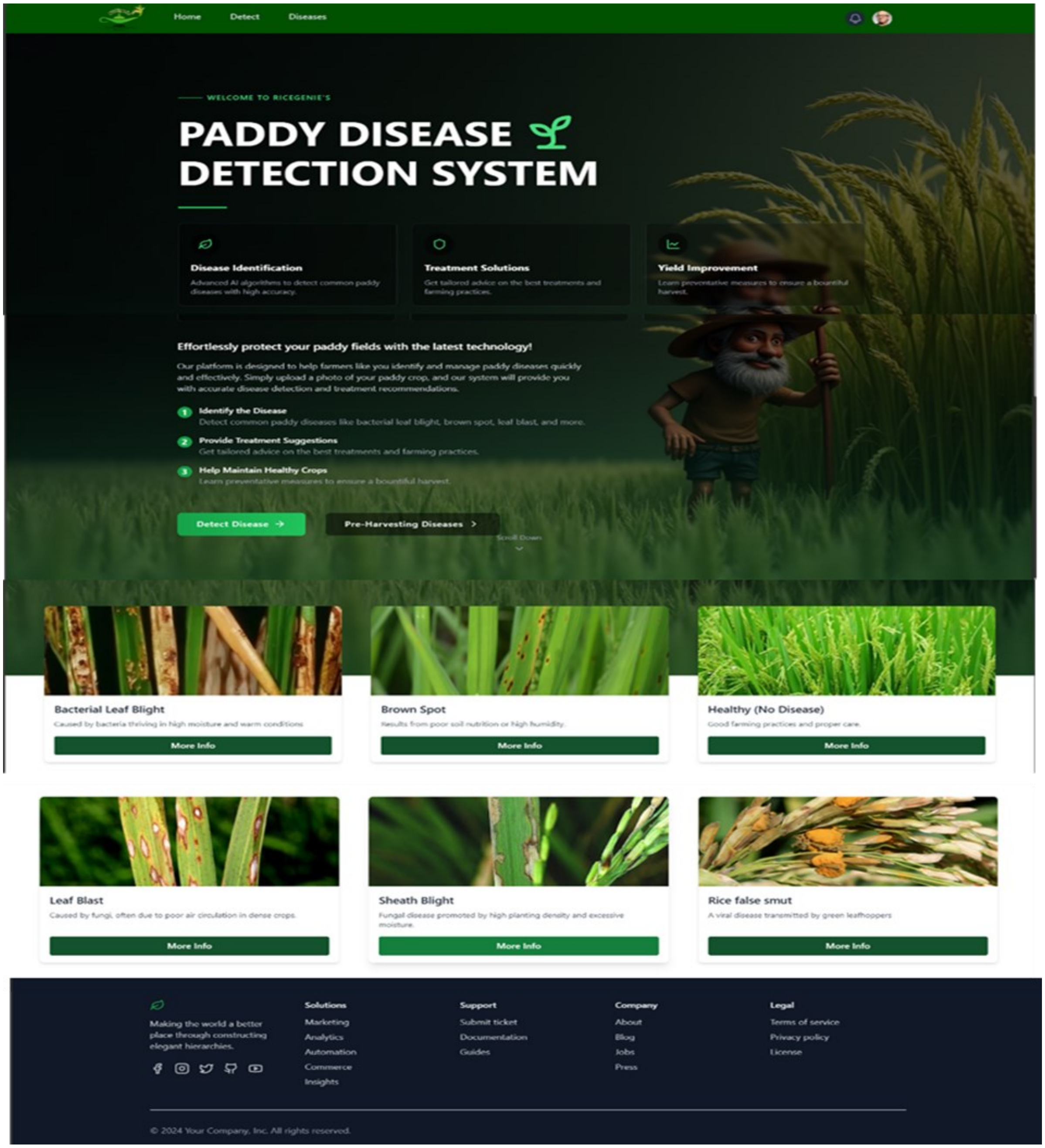

Figure 21: Home page of Paddy diseases Identification

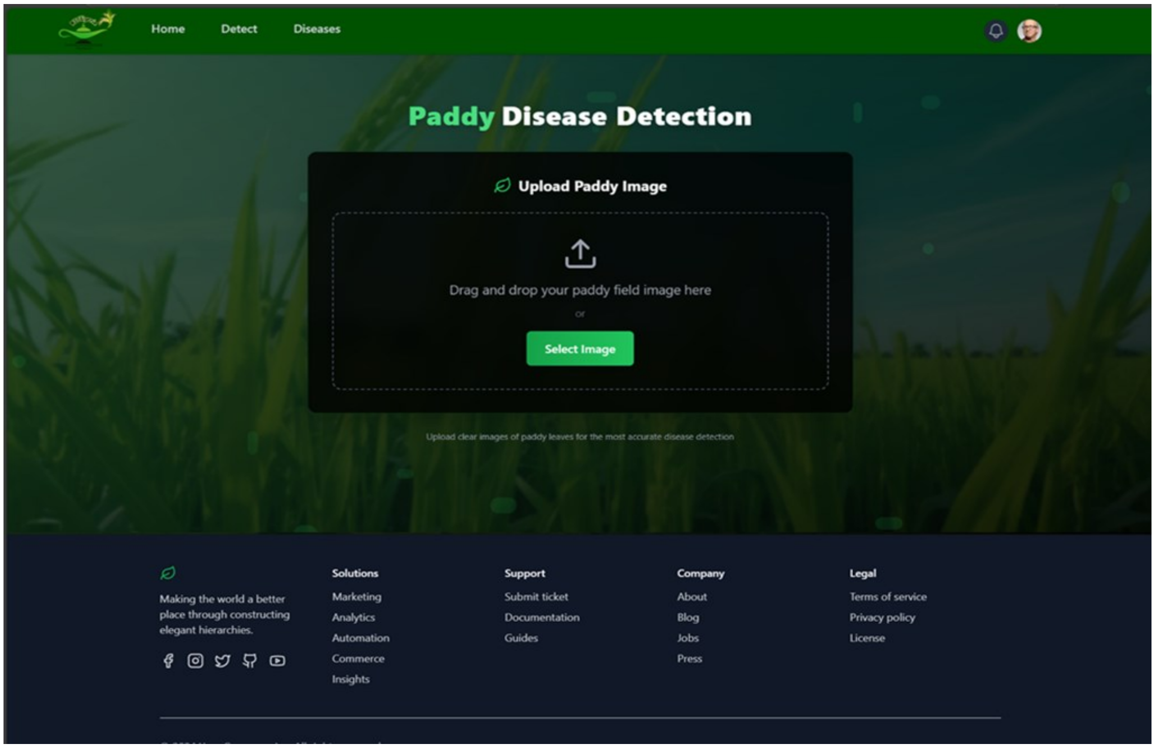

Figure 23: Interface for Upload Image

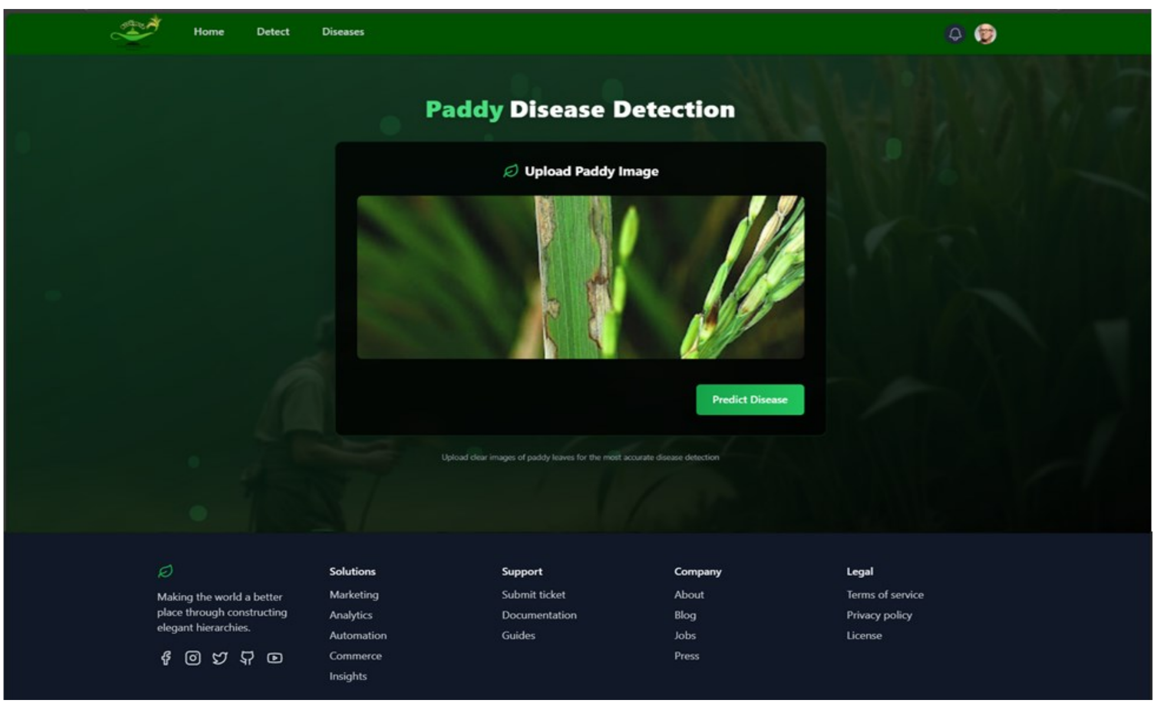

Figure 22: Interface of after uploading image.

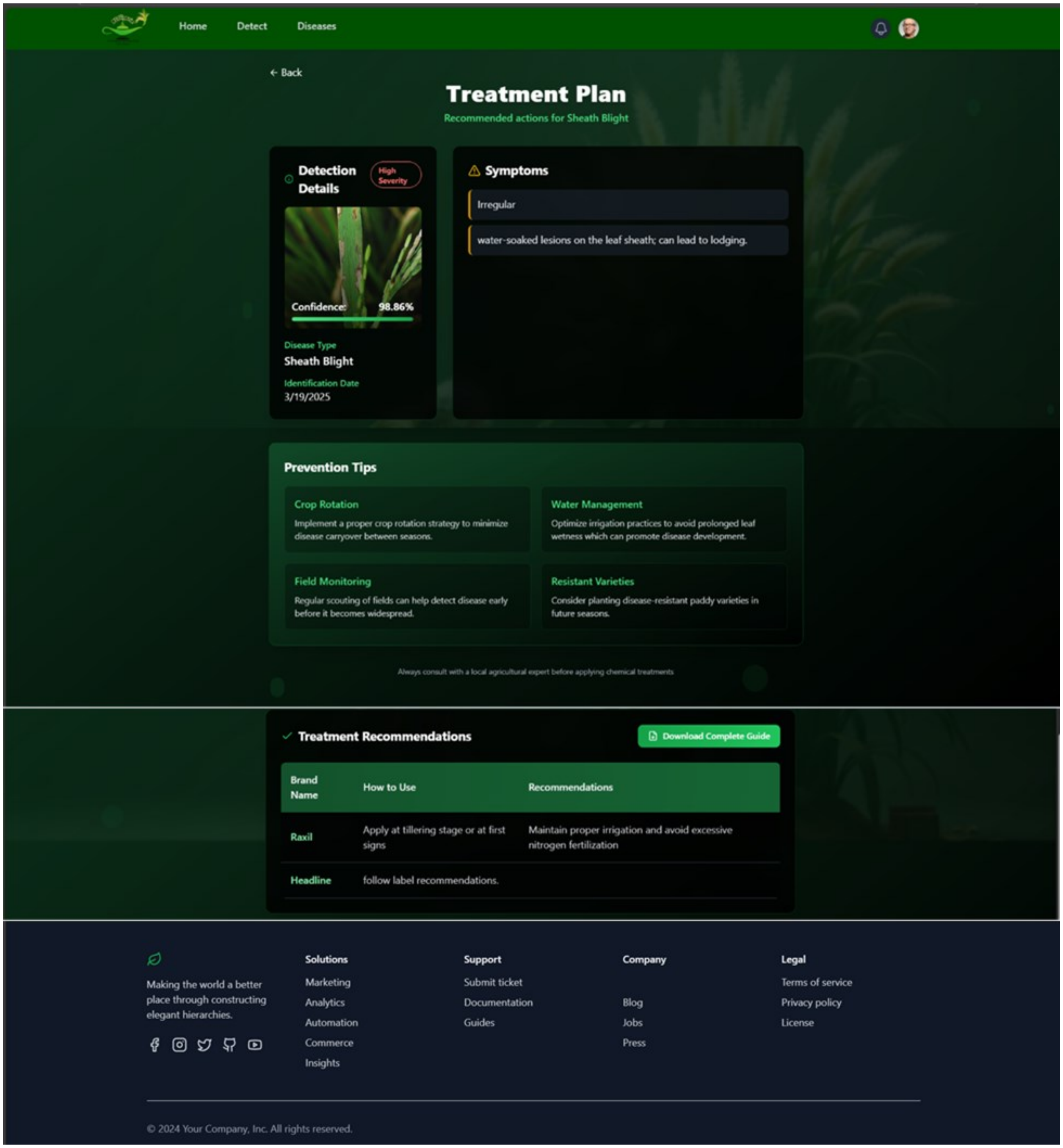

Figure 24: Interface of Treatment recommendation according to the user Inputs and disease Image

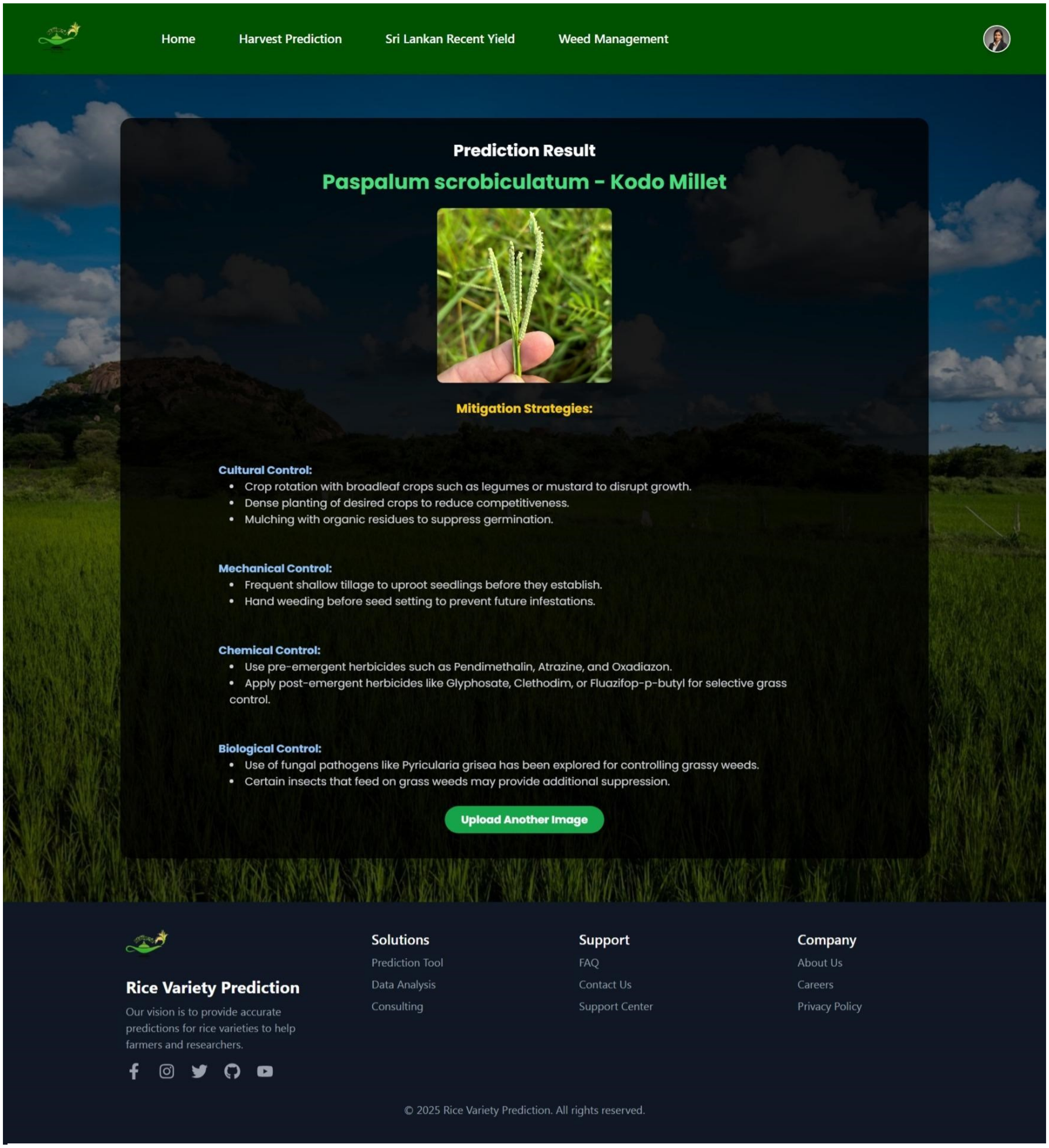

Figure 25: Interface of mitigation method recommendation according to the user Input

REFERENCES

[1] FAO, "Rice Market Monitor," Food and Agriculture Organization of the United Nations, 2018.

[2] S. Savary et al.., "Rice pest management: A global perspective," Global Rice Science Partnership (GRiSP) Annual Report, 2012.

[3] H. Zhang et al., "Development of a real-time and portable device to diagnose rice blast disease based on machine vision and machine learning," 2019.

[4] J. Picon et al., "Deep learning for images of paddy disease detection in paddy fields," 2019.

[5] K. H. C. H. Kumara, R. F. Hafeel, H. M. S. Herath, R. H. U. Gunawardhana, T. H. N. Sudarshana, and N. T. Wijesinghe, "Growth performances and grain yield response of rice varieties (Oryza sativa L.) established at different seed broadcasting rates in irrigated fields of the low country dry zone of Sri Lanka," Tropical Agricultural Research & Extension, vol. 26, no. 1, pp. 1-10, Mar. 2023, doi: 10.4038/tare.v26i1.5615.

[6] R. Rambukwella and E. A. C. Priyankara, Production and Marketing of Traditional Rice Varieties in Selected Districts in Sri Lanka: Present Status and Future Prospects, Hector Kobbekaduwa Agrarian Research and Training Institute, Research Report No. 195, May 2016.

[7] R. Berger, R. Weregoda, and V. Rathnabharathie, "Participatory rice variety selection in Sri Lanka," Practical Action South Asia, Colombo, Sri Lanka, 2007.

[8] V. N. T. Le, G. Truong, and K. Alameh, "Detecting weeds from crops under complex field environments based on faster RCNN," in 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Jan. 2021, pp. 350–355, doi: 10.1109/ICCE48956.2021.9352073

[9] P. Ekanayake et al., "Machine Learning Modelling of the Relationship Between Weather and Paddy Yield in Sri Lanka," 2021

[10] E.J.K.P. Nandani and T.T.S. Vidanapathirana, "Forecasting Paddy Yield in Sri Lanka Using Backpropagation Learning in Artificial Neural Network Model," 2024

[11] P. Ekanayake and W. Wickramasinghe, "Use of Regression Techniques for Rice Yield Estimation in the North-Western Province of Sri Lanka," 2020

[12] A. Amarasinghe et al., "Advancing Food Sustainability: A Case Study on Improving Rice Yield Prediction in Sri Lanka Using Weather-Based, Feature-Engineered Machine Learning Models," SN Applied Sciences, 2024.

[13] Muthu, Pabitha. (2021). Weed Detection Using Image Processing. 10.13140/RG.2.2.15116.64642.

[14] M. H. Asad and A. Bais, "Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network," *Information Processing in Agriculture*, vol. 7, no. 4, pp. 535–545, Dec. 2020, doi: 10.1016/j.inpa.2019.12.002.

[15] B. Liu and R. Bruch, "Weed detection for selective spraying: a review," *Current Robotics Reports*, vol. 1, no. 1, pp. 19–26, Mar. 2020, doi: 10.1007/s43154-020-00001-w.

[16] V. N. T. Le, G. Truong, and K. Alameh, "Detecting weeds from crops under complex field environments based on faster RCNN," in Proc. 2020 IEEE 8th Int. Conf. Commun. Electron. (ICCE), Jan. 2021, pp. 350–355, doi: 10.1109/ICCE48956.2021.9352073.

[17] Department of Census and Statistics Sri Lanka, “Paddy Statistics,” [Online]. Available: http://www.statistics.gov.lk

Accessed: Feb. 2025.

[18] Rice Research and Development Institute (RRDI), Bathalagoda. “Annual Reports and Variety Recommendations,” Department of Agriculture, Sri Lanka, 2021–2024.

[19] Labuduwa Rice Research Station, Galle. “Regional Paddy Breeding and Environmental Suitability Reports,” 2023.

[20] U.A. K. S. Udawela, M.C. Millawithanachchi, and J. Ponnamperuma Arachchi, “Expert consultations on rice varietal recommendation and agro-environmental mapping,” Personal communication, 2023–2024.

[21] Scikit-learn Developers, “Scikit-learn: Machine Learning in Python,” [Online]. Available: https://scikit-learn.org/

[22] H. He and E. A. Garcia, “Learning from Imbalanced Data,” IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263–1284, Sept. 2009. doi: 10.1109/TKDE.2008.239

[23] J. Brownlee, “Logistic Regression for Machine Learning,” Machine Learning Mastery, 2020. [Online]. Available: https://machinelearningmastery.com/logistic-regression-for-machine-learning/

[24] S. Patel, A. Yadav, and N. Sharma, “Rice Crop Recommendation System using Machine Learning Techniques,” International Journal of Computer Applications, vol. 177, no. 26, pp. 1–5, 2020.

[25] Ministry of Agriculture Sri Lanka, “Agro-Ecological Zones and Agricultural Planning Guidelines,” Policy Report, 2021.

[26] mKisan Portal, Government of India. [Online]. Available: https://mkisan.gov.in

[27]Vietnam Rice Knowledge Bank. [Online]. Available: https://www.knowledgebank.irri.org/vietnam/

[28] M. R. Islam, “Application of ICT in Agricultural Technology Dissemination: A Study on e-Krishok Service in Bangladesh,” Asian Journal of Agricultural Extension, vol. 18, no. 2, pp. 100–110, 2020.

GLOSSARY

Term	Definition
Paddy	A type of wet rice cultivation commonly grown in flooded fields.
Disease Detection	The process of identifying symptoms of plant diseases using image processing and machine learning.
Pre-Harvest Stage	The period before the rice crop is harvested is critical for managing diseases and improving yield.
Bacterial Leaf Blight (BLB)	A bacterial disease that causes wilting and yellowing of leaves, often spreading through wind and rain.
Brown Spot	A fungal disease that appears as brown lesions on leaves, reducing plant vigor and grain quality.
Leaf Blast	A highly destructive fungal disease affecting both young and mature rice plants.
Sheath Blight	A fungal infection causes the drying of leaf sheaths and reduces photosynthetic activity.

Deep Learning (DL)	A subset of machine learning using neural networks to model complex patterns in data.
Convolutional Neural Network (CNN)	A type of deep learning architecture effective in image classification tasks.
Data Augmentation	Techniques such as flipping, rotating, and scaling are used to artificially increase dataset diversity.
Grad-CAM	A visualization tool was used to interpret which parts of an image influenced the CNN's decision.
TensorFlow	An open-source deep learning framework developed by Google.
Keras	A high-level API for building and training deep learning models, often used with TensorFlow.
FastAPI	A modern web framework for building high-performance APIs using Python.
ReactJS	A JavaScript library used for building interactive and user-friendly web interfaces.
MySQL	A relational database management system is used for storing structured data.

Precision	A metric indicating the proportion of true positive predictions among all positive predictions.
Recall	A metric indicating the proportion of actual positives correctly identified by the model.
F1-Score	The harmonic mean of precision and recall, used to measure overall classification performance.
Model Deployment	The process of integrating a trained machine learning model into a production environment.
Treatment Recommendation	A system feature that provides specific suggestions (e.g., chemical brands, dosage) based on disease class.
Image Preprocessing	The process of preparing raw images (e.g., resizing, normalization) before feeding them into a model.