INFORMATICS INSTITUTE OF TECHNOLOGY
In Collaboration with
ROBERT GORDON UNIVERSITY ABERDEEN

Disease And Weed Detection of Paddy Plants and Remedy Recommendation System Using Image Processing Techniques

Group 7 Thesis by

Kavindya Koralegei - 20210575
Ronath Jayasuriya - 20210167
Dineth Hasaranga - 20210537
Mohomad Fazlan – 20210137

Supervised by

Ms. Sulochana Rupasinghe

Submitted in partial fulfilment of the requirements for the BSc (Hons)in Artificial Intelligence and Data Science degree at the Robert Gordon University.

February 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Chapter Overview

The proposed, Govimithuru app, is a tool to support paddy cultivators to detect the disease named Brown Spot in paddy plants and a weed named Wrinkle duck beak using image processing techniques and recommend remedies to them. A friendly chatbot and disease severity check system is included in the tool. The project's mission is to create a platform that can help farmers detect the weed, Wrinkle duck beak, and the disease, Brown Spot, and receive remedies for their problems in real-time in absence of a human expert. Cutting-edge AI techniques will be integrated to identify the disease and the weed and recommend remedies.

1.2 Problem domain

1.2.1 Background

Sri Lanka has been an agriculture-based country since time immemorial. The main occupation was agriculture and land usage was mainly agricultural. The agricultural industry in Sri Lanka is classified as Plantations, fisheries, livestock, and forestry currently. The agriculture sector has a contribution of 7.4% to the national GDP (International Trade Administration, 2021). The top priority of the current regime is to support productivity in agriculture. Rice is considered the primary food crop in the country. Tea cultivation is done in the central highlands of the country which is the main source of foreign exchange. In Sri Lanka, 30% of the population engages in the agriculture sector (International Trade Administration, 2021). According to the World Bank collection of development indicators agricultural land area in Sri Lanka is recorded as 45.5% (The World Bank, n.d.). According to the agriculture department, 512,00 hectares of paddy have been cultivated during Yala season of the year 2022 and the number of farmers engaged was 490,515 (Ministry of Agriculture, 2022).

1.2.1.1 Common Issues of Paddy Agriculture

The self-sufficiency of the country's paddy cultivation has recently faced difficulties, which have brought to light the necessity for creative solutions and collaborations. By addressing critical problems including climate change-related disasters, poor infrastructure, crop diseases & pest attacks and lack of mechanization, the nation would be able to increase crop production and eventually become self-sufficient in rice. According to the research carried down by the agriculture department crop diseases, weeds and pest attacks have been the major reason for the downfall of production.

There are numerous rice diseases that are brought on by nematodes, bacteria, viruses, and fungi, according to published research. While the effects of certain illnesses are minimal, others can cause major crop losses and reach epidemic levels. Rice blast, Sheath blight, Brown spot, False smut, Leaf scald, Sheath rot, Narrow brown leaf spot, and Bacterial leaf blight are some of the most common paddy diseases found in Sri Lanka. E- Common barnyard grass, E – Torpedo grass, Bata dalla, and forked fimbry are commonly found weeds in paddy agriculture.

1.2.1.2 Main diseases and weeds of paddy agriculture

Disease Name	Scientific Name	Distribution	Symptoms/Characteristics	Impact
Rice Blast	*Magnaporthe grisea*	Global	Spindle shape leaf spots, ashy centers, pointed ends, 1.0-1.5cm long and 0.3-0.5cm broad.	Nodes become black and rot. Rotten neck of panicle.
Sheath Blight	*Thanatephorus cucumeris*	Global	1-3cm long, Oval or ellipsoidal. Lesions initially appear near water and soil and then spread to leaf sheath.	Asexual overwintering structures develop on the surface of the leaf sheath.

Brown Spot	*Cochliobolus miyabeanu*	Global	Circular spots on coleoptiles, Colour vary from dark brown to reddish brown.	Attacking the young florets could impede grain development. Affected areas are coleoptiles, leaves, leaf sheaths, panicle branches, glumes, and grains.
Bacterial Leaf Blight	*Xanthomonas oryzae pv. oryzae*	Global	Lesions typically appear as orange stripes on leaf blades or leaf tips on mature plants.	Seedling stage and mature plant leaves are affected.
E- Common barnyardgrass	*Echinochloa crus-galli (L.) P. Beauv*	Global	Erect, stem- up to 200 cm tall, leaf- 10–40 cm long, green to purplish in color	Reduce the growth of paddy plants.
Wrinkle duck beak	*Ischaemum rugosum Salisb.*	Global	Erect, Stem- Up to 100 cm tall, Leaf- blades 10–30 cm long, **Inflorescence-** paired terminal spikes.	Hinder the growth of paddy.

Table 1: Selected paddy diseases & weeds

1.3 Problem Definition

Multiple plant diseases can be detected in a variety of methods, but plant pathology detection systems are less prominent in Sri Lanka. The diagnosis of plant diseases also makes use of a variety of laboratory-based methods. Polymerase Chain Reaction, Fluorescence *in-situ* Hybridization, Enzyme-Linked Immunosorbent Assay, Immunofluorescence, Thermography, Fluorescence Imaging and Gas Chromatography are some lab-based techniques (Fang and Ramasamy, 2015).

G. Anthonys and N. Wickramarachchi (2009) implemented a Paddy disease detection system using Image processing techniques. Researchers have used a mathematical morphology to segment the images and a classification method of membership function was used to differentiate the diseases. The system has an overall accuracy of 70%. Less no. of images used for the training process and the common disadvantage to the noise of segmentation algorithms can be cited as limitations of the system.

Edirisinghe et al.,(2022) implemented a system to identify weeds in paddy fields using Multispectral images obtained from an Agriculture drone. The system spots the area with the weeds but doesn't suggest a solution. However, the system couldn't configure grasses and sledges properly but was able to configure broad-leaf weeds.

According to the researchers, the FiX4D software's vegetative index-based image processing is insufficient for creating accurate weed maps. The researchers have mentioned the importance of implementing a more accurate image processing system using deep learning techniques to identify weeds.

There are many systems to detect paddy leaf diseases, but most systems are based out of Sri Lanka. Systems built to detect weeds in paddy fields are very rare in Sri Lanka and the existing systems are not able to achieve the success rate expected. Also, they do not provide the necessary solutions. The existing systems don't facilitate the option of solving other general problems in paddy agriculture.

1.3.1 Problem Statement

The current systems do not have the capability to identify weeds like wrinkled duck beak and paddy leaf diseases like Brown spot disease and recommend solutions within the same system.

1.4 Research Motivation

The main ambition of this research project is to develop an accurate, reliable Brown spot *(Cochliobolus miyabeanus)* paddy plant disease and Wrinkle duck beak *(Ischaemum rugosum Salisb)* weed detection in one mobile application while providing the severity of the disease and remedy recommendations to aid farmers in Sri Lanka.

Even though several technological advances have been involved in paddy agriculture in recent years in the world, new cultivators and farmers in Sri Lanka who have been in the agriculture field for many years are still relying on traditional mechanisms and experience along with receiving guidance from the regional paddy agricultural officials to detect paddy diseases and weeds. So, we as researchers were motivated to aid the farmers in developing such a system that would help discover two of the most troublesome paddy disease and weeds, giving the best solutions to recover from them and to farmers' queries regarding paddy agriculture.

1.5 Existing Works

In Sri Lanka, there are currently no technologies or applications in use that can accurately identify rice weeds and rice diseases, in particular, and offer more suitable remedies. However, there are also earlier techniques that have been put in place to detect weeds and diagnose rice diseases in Sri Lanka and other Asian countries with less precision.

Citation	Limitation	Technology/ Algorithm	Advantages
(Bandara and Mayurathan, 2021)	• Low accuracy due to lower images used in testing and training	SVM, k-NN	Color thresholding and different image preprocessing and classification techniques are compared to find the best-fitting method.
(Anthonys and Wickramarachchi, 2009)	• Wrong results due to poor image quality • Low accuracy due to lower images being used	Nearest neighbor classification	Machine learning techniques are employed. Can gain a general sense of whether machine learning or deep learning method produces the greatest outcomes.
(S.S.B.P.S. and V.G.T.N., 2022)	• Doesn't have the optimum accuracy due to the small dataset for testing • Diseases are not specifically recognized, only if it's diseased or not.	Decision tree, logistic regression, CNN	Use of Decision tree, logistic regression, and CNN models to compare and find the best-fitted method.

(Andrianto et al., 2020)	• Low test accuracy due to poor quality of images and noise disturbance	VGG165, CNN	Usage of VGG165 and machine learning techniques to get the optimum level of accuracy for the training dataset.
(Charitha et al., 2021)	• The weed is not recognized specifically, only whether a weed is in the crop field or not	CNN, YOLOV3, Image preparing	The usage of YOLO and image processing helped the images to have a corrected contrast and density
(Hu et al., 2021)	• Difficulty in identifying weeds from crops due to crop similarity and complex background • Lack of image pre-processing leading to low weed detection accuracy	CNN based on YOLOV4	Compared with YOLOV3, YOLOV4 has the advantage of accuracy improvement in network recognition and target detection
(Pabitha Muthu, Hemalatha and Atshaya G., 2021)	• Not specifically detecting the exact weed in the field	Image processing using MATLAB, K-means clustering	Edge detection and color segmentation are being used to compare and contrast several features in the images taken

Table 2: Existing Work

1.6 Research Gap

Paddy diseases and weeds can be regarded as a serious challenge in the agricultural field in the Sri Lankan context where employment of new technologies in the agricultural business is not at a satisfactory level. Developing a tool that can detect paddy illnesses and weeds in a one single mobile application is the research gap that will be addressed.

According to the above existing work, even though there have been several systems to detect paddy diseases in the past, currently no existing, active, and accurate application is being used in Sri Lanka to detect both paddy diseases and weeds and to give suitable remedy recommendations for those causes.

Taking rice weeds into consideration, there has not been any system or application in the Sri Lankan agriculture field that has made an attempt on identifying weed varieties in paddy fields. Manual weed identification and removal techniques are still employed for weed kinds, and they require a lot of human labor and time. Also, most farmers have little understanding of a specific disease and weeds and how to prevent it when a paddy plant becomes infected.

They are aware that some of them use conventional methods and others consult with field officers to get information and recommendations.

In this proposed system, Brown spot as the rice disease, and Winkle duck beak as the weed variety is being identified in the same application as these two have become the notable troublesome disease and weed found in Sri Lankan paddy fields. This system will also provide applicable expert remedy recommendations to improve the paddy and help them to have a real-time analysis by allowing a severity check on the diseased paddy.

1.7 Contribution to the Body of Knowledge

1.7.1 Technological Contribution

Machine learning algorithm and image processing technology will be combined in this paddy plant disease and weed detection tool. This implies that image pre-processing and image classification will be done using TensorFlow and Keras. Postman and React js are being used for web development and application testing and Jupyter will be used for model creation and representation.

When training the datasets and the system, to get a better accuracy percentage than other similar systems we can test using trial and error and find the best possible system layer structure.

1.7.2 Domain Contribution

Due to farmers In Sri Lanka being somewhat technologically illiterate, creating a user-friendly and easily accessible, and updatable mobile app, which could identify the disease or weed both together and provide remedies for those identified diseases and weeds would be of great convenience to the users. As this is available for the users as an affordable mobile application, any person who has access to a mobile phone is able to take advantage of the proposed application.

1.8 Research challenges

- Due to differences in the weed in Sri Lankan agricultural context and not being able to find an appropriate large data set for Wrinkle duck beak, we would have to make our own data set with enough images, such as around 400 images to train the system.

- Since all the project members aren't experts in handling data and have not participated in complex research projects before, the uncertainty in handling and cleaning datasets and working and training them has become a challenge in this project.

- Not having proper knowledge and understanding about the agricultural domain and especially paddy research domain and getting to know about different types of rice diseases and weeds has been another challenge we had to face.

- Choosing the best model or algorithm that's appropriate for our project is another challenge we faced as there are several machine learning technologies that we can use. So by evaluating and validating other research works we have to find the most suitable model for this project.

1.9 Research Questions

RQ1	What is the most suitable algorithm for differentiating diseased parts of a plant leaf or from paddy leaves to weeds?
RQ2	How would the tool differentiate brown spot disease from a disease that would show similar signs to the disease?
RQ3	How was the dataset consisting images of brown spot diseased paddy leaves and wrinkle duck beak weed collected?

Table 3: Research Questions

1.10 Research Aim

The aim of this research is to give farmers a tool to identify and provide recommendations of remedies to mitigate the damage caused by brown spot disease and wrinkle duck beak weed. After that identification, a severity check is done to check if the threat of the disease is high or low. If the farmer has questions regarding the paddy field an automated bot is given to answer those questions.

1.11 Research Objectives

Research Objectives	**Explanation**	**Learning Outcome**
Problem Identification	To detect brown spot disease and weeds accurately using image processing with the help of images. • Reading other research papers regarding the topic and identifying the main points and improving on that idea.	LO3
Literature Review	To identify the already existing paddy disease detection and weed detection in a paddy field using image processing techniques. • Reading research papers regarding the topics. After that the problems those researchers addressed, the solution used, and the limits of those solutions are identified.	LO1, LO3

	• Identifying an image processing algorithm.	
Data Gathering and Analysis	To gather datasets with high and mid-quality snapshots of brown spot disease and weeds in paddy fields. • Due to a plethora of peer-reviewed datasets for brown spot disease, there would be no issue in collecting those datasets. • There is a lack of datasets regarding the wrinkled duck beak weed, therefore the research group has contacted the Department of agriculture in creating and verifying those datasets.	LO1 , LO3
Research Design	Model development using gathered data resources and according to specified requirements. • Finding a way to integrate the different components of the research in a logical and coherent and logical way.	LO2 , LO3
Implementation	Implementing the developed system for disease and weed detection using image processing and integrate it into a physical system. • After the research design the process of putting the designed idea into effect.	LO2 , LO3 , LO4
Testing and Evaluation	Testing the system built with a range of unique and real-time inputs. • Testing the implemented system and checking and debugging errors.	LO2 , LO3 , LO4

Table 4: Research Objectives

1.12 Project Scope

1.12.1 In Scope

No	Description
1	Identifying the diseases as brown spot in paddy leaf using the uploaded image
2	Identifying the weed as wrinkle duck beak in a paddy field using the uploaded image
3	Recommending the remedies according to the weed or disease

4	If the farmers have an issue regarding the paddy field an automated bot was created to be there to answer those questions
5	According to the severity of the disease, classification would be given

Table 5: Project In-scope

1.12.2 Out Scope

No	**Description**
1	Ability to identify all the paddy diseases
2	Ability to identify all weeds in the paddy field
3	The user only will get a mobile app not a web application
4	Limit is for 1 disease and 1 weed
4	User–friendly language

Table 6: Project Out-scope

1.12.3 Prototype Diagram

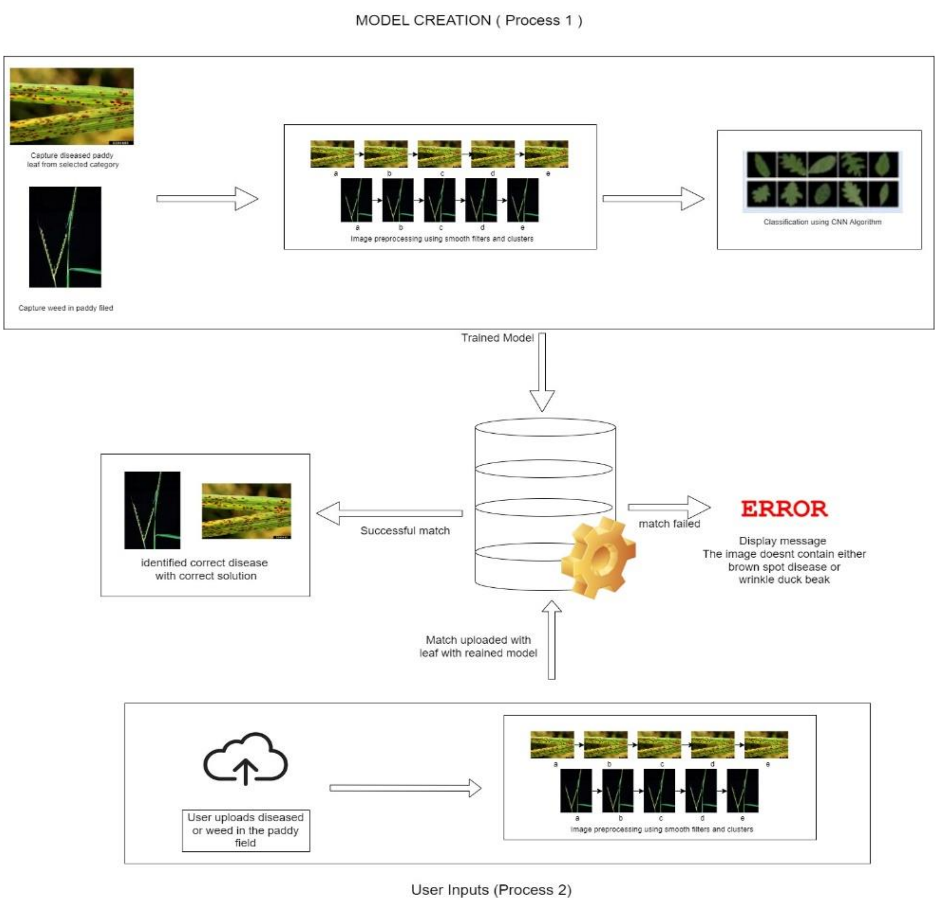


Figure 1: Prototype Diagram

1.13 Resource Requirements

1.13.1 Hardware Requirements

- **Laptop**
  - o Intel® i5 7th Gen
  - o RAM 16 Gb
  - o CPU 64-bit operating system

1.13.2 Software Requirements

- **MATLAB –** The program used to obtain the results, is utilized to identify diseased spots in paddy leaves as well as weeds by converting color to hue.
- **Python –** The primary language used to create the planned system
- **Tensor Flow –** Image processing and model training
- **Keras –** Image processing and model training
- **RASA –** For chatbot creation
- **Android Studio –** For app development and testing

1.13.3 Data Requirements

- Most of the data sets available are internationally facing paddy weeds and since this project provides a solution to Sri Lankan agriculture there are no suitable datasets, therefore, the dataset for Wrinkle duck beak was created by the team.
- For the paddy disease, we were able to find an appropriate brown spot disease dataset through "Kaggle"

1.13.4 Skill Requirements

- Research, Data collection and information skills
- Programming skill
- Data Cleaning
- Feature Engineering
- Model Prototyping

1.14 Chapter Summary

A disease and weed detection system for paddy plants is crucial for efficient and effective crop management. Such a system can quickly identify and diagnose problems with the plants, allowing farmers to necessary actions. This can include providing knowledge to apply the appropriate pesticides and fertilizers, Additionally, a disease and weed detection system can help to reduce the use of unnecessary chemicals, saving farmers money and minimizing the environmental impact of their operations. Overall, a disease and weed detection system for paddy plants is an important tool for ensuring optimal crops and protecting the health of the plants.

CHAPTER 2: LITERATURE REVIEW

2.1 Chapter Overview

Paddy cultivation in Sri Lanka is given utmost significance in the agriculture industry since rice is considered as the staple food of Sri Lanka and play's major role in the country's economy. Rice is recognized as the low cost and effective nutrient rich food available in the Asian region. In accordance to World Bank reports, consumption of rice will increase by 52% by 2025. The demand for rice has grown faster than the production in the majority of rice-producing countries. The main reason is the damage to crops due to pests' attacks, diseases, and the domination of weeds. Detection of disease and weeds has always been a challenging task. Currently, farmers use traditional procedures and seek the guidance of regional agriculture officers to detect the disease. Thus, this study attempts to identify paddy diseases and weeds and provide solutions using machine learning techniques with the support of Image processing. Also, we focus on introducing a Chatbot to help farmers to resolve their problems and a method to check the progress of the field using image processing techniques.

2.2 Concept Map

The project's information is systematically gathered using the concept graph.

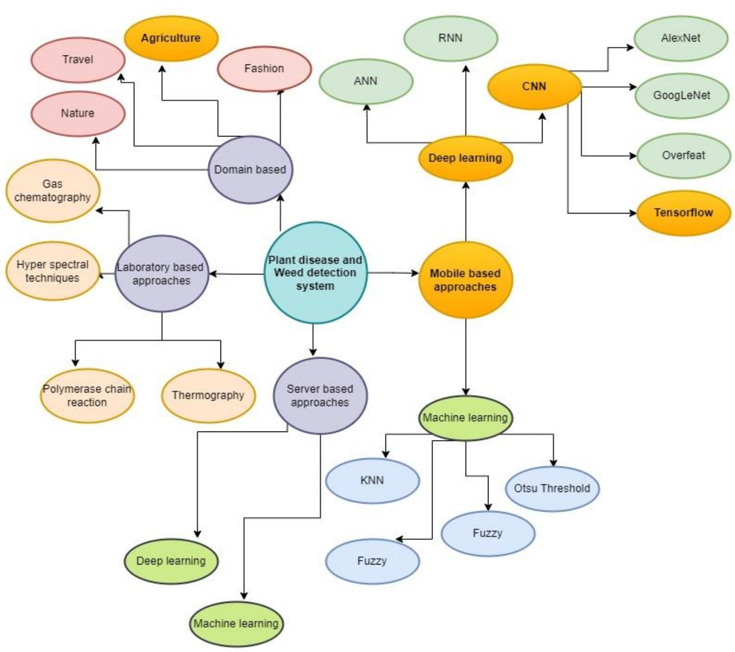


Figure 2: Concept Map

2.3 Relevant work

2.3.1 Disease and weed Identification using Image processing

2.3.1.1 Disease identification using image processing

Diseases in the paddy has serious negative effects on crop yield, therefore, to correctly identify the diseases is the key to avoid these effects. However, the existing disease diagnosis methods for paddy are neither accurate nor efficient, and special equipment is often required. With the use of deep learning based on a large dataset, we can identify those diseases without the before mentioned specialized equipment. There have been several research done to identify diseases using image processing. Convolutional neural network (CNN), Artificial neural network (ANN), and Support vector machines (SVM) are mostly used in the research thesis.

In 2017 theirs been research on the Recognition of diseases in paddy leaves using knn classifier by researchers name of M Suresha,K N Shreekanth, and B V Thirumalesh. where they identified Blast and Brown Spot diseases using image processing and a pattern recognition approach with accuracy of 76.59% where only fungal diseases have been considered (Suresha, Shreekanth and Thirumalesh, 2017).

Deep Learning-Based Classification for Paddy Pests &Diseases Recognition (Indonesia) by researches Ahmad Arib Alfarisy, Quan Chen and Minyi Guo where they detect both paddy pests and diseases using deep learning trained and fine-tuned to fit accurately on the dataset and creating paddy in 2018. This had an accuracy of 87% (Alfarisy, Chen and Guo, 2018).

Disease Identification in paddy leaves using CNN based Deep Learning by R Swathika, S Srinidhi,N Radha and K Sowmya where they did quick classification of paddy into diseased or healthy plants. If the plant is diseased, the area affected is identified using the image classification using CNN, ANN and SVM in 2021 and his program had 70% accuracy. The limit of this research is disease name is not recognized. Only whether it is healthy or not is recognized (Swathika et al., 2021).

Paddy Plant Disease Classification and Prediction Using Convolutional Neural Network by the researchers G K Sagarika,SJ Krishna Prasad, S Mohana Kumar, uses a system to detect the paddy leaf diseases and to classify them into 5 different classes of paddy diseases with minimal preprocessing steps with the help of MATLAB, CNN, and Hardware with an accuracy of 94.12%. The research thesis has not recommended suitable pesticides and fertilizer bontheir own based on the results found (Sagarika, Krishna Prasad and Mohana Kumar, 2020).

The researchers Rounak Talreja,Varsha Jawrani, Bhajan Watwani, Sharmila Sengupta,Prium Rohera and K.S.Raghuwanshi came up with an app called AgriCare for Detection of Paddy Diseases. This would allow limited-resource farmers to detect plant diseases at an early stage and eliminate incorrect fertilizer use which can harm the soil and plants. This methodology was created in 2022 with an accuracy of 99.48% where only 4 diseases can be recognized using this application (Talreja et al., 2022).

2.3.1.2 Weed identification using image processing

Weeds are unwanted plants that grow among crops. These weeds can significantly reduce the yield and quality of the farm output. Therefore, to increase to yield using the previously mentioned methodologies in the paddy disease identification we can identify those weeds and give them proper weedicides. There also have been several research done to identify weeds using image processing.

In 2020, the researchers Xue Yan,Xiangwu Deng and Jing Jin used SVM,DCNN and knn models to Classify weed species in the paddy field. The solution the thesis proposed is to the classify accuracy of weeds with the complex background and variable illumination. Due the to distinguishing ability of a single traditional handcrafted (HC) feature being limited, all shape, texture, and color features have been combined for weed detection to improve the recognition accuracy of weeds. This method had a 94.5% accuracy when testing (Yan, Deng and Jin, 2020).

Deep Learning Based Overcomplete Representations for Paddy Rice Crop and Weed Segmentation. This thesis uses CNN model to differentiate weed plants. UAV and ground-based sensing along with deep learning, especially CNN models are used to automate crop management. The authors of the thesis, G Ujwal Sai,N Tejasri,Ajay Kumar and P Rajalakshmi have stated to have 72.5% of accuracy. The traditional encoder-decoder models struggle in predicting segmented masks with noisy boundaries and overlapping crop-weed environments. (Sai et al., 2022).

In 2022 weed detection in paddy field using an improved RetinaNet network by the researchers Hongxing Peng, ZiheLi,ZhiyanZhou and Yuanyuan Shao used a drone which would fly around the field and detect those weeds in Realtime. This methodology has an accuracy of 94.1% (Peng et al., 2022).

In 2019 researchers S Umamaheswari, R Arjun and D Meganathan did a thesis called Weed Detection in Farm Crops using Parallel Image Processing where they used a Convolutional Neural Network, to get an accuracy of 91.1%. The target of this work is to use deep learning model to perform weed detection in rice crop images and to achieve accurate real-time detection and low machine cost resulting in widely used in practice. The limitation of these technique are not targeting other types of weeds like broad-leaf weed and sedges (Umamaheswari, Arjun and Meganathan, 2018).

2.3.2 Solutions – To diseases and weeds

Farmers in Sri Lanka mainly rely on the information given by the Agricultural instructors to decide the relevant fertilizers, weedicides and pesticides for their fields. Due to insufficient number of instructors, currently they face a huge issue to get timely instructions. Therefore, farmers are tempted to use their past knowledge in using fertilizers and other chemicals. This has resulted in over application or under application of fertilizers. Over application has resulted health issues and negative effects on bio-diversity. Under application has resulted in reduction of expected yield (Malwattha et al., 2014).

Even with familiarity and expertise in agriculture, human monitoring cannot detect the disease that has infected the crop. However, techniques such as deep neural networks can predict pest and disease attacks. Several research and studies have been conducted to explain the use of pesticides and fertilizers to treat crops affected with diseases. A Convolution Neural Network model was developed which predicted the pest or the disease infecting the crop and a method for managing pesticides which counts the number of pests at a given time by identifying the pixels of successive images for comparison (Miranda, Gerardo and Tanguilig iii, 2014)

The model for this solution had an accuracy rate of 94% and provided the farming community with personalized local language prediction and preventative measure recommendations and helps to predict the countermeasures for diseases by recommending the appropriate quality and quantity of pesticides for paving the way for sustainable agriculture but it doesn't output real time results and disease detection is limited only for two classes of vegetables (Miranda, Gerardo and Tanguilig iii, 2014).

Another research which was conducted in 2019 by a group of researchers from the Sri Lankan Institute of Information Technology focused on recommending the best crop based on the soil's fertility and advised a fertilizer as well as a fertilizer plan to maximize the number of fertilizers used for the suggested crops and educates an individual on the procedures through development of a cross-platform mobile application (Rahaman et al., 2017). This study emphasizes how technology may be used to achieve cultivation goals effectively by obtaining accurate information about the nutritional requirements for successful cultivation (Rahaman et al., 2017). The system achieved an over 80% accuracy rate after completing the end model with the trained data sets while the main limitation of this system being the sensor readings obtained from the sensors that produce humidity, pH, temperature, and EC can't guarantee the accuracy of 100% (Rahaman et al., 2017).

It was observed that usage of pesticides has increased drastically in the country. The continued application of permissible pesticides has paved the path to the contamination of the ecosystem. According to the research done by the University of Peradeniya, the most commonly used pesticide is herbicides which are 70%, few use fungicides which is 4% and 63% of them belongs to class II (moderately hazardous), 31% belong to class U (unlikely to cause hazard) and 6% comes under class III (slightly hazardous). The authors have found poor practices of pesticide usage among farmers. Only a few farmers (18%) reach the government officers and others rely on agrochemical dealers and their own experience. (Dissanayake et al., 2022).

2.3.3 Chatbot

Conversational agents, also known as chatbots have grown significantly over the past few years and in different sectors. With the basic premise that human interactions with the systems mimic typical talks, the term "Conversational Agent" has come to refer to a wide range of systems with varied capabilities and goals. In 1966, ELIZA, the first chatbot was created with the help of substitution methodology and pattern matching to demonstrate the communication between humans and the computer (Weizenbaum, 1983). Later, ELIZA inspired several chatbot applications in the agricultural domain mainly in Asia such as Farmer's assistant: Agriculture bot, E-AGRO, Agroxpert, Agribot and Krushi.

Farmer's assistance: Agriculture bot is used to offer two chatbot options to answer questions from farmers in a variety of languages where the agricultural multi-linguistic voice bot uses Google Translate, Pysttsx3, and Google search engines, while the suggestion bot offers a variety of suggestions to address queries about the crops, soil, weather, fertilizer, etc (M et al., 2021). It's also stated that the chatbot will allow the users to identify the crop diseases as well as the solutions while suggesting the most optimal crop and their harvesting stage using benchmarks such as climate, market price, and soil as a future modification to the system. According to Ekanayake and Saputhanthri (2020), E-AGRO, an intelligent chatbot is also implemented to connect and network farmers to have access and provide remedies to different problems through natural language processing techniques and Artificial Intelligence (AI) technologies.

Ekanayake and Saputhanthri (2020) also stated that the chatbot was trained with limited query inputs as well as a narrow domain which limited its capabilities to answer complex queries while also being a user-friendly decision-making tool to the farmers that provide solutions in real-time.

With the help of natural language processing, machine learning algorithms, and chatterbot python libraries, Agroxpert serves as chatbot assistance by effectively answering all of the farmers' questions and referring them to experts in the event that the proposed approach does not answer the farmers' inquiries (Nayak et al., 2021). Nayak et al. (2021) have also discovered that enabling both audio and video communication with their respective expert, sending queries in the form of pictures and video clips, and language translation would make the users more confident in using this application. Another objective similar chatbot is Krushi – the farmer chatbot, which is developed using AI and machine learning techniques is a top to bottom trainable model; a conversational system with little error and responds to inquiries about crop diseases, crop protection, soil tests, weather, fertilizer, etc. to make the best decisions to increase farmers yield (Momaya et al., 2021). The Farmer Chatbot is integrated into RASA X and according to preliminary findings, Krushi would be able to respond with expert remedies to 70% of all questions posed to the KCC hotline at any time of day with a 96.1% accuracy rate. (Momaya et al., 2021). As stated by Momava et al. (2021), they are hoping to make the application available to users in local languages for better convenience and enhance user-friendliness and adaptability for the farmers.

2.4 Severity Check

Detection and control of plant diseases are crucial factors for reliable production of food. Severity of rice disease is an important indicator for farmers to plan appropriate treatments for protecting crops from diseases which may cause damages. The researchers have used Deep Convolutional Neural Network to classify the diseases and was able to achieve an accuracy of 96.56% (Tendang and Chamnongthai, 2021).

The production of rice is hampered by various kind of paddy diseases. The main disease of paddy is leaf disease. It is very time-consuming and laborious for farmers of remote areas to identify paddy leaf diseases due to unavailability of experts. In some areas, though experts are available disease detection is performed by naked eye which causes inappropriate recognition sometimes. An automated system can minimize these problems. The research have used Kmeans clustering segmentation to segment the image and SVM to classify the images. The system has been able to achieve an overall accuracy of 92.06%. The system doesn’t suggest the most specific pesticide for some diseases, it mentions to use required fertilizers which can be seen as a limitation of the system. (Pinki, Khatun and Islam, 2017).

Edge detection is an image processing technique which is used to detect the boundaries of an object within images. This method works by detecting discontinuities in the brightness. Edge detection mechanism is integrated in the approach which removes image's worthless information and allows for quick feature selection. The results demonstrated that the proposed approach has a total accuracy of 97.692% (Dhiman and Saroha, 2022).

In order to check the severity of the disease a simple equation can be used as (disease affected area)/ (total area of leaf) *100. Such approach was taken by A.D. Nidhis and his team, U_(Ia/(Ia+Iu)) * 100, where Ia are white pixels of the affected area, Iu are white pixels of the unaffected area, and U is percentage value. The calculation was done by the MATLAB program. This calculation provides us the severity of the disease in terms of affected area percentage (Nidhis et al., 2019).

The severity check function helps the farmer to have a real time analysis of the field after the implementation of solutions to the disease affected. Thereby, helps the farmer to have a sound knowledge and proceed with the next steps.

2.4 Comparison table of relevant work

No	Research	Author	Year	Dataset	Model Used	Metric
Disease detection of paddy using image processing						
01	Recognition of diseases in paddy leaves using KNN classifier	M Suresha, K N Shreekanth, B V Thirumalesh	2017	Recognition of paddy diseases are made, RGB images were collected from Lousiana State University Auricular Center	Convolutional Neural Network, Googlenet-Overfeat model in TensorFlow framework	76.59 %
02	Deep Learning-Based Classification for Paddy Pests & Diseases Recognition (Indonesia)	Ahmad Arib Alfarisy, Quan Chen, Minyi Guo	2018	The ImageNet dataset have been used as a benchmark in computer vision image dataset that has 4,511 images for 13 class paddy pests and diseases	CaffeNet, Augmentor, PC, GPU	87%
03	Disease Identification in paddy leaves using CNN based Deep Learning	R Swathika, S Srinidhi, N Radha, K Sowmya.	2021	paddy leaves were assembled from public dataset platforms like Kaggle and other online platforms	image classification using CNN, ANN, SVM	70%
04	Paddy Plant Disease Classification and Prediction Using Convolutional Neural Network	G K Sagarika, SJ Krishna Prasad, S Mohana Kumar	2020	The dataset collected has around 60 images of paddy plant leaves. These images are collected from agricultural college web site and internet sources.	MATLAB, CNN, Hardware	94.12 %

Weed detection of paddy using image processing						
01	Classification of weed species in the paddy field with DCNN-Learned features	Xue Yan, Xiangwu Deng, Jing Jin	2020	Used a pretrained model with over one million images	SVM,DCNN,KNN	94.5%,
02	Deep Learning Based Overcomplete Representations for Paddy Rice Crop and Weed Segmentation	G Ujwal Sai, N Tejasri, Ajay Kumar, P Rajalakshmi	2022	The dataset was collected using two methods: static camera Sony RX 100 mounted on a tripod with a fixed height and angle to capture the data and UAV	CNNs	72.5%
03	Weed detection in paddy field using an improved RetinaNet network	Hongxing Peng ZiheLi ZhiyanZhou YuanyuanShao	2022	cranesbill seedling dataset, The dataset built in this work is containing more than six thousand images of rice and eight categories of weed.	SVM,GLCM,BA, ANN-BA	94.1%
04	Weed Detection in Farm Crops using Parallel Image Processing	S Umamaheswari, R Arjun, D Meganathan	2018	S. Haug, J. Ostermann and R. Bosch, "A Crop /Weed Field Image Dataset for the Evaluation of Computer Vision Based Precision Agriculture Tasks"	Convolutional Neural Network, Googlenet-Overfeat model in TensorFlow framework.	91.1%
Chatbot						
01	E-AGRO: Intelligent Chat-Bot. IoT and Artificial Intelligence to Enhance Farming Industry	Jayalath Ekanayake, Luckshitha Saputhanthri	2020	More than 50 test cases of E-AGRO users	-	-

02	Agroxpert – Farmer assistant using NPL, Chatterbot libraries and Django framework	Vandana Nayak, Pranav R Nayak N, Sampoorna, Aishwarya, N.H.Sowmya	2021	Data collected from Agrixpert users	Chatterbot	96%
03	Krushi – The Farmer Chatbot developed using Machine learning and artificial Intelligent techniques	Mihir Momaya, Anjnya Khanna, Jessica Sadavarte, Manoj Sankhe	2021	Data collected through KCC official website	RASA X	96.1%
04	Agribot: A Natural Language Generative Neural Networks Engine for Agricultural Applications	Bhavika Arora, Dheeraj Singh Chaudhary, Mahima Satsangi, Mahima Yadav, Lotika Singh, Prem Sewak Sudhish	2020	1000 iterations	-	98%
Severity Check						
01	Rice-Disease Severity Level Estimation Using Deep Convolutional Neural Network	Sayan Tepdang and Kosin Chamnongthai	2021	500 images were taken from each disease. Therefore, 2500 images were used in total.	CNN, K- mean segmentation	96.56 %
02	Detection of Severity of Disease in Paddy Leaf by integrating Edge Detection to CNN-based Model	Anupriya Dhiman, Vinod Saroha	2022	650 images	CNN	97.69 %

Table 7: Comparison table of relevant work

2.5 Chapter Summary

The literature review chapter has provided an overview of the current state of knowledge related to paddy cultivation in Sri Lanka. The significance of rice as a staple food in Sri Lanka has been highlighted, and the challenges faced in paddy cultivation have been discussed, including the damage caused by pests, diseases, and weeds. Overall, the literature review highlights the need for innovative solutions to address the challenges faced by farmers in paddy cultivation and provides a framework for the research approach adopted in this study.

CHAPTER 3: METHODOLOGIES

3.1 Chapter Overview

The methodology chapter provides a detailed account of the research design, data collection methods and sources, and the machine learning and image processing techniques used in the study. This chapter also describes the proposed chatbot and field progress checking methods. The research design outlines the overall approach, including the sample selection, data collection procedures, and analysis techniques. The data collection methods and sources describe the process of gathering data, including the types of data collected and the tools and instruments used. The machine learning and image processing techniques used in the study are described, highlighting their strengths and limitations in detecting paddy diseases and weeds. Finally, the proposed chatbot and field progress checking methods are presented as potential solutions to address the challenges faced by farmers in paddy cultivation.

3.2 Research Methodology

Research Philosophy	- A research philosophy describes how data about a phenomenon should be collected, analyzed, and used. There are four main philosophies: pragmatism, positivism, realism, and interpretivism (Dudovskiy, 2011). - This work adheres to positivism, which believes that only "factual" knowledge obtained through observation, including measurement, is reliable. Research in positivism focuses on collecting and interpreting data objectively. - In our research, we accumulate images related to diseases, weeds, and facts related to paddy agriculture which resembles positivism.

Research Approach	- There are two main approaches during research termed as inductive and deductive. Below depicts the Inductive approach  *Figure 3: Inductive approach* - The most suitable approach is the Inductive approach because observations of disease and weed are done first, and a pattern is recognized to achieve the final conclusion.
Research Strategy	- A research strategy is the general phrase for the techniques utilized to carry out quantitative and qualitative approaches. - In this research, the researchers collect data in the form of images of diseased leaves and weeds related to paddy plants. - Therefore, this research utilizes the qualitative approach, where the exterior features of the leaves are used for the classification of weeds, disease, and its severity level.
Research Choice	- Research choice methodology consists of mono, multi and mixed methods. - The technical knowledge required is gathered through in-depth interviews done with field experts, farmers, and research institutes. Therefore, this research project would be utilizing the mono-research choice method.
Time zone	- This establishes the research time frame as splitting into two types as cross sectional and longitudinal. - The cross-sectional method would be fitting this project as the data gathering will be covering only a certain time zone.

Table 8: Research Methodology

3.3 Development Methodology

Life-Cycle Model

A system for task planning, evaluation techniques, and control is known as a software development methodology. A productive software development methodology describes how all our tools, strategies, and procedures come together to make an effective solution. Out of which the **prototyping model** would be the suitable model to be implemented in this project as the system is only partially implemented. A prototype of the product is first created, tested, and refined based on customer feedback repeatedly until a final acceptable prototype is achieved, which serves as the basis for creating the final solution.

Design Methodology

Design methodology describes the process of creating a system or approaches for a certain circumstance (Learn.org, 2016), the two main concepts of this methodology are Object Oriented Analysis & Design (OOAD) & Structured System Analysis and Design Modelling (SSADM). Out of the two models the OOAD fits this project since the SSADM model is designed for large scale companies, and it is time consuming when compared to OOAD (Schumacher, 2002). The **OOAD model** is flexible and able to tackle more challenging problem domains.

Evaluation Methodology

There are different types of evaluation metrics available for testing models. Evaluation metrics are used to gauge the quality of the machine learning model. Classification accuracy, logarithmic loss, confusion matrix, and other metrics are some of these. The number of correct and incorrect predictions made by the model in relation to the actual classifications in the test set, as well as the kind of errors being made, are displayed in a table called a confusion matrix. This matrix shows how well a classification model performs when tested on test data with known true values.

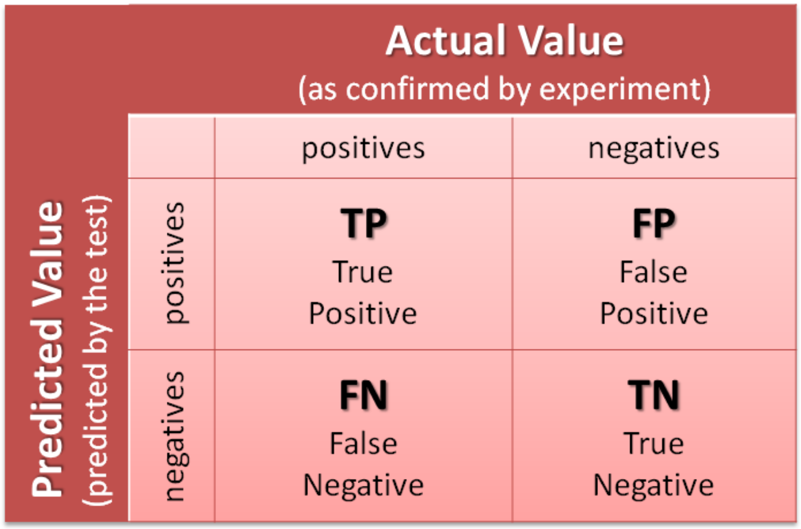


Figure 4: Confusion matrix

In above image, rows reflect the count of predicted classifications of the model, while columns represent the count of actual classifications in the test data.

Accuracy of the model can be calculated as below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision of the model can be calculated as below.

$$Precision = \frac{TP}{TP + FP}$$

3.4 Project Management Methodology

Gantt Chart

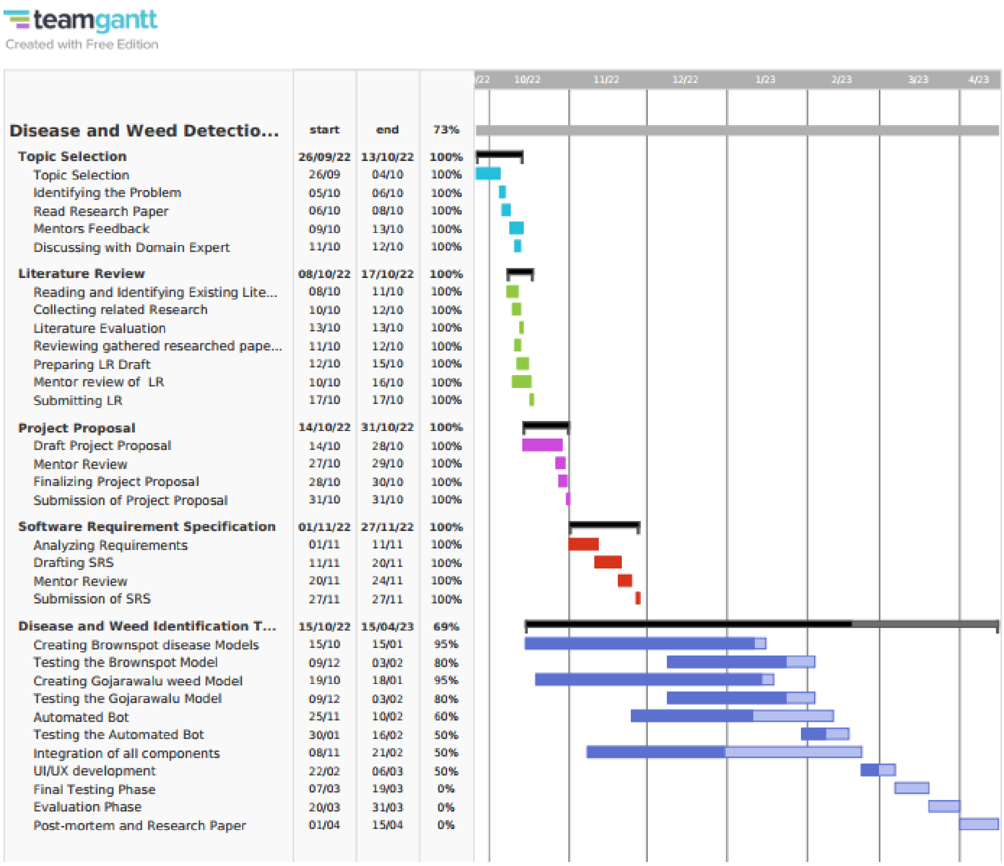


Figure 5: Gantt Chart

Deliverables, milestones and due of deliverables

Deliverables / Milestones	Due Date
Submission of Literature Review	17th October 2022
Submission of Project Proposal	31st October 2022
Software Require Specification Submission	27th November 2022
UI/UX Development	25th December 2022
Prototype development	5th March 2023

Test Report and testing phase	19th March 2023
Final Project Report Submission	9th April 2023

Table 9: Deliverables Due dates

3.5 Chapter Summary

The methodology chapter outlines the research design, data collection methods and sources, and the machine learning and image processing techniques used in the study. The chapter begins by describing the research design, including the sample selection, data collection procedures, and analysis techniques. The data collection methods and sources are then presented, outlining the process of gathering data, including the types of data collected and the tools and instruments used. Finally, the machine learning and image processing techniques used in the study are described, providing a detailed account of the research approach, and justifying the methods used to achieve the study's objectives.

CHAPTER 4: SOFTWARE REQUIREMENTS SPECIFICATION

4.1 Chapter Overview

This chapter investigates the procedure of collecting the system requirements and the steps taken to analyze the information collected throughout the process. A stakeholder analysis is shown at the beginning of the chapter by listing down a description of the responsibilities they have towards the system. Then the different methods for reviewing the requirements gathered are pointed out by differentiating the advantages and disadvantages of each method used, Furthermore, the use case diagram and its descriptions are included in the requirement analysis part. At the end a scope definition is used which specifies the functional and non-functional requirements of the system and is categorized based on their priority.

4.2 Rich Picture

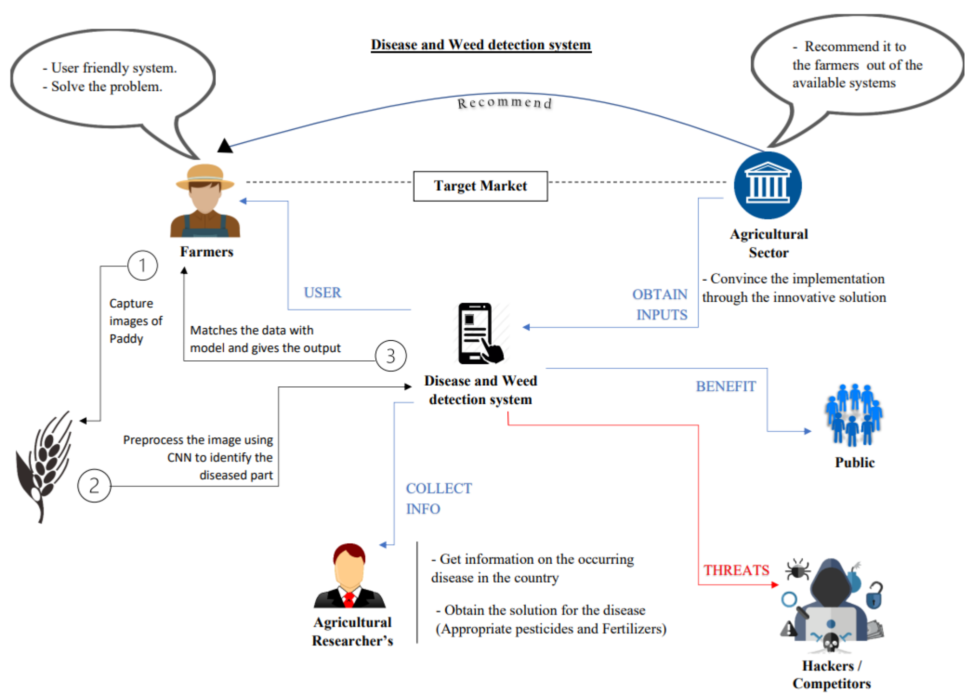


Figure 6: Rich Picture

4.3 Stakeholder Analysis

The system's stakeholder groups, together with their corresponding functions and positions, are described in the onion model.

4.3.1 Onion Model

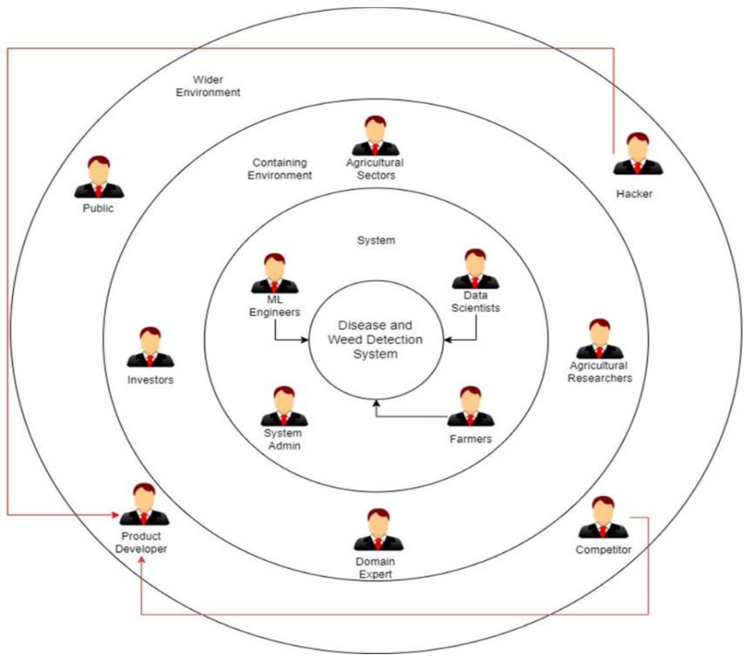


Figure 7: Onion Model

4.3.2 Stakeholder Viewpoints

Stakeholder	Roles	Benefits
System		
o Data Scientists o Farmers o System Admins o ML Engineers	The whole system is maintained by these stakeholders	Designing the paddy disease and weed detection model and implement them
Containing Environment		
o Agricultural Sectors	Quality Beneficiary	Make sure that the application is up to standards and is of a certain quality
o Agricultural Researchers	Benefits from the Education	Research on the current available systems and plans out the possible approaches and techniques to implement the program.

o Domain Expert	Experts	Detects whether the plan of implementation is feasible and possible
o App User o APIs o Dependent Systems	Functional Beneficiaries	Uses the developed apps in different instances and implement it in other scenarios
o Investors	Financial Beneficiaries	Shareholders of the Entire System
Wider Environment		
o Public	Negative	Provides in-depth knowledge, suggests alternative approaches, and provides feedback related to the techniques, algorithms and methodologies used in the system.
o Product developer	Operational Maintenance Head of Development	In charge if the system making and maintenance
o Hacker	Negative Stakeholder	A person or group of people gains unauthorized application access in order to make changes to its data or contains
o Competitors	Negative Stakeholder	Develops the system almost the same as this and brings potential competition among the market

Table 10: Stakeholder Viewpoints

4.4 Selection of Requirement Elicitation Techniques

Researching and learning the needs of a system from users, clients, and other stakeholders are known as requirements elicitation. To gather information and determine the project's needs, important stakeholders are consulted during the requirement elicitation process. Various requirement elicitation strategies were used in this study to collect needs and this section below provides an analysis of the applicability of various strategies and the selection justification and detailing the merits and drawbacks of each.

4.4.1 Literature Review and Existing System Documentation Observation

To identify the need for updates or enhancements, this method is used to review the current documentation. Before organizing further in-depth needs elicitation sessions or stakeholder interviews, existing system observation is carried out. As this is an extension of the domain research, a literature review is done to identify and validate the research gap. Thus, gaps in the Agricultural domain, the best CNN approaches, and image processing techniques were investigated in LR.

Advantages	Disadvantages
• Helps to identify the components that need more enhancement or that have gaps so that they can be used as a foundation for further study • Helps to identify limitations and gaps in the existing work • Can get a proper idea on how to start the implementation process while searching for improvements in the systems	• It's a time-consuming task as finding usable information would require sifting through a lot of unnecessary data • As different researchers have diverse objectives to achieve, existing system reviews can be a complicated method

4.4.2 Surveys and Questionnaires

Surveys and questionnaires are useful tools for connecting with stakeholders and gathering information that may be utilized to collect, examine, and comprehend the perspectives and opinions of a group of individuals. The major goal of the survey/questionnaire was to validate the stated problem by gathering requirements and necessary information, particularly from the target stakeholders, farmers.

Advantages	Disadvantages
• Can reach out to a larger group of people around the country to collect data	• All the needed questions might not be covered in the questionnaire

• Takes much less time for the participants to answer the questions • Analyzing the gathered data can be done easily	• The participants might misunderstand a question • Expected answers might not be received from every participant which would complex the analysis of the data

4.4.3 Interviews

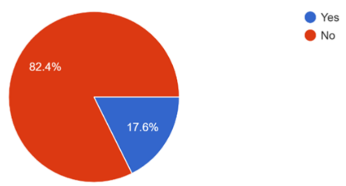
Researchers, specialists in the paddy research and agriculture fields, as well as academic and industrial experts, were taken as the target audience for the interviews since they have a thorough understanding of the subject matter, its existing practices, and its limits. Unstructured one-on-one interviews were chosen as a result because it is feasible to reach the relevant target audience.

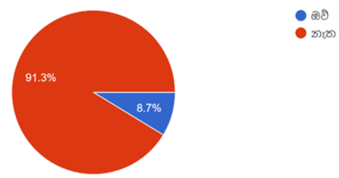
Advantages	Disadvantages
• Interacting directly with the audience is efficient • Can clarify doubts in an effective manner	• It's a time-consuming process as meeting a large number of people individually takes a lot of time • The answers given might not be straightforward

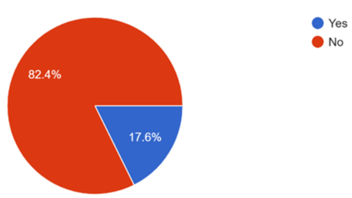
4.5 Discussion of Results – Questionnaire

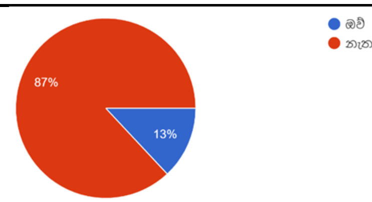
With the help of the main targeted stakeholders; farmers in paddy fields, using separate Sinhala and English language questionnaires we were able to validate the identified limitations and gather their thoughts on introducing a paddy and weed detection mobile application.

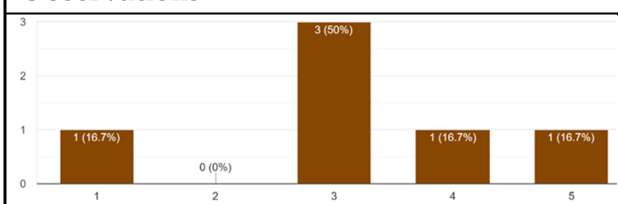
Question	Have you ever used any paddy disease detection mobile applications?
Aim of Question	To Identify the popularity of paddy disease detection systems
Observations	

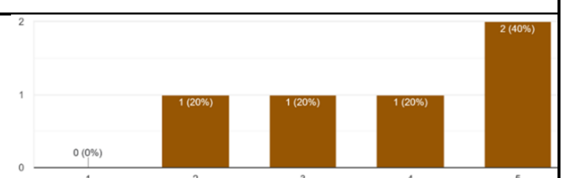
<table>
<tr><td colspan="2">



It is observed that 88% of the participants have not used any paddy disease detection application in their life while 12.5% have used such application in the field.</td></tr>
<tr><td colspan="2">Conclusion</td></tr>
<tr><td colspan="2">According to the questionnaire findings, we can see that most farmers have never used any paddy disease mobile applications while working in the field and it shows that Sri Lankan farmers are not familiar working with the technology.</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Question</td><td>Have you ever used any paddy weed detection mobile applications?</td></tr>
<tr><td>Aim of Question</td><td>To Identify the popularity of paddy weed detection systems</td></tr>
<tr><td colspan="2">Observations</td></tr>
<tr><td colspan="2">



It is observed that 84% of the participants have never used any paddy weed detection application while working in the field to recognize paddy weeds while only 15% have used such application in their life</td></tr>
<tr><td colspan="2">Conclusion</td></tr>
<tr><td colspan="2">When looking into the usage of paddy weed detection mobile applications among Sri Lankan farmers, we can also see that most farmers have never used any such applications in the field. This shows that the farmers do not tend to use technology in their field of work.</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Question</td><td>If applicable, rate the satisfactory level of the mobile application you used</td></tr>
<tr><td>Aim of Question</td><td>To Identify the satisfaction of the users of the applications they used</td></tr>
<tr><td colspan="2">Observations</td></tr>
<tr><td colspan="2">

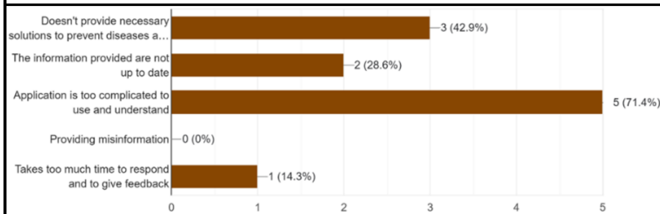


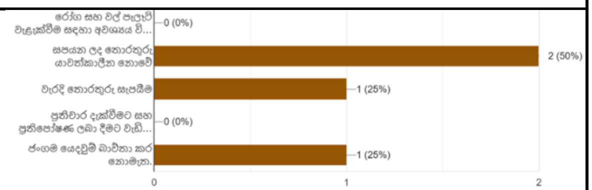
According to the above graphs, we can observe that 36% of the application users have a neutral level of satisfaction with the usage of paddy disease or weed disease applications they used while 27% have high satisfaction with their used applications.</td></tr>
<tr><td colspan="2">Conclusion</td></tr>
</table>

From the above findings, we can conclude that most of the participants who have used paddy disease or weed detection applications before, have a neutral level of satisfaction which indicated that they have not found those applications to be at their best while some participants have high satisfaction on the applications they used.

Question	If you were not satisfied, what are the issues you found in the mobile application you used?
Aim of Question	To Identify the issues of the used applications the users found

Observations





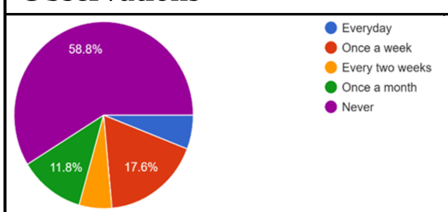
It is observed that 34% of the participants consider that the applications they used were too complicated to use while 26% and 20% of the participants think that the information provided on the applications is out of date and does not provide necessary solutions to prevent paddy diseases and weeds respectively. 7% of the participants each have also thought that misinformation being provided, and slow feedback has also been some of the issues they found.

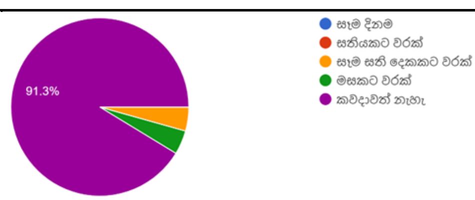
Conclusion

According to the above findings, we can conclude that the users of the paddy disease and weed detection applications have considered the lack of user-friendliness of the applications to be the biggest issue in the existing applications along with not providing solutions for the diseases and weeds found and the information that has been provided is outdated.

Question	How often do you use such applications to detect paddy diseases or weeds?
Aim of Question	To Identify the frequency of the usage of such applications in the field

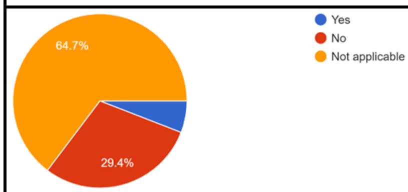
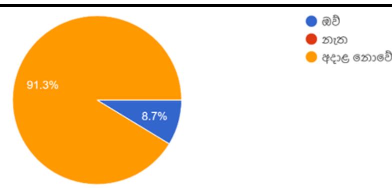
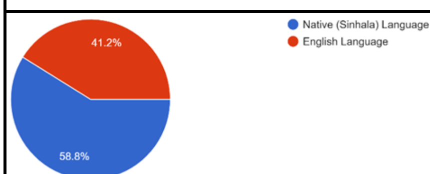
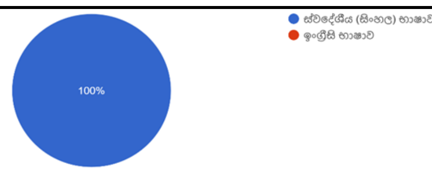
Observations

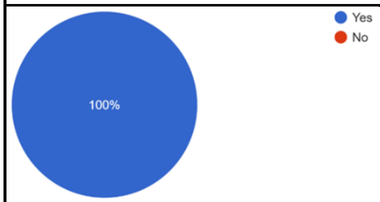
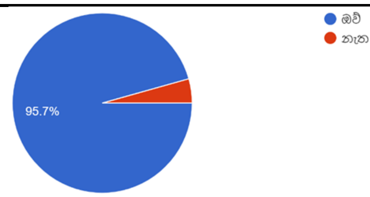
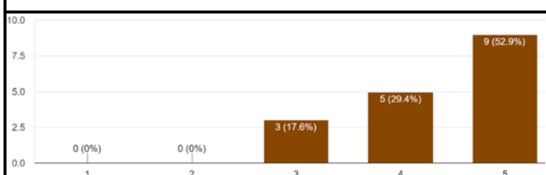
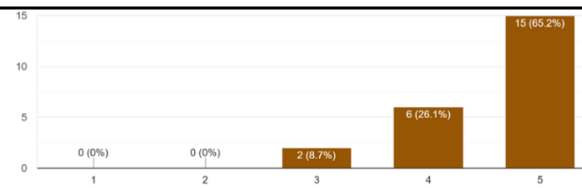


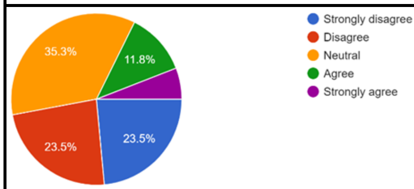


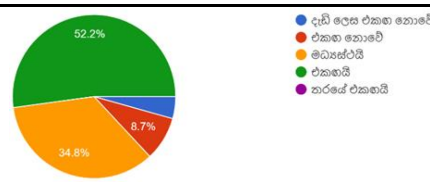
It is observed that 78% of the participants have never used any paddy disease or weed detection applications while 7.5% of the participants have been using such applications once a week and once a month respectively. Also, 5% and 3% of the participants use such applications every two weeks and every day respectively.

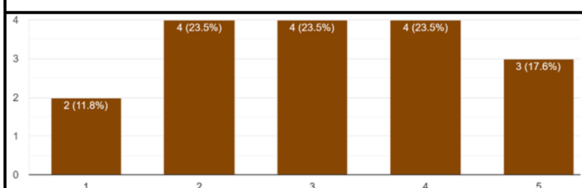
Conclusion

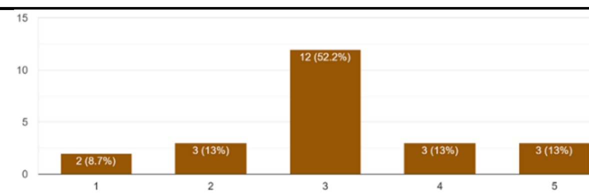
From the above pie chart, we can conclude that most of the participants have never used any kind of paddy disease and weed detection applications and the participants who use such applications also do not use them regularly and have difficulty using them in their day-to-day field activities.	
Question	Is the application you used were in Native Language? Eg: Sinhala
Aim of Question	To Identify whether the applications used were available to the users in their native language
Observations	
  It is observed that 80% of the participants mentioned that the above question asked is not applicable while 13% of the participants answered that the applications they have come across were not available in their native (Sinhala) language. 8% of the participants mentioned the applications were available in their native language.	
Conclusion	
According to the above survey, we can see that the participants choosing the "not applicable" option was due to them not using any applications regarding paddy disease and weed detection applications. We can also see that Sinhala language support on such applications are also in low frequency.	
Question	Would you prefer such application to be in Native language(Sinhala) or English language?
Aim of Question	To Identify the preference of the users of the application's language
Observations	
  It is observed that 83% of the participants prefer paddy disease and weed detection applications to be in their native (Sinhala) language while 18% of the participants prefer such applications to be in the English language	
Conclusion	
From the above pie chart, we can conclude that most of the farmers are more comfortable in using applications that would be familiar to their mother tongue as it's more convenient and easier to use in any given time. The reason must be that most farmers are in their late 50s, their English literacy is lower than Sinhala literacy.	

Question	Would you like to use an application which helps you to detect Paddy disease and weeds accurately in the same application?
Aim of Question	To Identify the preference of the users to be able to have an application which would allow detecting both paddy disease and weeds.
Observations	
  98% of the participants have given a positive response in relation to using an application that would help detect paddy disease and weeds while 2% have given a negative response.	
Conclusion	
According to the above survey findings, we can say that almost all the participants prefer to use an application that would aid them in detecting paddy disease and weeds using the same application if a new application were to be introduced in the future.	
Question	How useful do you think it will be to have an application that accurately detects paddy disease and weeds?
Aim of Question	To Identify the users' thoughts on the usefulness of a paddy disease and weed detection application
Observations	
  It is observed that 60% of the participants agreed that having an application that would accurately detect paddy diseases and weeds would be extremely useful while 28% and 13% of the participants think it would be very useful and moderately useful respectively. No participants think that such an application would not be useful.	
Conclusion	
According to the above-collected data, we can see that overall, 88% of participants think that it is useful to have an application that would aid them in detecting paddy disease and weed accurately. So comparing Questions 1 and 2 with the current findings of the question, we can conclude that most of the participants are willing to use such an application given the opportunity.	
Question	Is getting professional help from officials and getting their feedback an easy task for you? Eg: easy to contact, easy to meet..

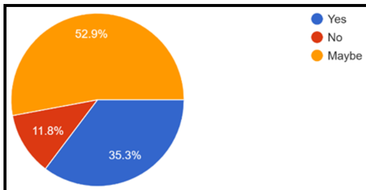
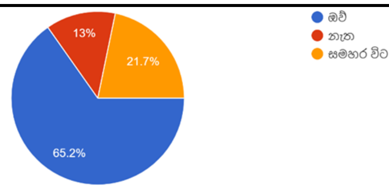
<table>
<tr><td>Aim of Question</td><td>To Identify whether the users are able to gather relevant information from professionals in an efficient manner</td></tr>
<tr><td colspan="2">Observations</td></tr>
<tr><td colspan="2">



It is observed that 35% each agree and have a neutral opinion that they are able to contact the field experts and professionals to gather relevant information effectively while 15% and 13% of participants disagree and strongly disagree on the matter. There are 3% of the participants strongly agree on the matter as well.</td></tr>
<tr><td colspan="2">Conclusion</td></tr>
<tr><td colspan="2">It is shown that currently, most farmers are not happy; have less positive opinions regarding the fact that they are not able to contact and receive relevant information regarding paddy agriculture from the field professionals and officers.</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Question</td><td>Rate how rapidly you were able to get professional solutions to questions regarding paddy agriculture</td></tr>
<tr><td>Aim of Question</td><td>To Identify the competency of getting solutions to users problems through professional aid</td></tr>
<tr><td colspan="2">Observations</td></tr>
<tr><td colspan="2">



40% of the participants have a neutral opinion on how rapidly they were able to receive solutions, feedback and professional aid regarding the problems arising in paddy agriculture while 18% of each have low and high opinion on the matter. While 15% of participants have an extremely high opinion, 10% have an extremely low opinion on the above matter.</td></tr>
<tr><td colspan="2">Conclusion</td></tr>
<tr><td colspan="2">According to the above results, we can see that just as the previous question's overall result of farmers not being happy about not being able to contact the field experts, getting rapid solutions and feedback from the experts and professionals are also not in a favorable position.</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Question</td><td>Are you comfortable in providing your personal information if the professionals were to reach back to you?</td></tr>
<tr><td>Aim of Question</td><td>To Identify whether the users have concerns regarding their privacy</td></tr>
<tr><td colspan="2">Observations</td></tr>
</table>

 While 53% of participants are comfortable with providing their personal information to the application, 35% of participants are in a doubtful situation where they are not fully comfortable. 13% of participants are not comfortable providing their personal information.
Conclusion
As keeping the customer's privacy safe is the responsibility of the application developer, the customers/users may feel doubtful about providing their personal information to any application provider. Even though almost half of the participants are comfortable in providing their information, the other half are either doubtful or not comfortable in giving out their information.

4.6 Summary of Findings

Findings	Literature Review	Questionnaire	Interviews
There are very few applications active in Sri Lanka that can be used to detect paddy disease and weeds	X	X	X
There are no existing applications in Sri Lanka that support both detections of paddy disease and weeds in a single application	X		X
There is a lack of Sinhala language-supported applications available	X	X	X
It is more effective to use CNN for feature extraction than other conventional techniques.	X		
GUI should be easy to use and understand for the users		X	X
The application should provide effective and efficient ways and means that would provide solutions, professional aid, and feedback	X	X	X
Lak of datasets available in the context of Sri Lankan paddy weed species is a major issue	X		X

Table 11: Summary of Findings

4.7 Context Diagram

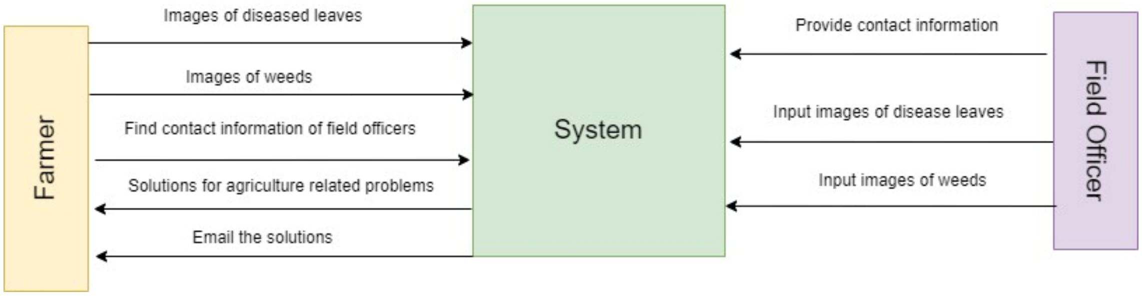


Figure 8: Context Diagram

4.8 Use case Diagram

Figure 9: Use case Diagram

4.8.1 Use Case Descriptions

Only one use-case explanation is offered here due to page constraints. Please see the appendix for details of more use cases.

<u>Use Case Description 01 – Farmer, Field Officer</u>

Use case name	To determine the disease.
Actors	Farmer, Field officer
Purpose	To determine the name of the disease infected to the plant
Overviews	Farmer needs to determine the disease of the Farmer inputs an image to the app Application generates the result
Pre- conditions:	Farmer and Field officer should register to the app. Farmer and Field officer should have a diseased plant.
Post- conditions:	Determined the disease, display the severity level and provide necessary recommendations.

Typical course of events:

Actor action	System response
(01) Farmer logins to the application (03) Farmer inputs an image to the application (05) Farmer inputs an image of incorrect format	(02) Application allows to login (04) Application generates the result (05) Application displays an error message.
Alternative Case	No diseased plant
Includes	Input image of infected leaf Validate image Check for severity of diseased leaf Recommend solution
Extend	Email the solution

4.9 Functional Requirements

The system's functional requirements are listed in the table below, along with their priority level.

	Requirement and Description	Priority
FR-01	Capture image *The system allows the user to take images*	Critical

FR-02	Accepting images taken	Critical
	As an input, the app must be able to accept images taken	
FR-03	Isolating the diseased part of the paddy plant	Critical
	Since the normal parts of the paddy plant would interfere / mislead the tool, isolation of the diseases part of the plant should be done	
FR-04	Giving a severity level of the paddy disease according to the uploaded image	Critical
	After the disease identification is done a second scan is done to identify the severity level of the disease (different data set is provided according to the level of the disease)	
FR-05	Generate evaluation results to the model	Critical
	Even though for end user, it is not important function, in research perspective the accuracy and performance evaluation are an must feature.	
FR-06	Support different qualities of images	Non-Critical
	The quality of the image should be varied, due to users having different types of phones.	
FR-08	GUI and other Interface support	Critical
	User friendly UI and API interfaces to connect with the system	
FR-09	Produce Result	Critical
	produce outcomes based on image processing.	

Table 12: Functional Requirements

4.10 Non-Functional Requirements

Accuracy

Accuracy of the system is very important, and it should identify the diseases or weed correctly to supply the needed remedy. more errors in text transcript means, the system or the model is not usable.

Performance

The data size of the train and test set is very large (diseased paddy and the weed) and will increase with time. So that the model training time will be longer when using more data.

Usability

The whole system could be operated via command prompt, but in-order to make the tool user friendly a UI with the ability to upload the image which is saved in the phone.

Security

Since the users have to login using username and password, the system should be secured to avoid unauthorized access and misuse of data.

	Requirement and Description	Specification	Priority
NFR - 01	The accuracy of the model should be high	Accuracy	Important
NFR - 02	Image processing to remedy recommendation should not take a very long time	Performance	Important
NFR - 03	The process should be done with minimum hardware utilization	Performance	Important
NFR - 04	User friendly interface for image uploading	Usability	Non-Important
NFR - 05	User data should be secured to prevent authorized access	Security	Non-important
NFR - 06	Should be able to increase the hardware configuration	Scalability	Non-Important

Table 13: Non-Functional Requirements

4.11 Chapter Summary

A concise summary of the Software Requirement Specification has been provided in this chapter. In this chapter, the identification of internal participants in the project and the extent of their involvement was examined. Also explained were the requirements for the elicitation, the results, and the summary of findings. Additionally, priority lists for both functional and non-functional criteria were created.

CHAPTER 5: SOCIAL, LEGAL, ETHICAL, AND PROFESSIONAL ISSUES

5.1 Chapter Overview

All different kinds of external impacts on the developed product can be evaluated with the aid of the SLEP analysis which makes it easier to view the overall, macro-environment. All of the legal, social, ethical, and professional elements of our project were thoroughly examined in this chapter.

5.2 SLEP Issues and Mitigation

5.2.1 Social Issues

- Since the majority of Sri Lankan farmers lack the ability to read English fluently, the app would be available in both English and Sinhala to increase the user experience
- The images taken and uploaded to the app would only be used to identify the disease or the weed, and after that part is done, those images would be deleted

5.2.2 Legal Issues

- Property rights: The ownership of the technology and data used in the development of the plant disease and weed detection system may be subject to legal issues
- Regulation compliances: The system must comply with Sri Lankan regulations related to the use of AI technology in agriculture, data privacy, and intellectual property rights
- Data privacy laws and regulations must be considered when collecting and storing data from the system

5.2.3 Ethical Issues

- Assure all users that any information they provide through Google forms will be kept private and never be disclosed.
- Only the datasets that adhered to a solid ethical framework were chosen and developed in order to build the AI model.
- The dataset which we self-made by the group members was also validated by a professional and created according to proper ethics.

5.2.4 Professional Issues

- There may be a lack of technical expertise in developing and implementing the system in Sri Lanka
- Collaboration with local agricultural professionals, such as extension officers, may be crucial in developing effective and sustainable solutions
- Training and capacity-building initiatives may be necessary to ensure farmers can effectively use the technology

5.3 Chapter Summary

This chapter outlines clearly the product's social, legal, ethical, and professional problems as well as possible solutions to overcome those issues.

Referencing

Alfarisy, A.A., Chen, Q. and Guo, M. (2018). Deep learning based classification for paddy pests & diseases recognition. *Proceedings of 2018 International Conference on Mathematics and Artificial Intelligence*. doi:10.1145/3208788.3208795.

Andrianto, H., Suhardi, Faizal, A. and Armandika, F. (2020). *Smartphone Application for Deep Learning-Based Rice Plant Disease Detection*. [online] IEEE Xplore. doi:10.1109/ICITSI50517.2020.9264942.

Anthonys, G. and Wickramarachchi, N. (2009). An image recognition system for crop disease identification of paddy fields in Sri Lanka. *2009 International Conference on Industrial and Information Systems (ICIIS)*. doi:10.1109/iciinfs.2009.5429828.

Arora, B., Chaudhary, D.S., Satsangi, M., Yadav, M., Singh, L. and Sudhish, P.S. (2020). *Agribot: A Natural Language Generative Neural Networks Engine for Agricultural Applications*. [online] IEEE Xplore. doi:10.1109/IC3A48958.2020.233263.

Bandara, D. and Mayurathan, B. (2021). *Detection and Classification of Rice Plant Diseases using Image Processing Techniques*. [online] Sabaragamuwa University of Sri Lanka. Available at: http://repo.lib.sab.ac.lk:8080/xmlui/handle/123456789/1733 [Accessed 20 Oct. 2022].

Charitha, P.L., Mydhili, M., Khyathi, N., Pavithra, P. and Anuradha, G. (2021). Detection of Weed Plants Using Image Processing and Deep Learning Techniques. *Lecture Notes in Networks and Systems*, pp.523–532. doi:10.1007/978-3-030-84760-9_44

Dhiman, A. and Saroha, V. (2022). Detection of Severity of Disease in Paddy Leaf by Integrating Edge Detection to CNN-Based Model. *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*. doi:10.23919/indiacom54597.2022.9763128.

Dissanayake, D.M.P.N.K., De Silva, S.N.T., Pathmarajah, S., Kodagoda, K.A.D.A., Chandimal, T.M.R. and Herath, H.M.T.D. (2022). Pesticide Usage Pattern in Rice Cultivation in Trincomalee District in Sri Lanka. *Tropical Agricultural Research*, 33(2), p.173. doi:10.4038/tar.v33i2.8474.

Dudovskiy, J. (2011). *Research Philosophy - Research Methodology*. [online] ResearchMethodology. Available at: https://research-methodology.net/research-philosophy/.

Edirisinghe, T., Nuwan, S., Silva, D., Swarna Herath, H., Amarasingha, R., Ariyarathne, M., Marambe, B. and De, S. (2022). *EasyChair Preprint Identification of Major Weed Types in Broadcasted Paddy Fields Using Multispectral UAV Images Identification of Major Weed Types in Broadcasted Paddy Fields Using Multispectral UAV Images*. [online] Available at: https://easychair.org/publications/preprint_open/dlcB [Accessed 29 Oct. 2022].

Ekanayake, J. and Saputhanthri, L. (2020). E-AGRO: Intelligent Chat-Bot. IoT and Artificial Intelligence to Enhance Farming Industry. *Agris on-line Papers in Economics and Informatics*, 12(01), pp.15–21. doi:10.7160/aol.2020.120102.

Fang, Y. and Ramasamy, R. (2015). Current and Prospective Methods for Plant Disease Detection. *Biosensors,* [online] 5(3), pp.537–561. doi:10.3390/bios5030537.

Hu, D., Ma, C., Tian, Z., Shen, G. and Li, L. (2021). *Rice Weed detection method on YOLOv4 convolutional neural network*. [online] IEEE Xplore. doi:10.1109/CAIBDA53561.2021.00016.

International Trade Administration (2021). *Sri Lanka - Agricultural Sector*. [online] www.trade.gov. Available at: https://www.trade.gov/country-commercial-guides/sri-lankaagricultural-sector

Learn.org (2016). *What Is Design Methodology?* [online] Learn.org. Available at: https://learn.org/articles/What_is_Design_Methodology.html.

M, K.D., M S, D., V, V.K., Jaincy D E, M., R A, K. and Kumar R M, S. (2021). FARMER'S ASSISTANT using AI Voice Bot. *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*. doi:10.1109/icspc51351.2021.9451760.

Malwattha, S.S., Weliwita, T.C., Rathnayake, U.A., Ekanayake, Y. and Sandaruwan, D. (2014). A novel approach to estimate nitrogen content of paddy. *2014 14th International Conference on Advances in ICT for Emerging Regions (ICTer)*. doi:10.1109/icter.2014.7083917.

Ministry of Agriculture (2022). *Ministry of Agriculture - Sri Lanka - The Department of Agrarian Development emphasize that even a single farmer in Sri Lanka has not kept away from farming during the past few years*. [online] www.agrimin.gov.lk. Available at: https://www.agrimin.gov.lk/web/index.php/news-scroll/1828-11-10-2022-3e?lang=en [Accessed 29 Oct. 2022].

Miranda, J.L., Gerardo, B.D. and Tanguilig Iii, B.T. (2014). *(PDF) Pest Detection and Extraction Using Image Processing Techniques*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/271305520_Pest_Detection_and_Extraction_Using_Image_Processing_Techniques [Accessed 6 Oct. 2022].

Momaya, M., Khanna, A., Sadavarte, J. and Sankhe, M. (2021). Krushi – The Farmer Chatbot. *2021 International Conference on Communication information and Computing Technology (ICCICT)*. doi:10.1109/iccict50803.2021.9510040.

Nayak, V., Nayak, P., Sampoorna, Aishwarya and Sowmya, N.H. (2021). Agroxpert - Farmer Assistant. *Global Transitions Proceedings*. doi:10.1016/j.gltp.2021.08.016.

Nidhis, A.D., Pardhu, C.N.V., Reddy, K.C. and Deepa, K. (2019). Cluster Based Paddy Leaf Disease Detection, Classification and Diagnosis in Crop Health Monitoring Unit. *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, pp.281–291. doi:10.1007/978-3-030-04061-1_29.

Pabitha Muthu, P., Hemalatha, M. and Atshaya G. (2021). WEED DETECTION USING IMAGE PROCESSING. *Research Gate*. [online] doi:10.13140/RG.2.2.15116.64642.

Peng, H., Li, Z., Zhou, Z. and Shao, Y. (2022). Weed detection in paddy field using an improved RetinaNet network. *Computers and Electronics in Agriculture,* 199, p.107179. doi:10.1016/j.compag.2022.107179.

Pinki, F.T., Khatun, N. and Islam, S.M.M. (2017). Content based paddy leaf disease recognition and remedy prediction using support vector machine. *2017 20th International Conference of Computer and Information Technology (ICCIT)*. doi:10.1109/iccitechn.2017.8281764.

Rahaman, S., M, H., R, A., YR, B. and M, C. (2017). Detection of NPK Ratio Level Using SVM Algorithm And Smart Agro Sensor System. *International Journal of Latest Research in Engineering and Technology (IJLRET)* ||, [online] 03(07), pp.11– 15. Available at: http://www.ijlret.com/Papers/Vol-3-issue-7/3-B2017181.pdf.

S.S.B.P.S., P. and V.G.T.N., V. (2022). Identification of Paddy Leaf Diseases using Machine Learning Techniques. *International Journal of Computer Applications*, 183(49), pp.1–5. doi:10.5120/ijca2022921898.

Sagarika, G.K., Krishna Prasad, S. and Mohana Kumar, S. (2020). Paddy Plant Disease Classification and Prediction Using Convolutional Neural Network. *2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*. doi:10.1109/rteict49044.2020.9315634.

Sai, G.U., Tejasri, N., Kumar, A. and Rajalakshmi, P. (2022). Deep Learning Based Overcomplete Representations for Paddy Rice Crop and Weed Segmentation. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. doi:10.1109/igarss46834.2022.9884107.

Schumacher, M. (2002). *The use of SSADM (Structured Systems Analysis and Design Methodology) as a standard methodology on Information Systems Projects*. [online] *www.grin.com*. Available at: https://www.grin.com/document/106034. [Accessed 29 Oct. 2022].

Suresha, M., Shreekanth, K.N. and Thirumalesh, B.V. (2017). Recognition of diseases in paddy leaves using knn classifier. *2017 2nd International Conference for Convergence in Technology (I2CT)*. doi:10.1109/i2ct.2017.8226213.

Swathika, R., Srinidhi., S., Radha, N. and Sowmya., K. (2021). Disease Identification in paddy leaves using CNN based Deep Learning. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. doi:10.1109/icicv50876.2021.9388557.

Talreja, R., Jawrani, V., Watwani, B., Sengupta, S., Rohera, P. and Raghuwanshi, K.S. (2022). AgriCare: An Android Application for Detection of Paddy Diseases. *2022 3rd International Conference for Emerging Technology (INCET)*. doi:10.1109/incet54531.2022.9825038.

Tendang, S. and Chamnongthai, K. (2021). Rice-Disease Severity Level Estimation Using Deep Convolutional Neural Network. *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*. doi:10.1109/itc-cscc52171.2021.9501485.

The World Bank (n.d.). *Agricultural land (% of land area) - Sri Lanka | Data*. [online] data.worldbank.org. Available at: https://data.worldbank.org/indicator/AG.LND.AGRI.ZS?locations=LK

Umamaheswari, S., Arjun, R. and Meganathan, D. (2018). Weed Detection in Farm Crops using Parallel Image Processing. *2018 Conference on Information and Communication Technology (CICT)*. [online] doi:10.1109/INFOCOMTECH.2018.8722369.

Weizenbaum, J. (1983). ELIZA --- a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 26(1), pp.23–28. doi:10.1145/357980.357991.

Yan, X., Deng, X. and Jin, J. (2020). Classification of weed species in the paddy field with DCNN-Learned features. *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*. doi:10.1109/itoec49072.2020.9141894.

Appendix

Use Case Description 02 – Farmer, Field Officer

Use case name	To determine the weed.
Actors	Farmer, Field officer
Purpose	To determine the name of the weed in the cultivation
Overviews	Farmer needs to determine the disease of the Farmer inputs an image to the app Application generates the result
Pre- conditions:	Farmer and Field officer should register to the app. Farmer and Field officer should have a weed
Post- conditions:	Determined the weed and provide the recommendations

Typical course of events:

Actor action	System response
(01) Farmer logins to the application (03) Farmer inputs an image to the application (05) Farmer inputs an image of incorrect format	(02) Application allows to login (04) Application generates the result (05) Application displays an error message.
Alternative Case	No weeds in the cultivation
Includes	Input image of a weed Validate image Recommend solution

Use Case Description 03 - Farmer

Use case name	Use of chatbot
Actors	Farmer
Purpose	To resolve the problems related to paddy agriculture and contact the field officers
Overviews	Farmer needs to clarify problems related to agriculture practices Farmer enters the query to the chatbot Chatbot provides the results
Pre- conditions:	Farmer should have a problem
Post- conditions:	Solved the problem and contacts the relevant authority

Typical course of events:

Actor action	System response
(01) Farmer enters the query to the chatbot (02) Farmer enters an incorrect query	(03) Chatbot provide the results (04) System displays an error message
Alternative Case:	Farmer has no problems Chatbot doesn’t contain relevant contact information
Include:	Resolve agricultural questions Find contact info of officials

Use Case Description 04 – Farmer, Field Officer

Use case name	Registration and Login
Actors	Farmer, Field officer
Purpose	To resolve the problems related to paddy agriculture and contact the field officers
Overviews	Farmer and Field officer needs to clarify problems related to agriculture practices, diseases and weeds. Farmer enters the relevant details for registration and login process. Farmer is allowed to access the functions of the application.
Pre- conditions:	Farmer should have a problem related to paddy agriculture.
Post- conditions:	Solved the problem through the functions of the application.

Typical course of events:

Actor action	System response
(01) Farmer and Field officer enters the personal information for registration process (02) Farmer logins to the system	(03) Application check for validity of information and registers the details (04) System allows to login for registered members
Alternative Case	Farmer has no problems Chatbot doesn’t contain relevant contact information

<u>Use Case Description 05 - Field Officer</u>

Use case name	Provide details
Actors	Field officer
Purpose	Field officer details are required for creating the chatbot
Overviews	System developer needs to create the chatbot System developer gets the data through the field officers System developer creates the chatbot
Pre- conditions:	Field officer should volunteer for providing information
Post- conditions:	Created the chatbot

Typical course of events:

Actor action	System response
(01) Farmer and Field officer enters the personal information for registration process (02) Farmer logins to the system	(03) Application check for validity of information and registers the details (04) System allows to login for registered members
Alternative Case	Farmer has no problems Chatbot doesn’t contain relevant contact information

Use Case Description 06 – Administrator

Use case name	Manage and update the system
Actors	Administrator
Purpose	For the maintenance and updates relevant to the system
Overviews	Administrator needs update the system Administrator checks for the bugs Administrator fix the bugs and update the system
Pre- conditions:	The system should have a technical issue
Post- conditions:	Fixed the bugs and updated the system

Typical course of events

Actor action	System response
(01) Check for bugs and resolve the issues. (03) Check for login issues (05) Update the system	(02) Works appropriately (04) Allows the login request (06) Works appropriately
Alternative Case	No bugs in the system.

<u>Extend Use case – Email the Solutions</u>

Use case name	Email the solutions
Actors	System
Purpose	To send a report of solutions according to the request of user
Overviews	User needs a report of solutions User request for a report System emails the report
Pre- conditions:	System should recommend solutions
Post- conditions:	Report is emailed to the user
Alternative Case	No solutions recommended