DISEASE IDENTIFICATION AND MANAGEMENT SYSTEM FOR GHERKIN PLANTS

TMP-R24-010

Project Proposal Report

Perera H.M.A.G - IT21022388

BSc (Hons) in Information Technology

Specializing Information Technology

Department of Computer Science, Information Technology, Sri Lanka Institute of Information Technology Sri Lanka

February 2024

DISEASE IDENTIFICATION AND MANAGEMENT SYSTEM FOR GHERKIN PLANTS

TMP-R24-010

Project Proposal Report

Perera H.M.A.G. - IT21022388

BSc (Hons) in Information Technology Specializing in Information Technology

Department of Computer Science, Information Technology, Sri Lanka Institute of Information Technology Sri Lanka

February 2024

DECLARATION

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

Signature: …….. Date:…….

Signature of the Supervisor: …… Date:…….

Signature of the Co-Supervisor: …… Date:…….

ABSTRACT

Gherkins are a popular vegetable worldwide, yet they face several threats from diseases. Accurately identifying diseases and implementing effective management strategies are crucial for maintaining optimal production levels. Identifying diseases affecting gherkin plants is a critical challenge for farmers, impacting both crop quality and yield. Achieving high-quality gherkins is crucial, especially for export. Misidentification of diseases using traditional methods can significantly reduce both quality and quantity of harvest. Manual identification often suffers from limitations of time consuming, rely on visual inspections, experiences, and expertise requirement. However, traditional methods often lack precision and speed, resulting in lower yields and higher costs.

Gherkin cultivation faces numerous challenges related to disease management, leading to reduced yields and increased costs. This system offers a robust and reliable tool for gherkin disease identification and management. Its ability to identify diseases and categorize disease types, estimate diseases impact levels, forecast disease stages, and suggest appropriate solutions makes it an asset for farmers seeking to maximize their yields while minimizing costs quickly and accurately.

The aim of the research is to develop a system that utilizes CNNs to analyze images of infected gherkin plants, distinguishing between diseases caused by fungi, bacteria, or environmental factors. It then estimates the impact level of identified diseases, categorizing them as critical, severe, or minimal. Additionally, the system forecasts the current stage of diseases using RNN to help farmers intervene at the right time. Eventually the system provides customized solutions for disease management, drawing upon a vast repository of expert advice accessed through a React Native mobile application. The system aims to optimize Gherkin cultivation practices, mitigate crop losses, prevent, and protect gherkin plants from diseases and enhance overall agricultural productivity. As such, we strongly recommend its adoption in gherkin cultivation globally.

Keywords: Gherkin Cultivation, Disease Management, Image Analysis, Impact Assessment, Disease Stages, Solutions Provisioning.

Table of Contents

LIST OF ABBREVIATIONS

Abbreviation	Description
ML	Deep Packet Inspection
CNN	Threat Intel
RNN	Indicators Of Compromised

1. INTRODUCTION

The significance of the current state of gherkin production in Sri Lanka is that the country exports 6984 tons of gherkins worth Rs.189 million [1]. Gherkins are grown in diverse agro-ecological regions according to the dry zones of low-county wetness to mid county wetness. They are popular in Sri Lanka due to the unique ability to grow in a wide range of soil types and weather conditions [1]. However, diseases like leaf miner, aphids, and thrips pose a major challenge to gherkin production, leading to reduced yields and increased costs [1].

The distribution of cucumber production has broadened within the agro-ecological zones of the country, and the private companies offer the required technology to growers [1]. However, there were problems with gherkin production, but the industry was very well-known in several agro-ecological regions, such as the dry zones of the low country and the intermediate zones of the mid country [1]. Gherkin farming is a key feature of the Sri Lanka agribusiness, contributing a lot to the country's economy. Yet, the diseases entirely happen to be the disturbing factor in the gherkin production that subsequently cause a reduction in yield and extra costs. To deal with these obstacles, researchers have built various tools and approaches. Traditional practice for the diagnosis of gherkin plant diseases involved simple techniques such as visual examination. While they do offer a risk for disease management, they are also labor-intensive, time-consuming, and prone to errors limiting their effectiveness in disease management.

To tackle these challenges, researchers have employed machine learning (ML) techniques and convolutional neural networks (CNNs). These technologies have shown promise in accurately identifying diseases in gherkin plants, predicting their impact levels, forecasting their growth stages, and providing suitable solutions.
At present, researchers are studying to use Transfer Learning which increases the efficiency of CNN models without labelled data to be too much. In addition, they are exploring ROI-based methods, which help in the fine-tuning of the predictability of diseases.

Many projects have been developed as a solution to these issues with the aim to make the diagnosis of gherkin diseases. Though there were models that managed the production line, those were not usually accurate and fast, they had less yield and thus required more cost. The basis of our method will strive to expand upon these previous studies by applying the advanced ML strategies and CNNs to get into more detail detection, implication degree prediction, and stage forecasting. With our method we will use ML algorithms and CNNs to accurately recognize diseases gherkin plants, gauge their threat levels, forecast their growth stages, and recommend proper measures. This will provide the farmers with decision-making tools to act in a timely and efficient manner, allowing them to prevent gherkin plant diseases, thus leading to better production results.

This study will use modern ML tools to develop more sophisticated identification methods (disease) which will in turn increase accuracy of disease-impact prediction and stage prediction, too. This will speed up production action time and help them timely and efficiently solve problems on gherkin plants and optimize production.

1.1. BACKGROUND & LITERATURE SURVEY

Gherkin (Cucumis anguria), a cucumber-like vegetable, holds significant economic importance in agriculture worldwide due to its versatile culinary uses and high market demand. Sri Lanka, with its favorable climatic conditions, cultivates gherkins extensively, contributing to the country's agricultural exports. Gherkin cultivation is a significant agricultural activity in regions like Sri Lanka and Karnataka. In Sri Lanka, the gherkin industry has seen substantial growth, with exports increasing over the years [1]. The cultivation of gherkins is widespread across different agro-ecological zones, showcasing the adaptability of this crop to various soil types and climates [2]. Varietal screening plays a crucial role in determining the most suitable gherkin varieties for different environmental conditions, aiming to enhance yield potential under local settings [1].

Disease identification is vital in gherkin cultivation to ensure optimal plant health and productivity. However, there is limited information available specifically on disease identification methods used by Sri Lankan farmers. Understanding common diseases affecting gherkin plants and implementing effective disease management strategies are essential for sustainable cultivation practices. Despite its economic significance, gherkin cultivation is vulnerable to various diseases that can significantly impact plant health and yield. Common diseases affecting gherkin plants include fungal infections like powdery mildew and downy mildew, bacterial diseases such as bacterial wilt, and viral infections like cucumber mosaic virus. Effective disease management strategies are crucial for minimizing crop losses and ensuring sustainable gherkin production.
Traditionally, Sri Lankan farmers rely on manual methods for disease identification in gherkin plants. This process involves visual inspection of plants for symptoms such as leaf discoloration, lesions, and abnormal growth patterns. Farmers use their experience and knowledge to diagnose diseases based on observed symptoms, often resorting to trial-and-error approaches, and consulting agricultural extension officers for guidance.

Research on gherkin cultivation in Karnataka highlights the importance of contract farming and SWOT analysis in enhancing production efficiency and market access[2]. The study emphasizes the benefits of contract farming for small and medium farmers, offering reduced risk, guaranteed returns, and technical support. This research provides valuable insights into the strengths, weaknesses, opportunities, and threats associated with gherkin cultivation in Karnataka.
Several research studies have explored automated methods for disease identification in crops, including gherkins, using technologies such as machine learning and image processing. However, existing research primarily focuses on larger-scale crops like rice, wheat, and maize, with limited emphasis on gherkin cultivation. While some studies propose automated solutions using drones and deep learning techniques, these approaches may not be readily accessible or affordable for small-scale gherkin farmers.

This research focusing on disease identification methods in Sri Lankan gherkin cultivation fills a crucial gap in existing literature. By exploring manual disease identification practices used by farmers and comparing them with technological advancements, your study can contribute significantly to improving disease management strategies. This research is essential for enhancing crop health, increasing yield, and ensuring sustainable gherkin production in Sri Lanka. Incorporating advanced technologies like image processing, and machine learning tools can revolutionize disease identification processes in gherkin plants[3]. These technologies offer precise and efficient methods for early disease detection, enabling timely interventions to prevent crop losses. Implementing such tools can enhance the accuracy and speed of disease identification compared to manual methods.

The proposed disease identification and management system for gherkin cultivation fill a crucial gap in existing research by providing a practical and accessible solution tailored to the needs of Sri Lankan farmers. By leveraging machine learning algorithms, such as Convolutional Neural Networks (CNNs), for image analysis,

the system offers accurate and timely disease diagnosis, enabling farmers to take proactive measures to mitigate crop losses effectively.

To overcome the limitations of manual disease identification and management, there is a growing interest in leveraging technology, particularly machine learning (ML) and convolutional neural networks (CNNs). These advanced techniques offer the potential to automate disease detection, improve accuracy, and provide timely insights for farmers.

By integrating ML and CNNs into disease identification systems, researchers aim to enhance the efficiency and effectiveness of disease management practices in gherkin cultivation. These systems utilize image analysis algorithms to analyze plant images, identify disease symptoms, and classify diseases based on their causal agents. However, existing literature on disease identification in gherkin plants using ML and CNNs is limited. While some studies have explored similar approaches in other crops, there is a lack of research specifically tailored to the unique characteristics and challenges of gherkin cultivation. Thus, there is a need for further research to develop robust disease identification and management systems tailored to the specific needs of gherkin farmers.

Considering these considerations, this research proposal aims to bridge the gap between traditional manual methods and modern technological solutions by developing a comprehensive disease identification and management system for gherkin cultivation. By leveraging ML, CNNs, and mobile application technology, the proposed system seeks to provide farmers with accurate disease diagnosis, impact assessment, disease stage forecasting, and actionable solutions, ultimately improving crop health and productivity.

1.2. Research Gap

A research gap in the Gherkin disease identification and management system is observed through the application of different technologies which identify diseases accurately and measure their levels of impact and progress stages. Despite the concentration on the disease's recognition by conventional diagnostic ways like visual examination, these techniques are laborious, time-taking and erroneous. What makes it a challenge as well, is that there is a gap in the prediction of the force of diseases on gherkin plants and the determination of their growth stages, which can severely affect the health of the crop and estimate its productivity.

Even though there are studies by researchers that investigated automatic disease differentiation in crop using machine learning methods with CNNs, there is quite a gap in their use in the case of gherkin cultivation. The prevailing research generally emphasizes greater proportioned grains such as rice, wheat, and maize but, they are slightly overlooking okra. The consequence of this background information therefore is that machine learning and CNNs are not yet adapted to the distinctive features and plant pattern-sides of gherkin plant.

Lack of Impact Level Prediction: Lack of Impact Level Prediction:
The common disease detection research in plants used nowadays is based on the ability to discover the presence of diseases but it has no potential for predicting the impact level on the plant health. Introducing this component, our novel technology helps to estimate interspecies disease transmitting pathways and disease consequences in gherkin plants, ranging from critical and severe to mild and moderate. This implies a major research gap in the existing literature as the disease impact level forecasting achievements are essential for farmers taking into consideration the management priorities due to its accuracy.

Absence of Disease Stage Forecasting: Absence of Disease Stage Forecasting:
Another research gap in the already exiting literature is the absence of disease forecasting for the main stages of gherkin plants. Contrarily, some research could be found that identify the diseases, but do not give information about current staging of the disease development. In this part, we tackle the deficiency by adding a novel function to forecast the stages of a disease at varying heights. This could help farmers find out when they need to intervene and select proper practices earlier.

To achieve the first novelty of forecasting disease susceptibility in gherkin plants, you need to employ a CNN multi-class classifier. Teach the model this dataset which includes pictures of different levels of damage shown by the disease. This will be a way to allow the model to identify between minor, moderate, and critical influences to make viable decisions thereon.

The second novelty can be applied using the ROI approach which can select the region with highest intensity. Make the model emphasize the areas on the image where the lesion is and not the whole image. Thanks to this the model will be able to differentiate between those images newly encountered as impacted or healthy with reasonable precision.

The research area has been mostly concerned with disease identification using classical approaches such as visual confidentiality which is a work of art, takes time, and results in human errors. Consequently, these old methods cannot give the wanted result because they lack precision and speed that allow them to work lower yields at the high rate. The goal of your this study is to enhance the earlier researches through advanced ML techniques of identifying the diseases with remarkable accuracy, while the CNNs could predict disease impact and level with more precision. To cope with this information gap only the recommended part about research will have recourse to machine learning systems using CNNs in disease identification and management in gherkin farming.

Through this research proposal, we aim to

- Develop an automated disease detection and management device, which will match gherkin production.
- Offer, realistically, the extent and severity of the disease effects, thus allowing farmers to step up the management of diseases.
- Institutionalize research gaps that are not answered by the current literature to support agricultural progress in gherkin cultivation.

Research aiming to bridge the gap in current literature will do that by using the ML algorithms and CNNs to improve the precision in disease identification, impact intensity level prediction, and forecasting stage of the disease. This will empower farmers to take proper decisions soon enough and guide enable them in rightly managing diseases on the plants of gherkin through the good quality of crops productions.

1.3. Research Problem

The prime research problem relates to the building of an easy system which would have the ability to precisely identify and segment the diseases of the gherkin plants by using advanced image processing techniques as the main technique. It implies surmounting problems, namely: symptom variability of disease, impact of environment upon plant health, and creation well-developed algorithms to process multiple image data in an effective way.

A very important part of my search is the development of methodologies for forecasting of disease progression stages and gherkin plant's efficacy level. This is achieved using machine learning to mine the historical data and accurately predict disease across different stages. Also, there ought to be a categorization of the disease impact severity, ranging from critical, to severe and minimal to aid in timely and efficient management strategies implementation.

As a critical research step, optimization of the created system is imperative to guarantee exactness, effectiveness, and attain peak performance. It includes the subsequent adjustment of machine learning models, optimization of algorithms for speed and scalability and the implementation of user-friendly interfaces for a simple interaction with agricultural market players. Besides the resource constraints problem where computational power and data storage are considered, the other fundamental research problem is the ability to achieve maximum performance while operating the system optimally.

The research problem also relies on you engaging the growers, the agricultural experts, and other stakeholders so by doing so you individually could ensure the system's effectiveness and its usability in real-world settings. (This implies that it involves the conduct of the user studies to collect the insights on the user preferences and needs, as well as the refinement of the system based on the stakeholder feedback.)

To this end, it involves overcoming scaling and adaptability issues, which guarantee that the engineered system can be implemented in different climatic zones, where cultivation of an eggplant takes place. It encompasses incorporating issues like geographical variability, climatic conditions, and farming practices in the design of a system capable of being dynamic and responsive to meet the needs of the growers of different regions.

To sum up, the problem is designing a comprehensive disease identification and management system for gherkin cultivation that tackles the matter of precision, prediction, optimalization, stakeholders integration, and scalable approaches. In order to address these research areas, the proposed research will seek to develop innovative technologies for gherkin production and support a sustainable and productive system of production.

2. OBJECTIVES

2.1. Main Objectives

The main objective of this component is to develop a mobile application to identify diseases of gherkin plant leaves and provide solutions. Gherkin farmers can easily identify the diseases and prevent them. This full component is completely relying on machine learning and deep learning architectures like CNNs to provide a more effective disease fighting system in Gherkin growing operation. This 'automated disease identification system' will be able to precisely diagnose diseases in gherkin that are present in images, determine the type of impact on plant health, forecast the current growth stage of disease, and suggest appropriate solutions. Through sophisticated applications of technologies and techniques, their aesthetic goal is that it increases gerbera yield quality and quantity cuts disease losses, all in a bid to grow agricultural productivity using gherkin production.

2.2. Specific Objectives

There are three specific objectives that must be reached to achieve the overall objective described above.

Disease Identification:
Collect a set that involves a variety of gherkin plant images as well as both one of a kind healthy and those affected by different diseases. Apply preprocessing of the image data by rescaling, normalization of the pixel-values and increase the size of the dataset for better model robustness. Exploit Convolutional Neural Networks (CNNs), that is, networks in which the feature extraction is adaptively performed on the succession of convolutional layers, which is perfectly suitable for the problem of image classification. Valid the trained model's performance metrics such as accuracy or precision through use of a specific validation dataset, so the disease diagnosis result can be evaluated reliably.

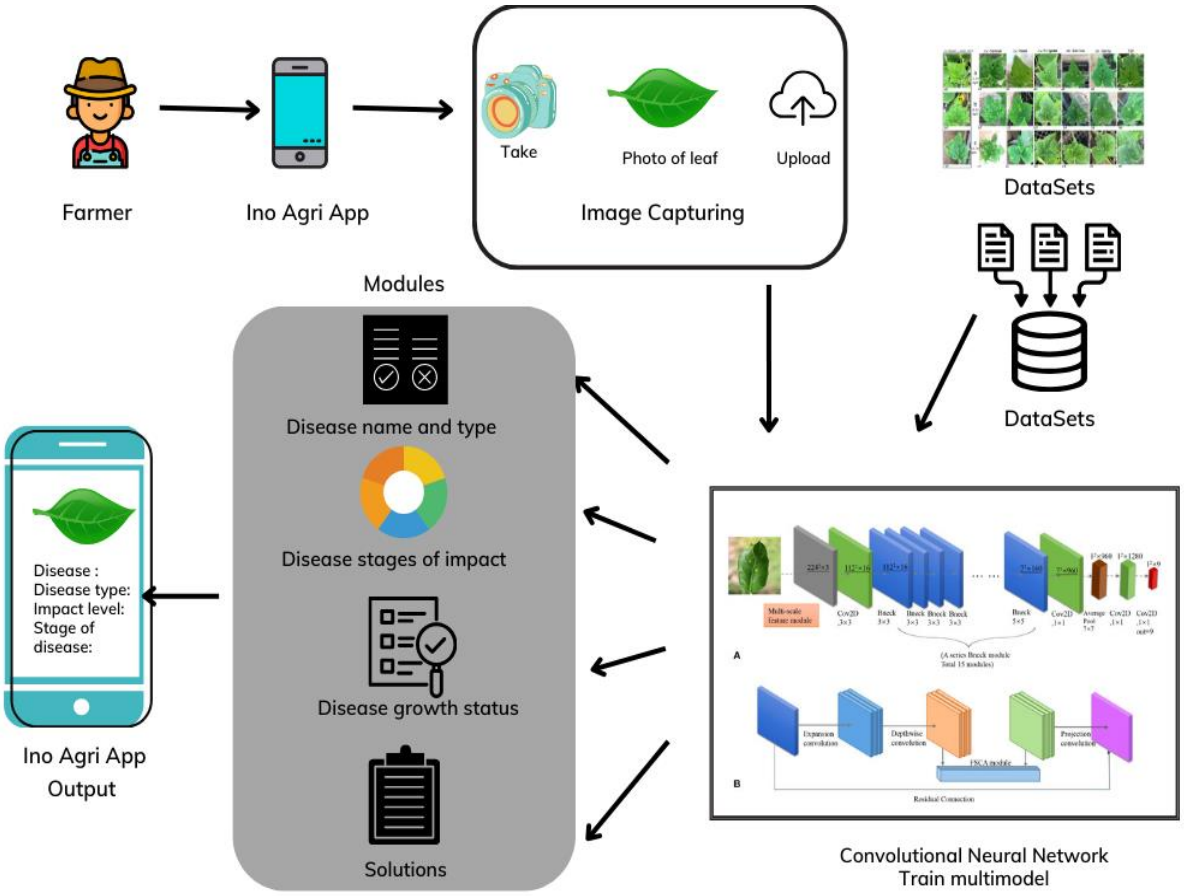
Impact Level Prediction:
Create a dataset that encompasses a variety of diseases, their showed effects represented by critical, severe, or minimum impact levels. On the disease with the use of different types of leaves infected train models and labeling. First, choose an appropriate machine learning algorithm for multi-class classification (e.g. "CNN") Acquire the selected classifier. Output the dataset in the suitable way to the algorithm and manage class imbalance, if necessary. Assess the trained classifier’s competence with the help of matrices such as accuracy and precision. These metrics will help you understand how well the classifier can identify disease impact levels correctly.

Disease Stage Forecasting:
Collect a one-dimensional dataset encompassing the historical data on disease progress in the gherkin plants since the development of the plant influenced different stages, with different observations. Preprocessing of the time-series data entails the establishment of the best way to handle missing values and eliminate outliers accurately. Select LSTM (Long Short-Term Memory) networks, a group of RNN (Recurrent Neural Networks), as they are very efficient at dealing with the dependencies in the data that is sequentially arriving. This, in turn, makes them to be suitable for tasks dealing with time-series forecasting. Train an LSTM model on the preprocessed time series dataset. Vary model architecture and hyperparameters to arrive at the model that features high accurate disease stage prediction.

Solution Provisioning:
Adapt external APIs or databases with expert consultation and solutions for disease control as being open source in gherkin. Reach for the API integration by featuring a function within the React Native mobile application to view and even take the solutions available with the application.

3. METHODOLOGY

3.1. System Diagram



4. PROJECT REQUIREMENTS

4.1. Functional Requirements and Non-Functional Requirements

Functional Requirements	Non-Functional Requirements
System should allow user to upload real time images of gherkin plant leaves affected by diseases.	User-friendliness: The system should offer cross-platform application with a visually appealing, responsive, and well-organized interface.
Utilizing CNN, the system should accurately identify diseases and disease types from the uploaded images.	Reliability: Throughout the procedure, the system shouldn't malfunction or become stuck. When utilizing the program, consumers ought to have confidence and security. All private data needs to be safeguarded.

The system should predict the impact level of the disease on the gherkin plant.	Availability - Users should be able to access the application. It ought to be usable whenever necessary.
By analyzing datasets, the system should forecast the current growing stage of the diseases.	Performance: The system should function well by giving users quick and accurate results.
Providing solutions to users via a React Native mobile app and API s for additional information.	

4.2. Expected Test Cases

1. Test the accuracy of disease identification by comparing predicted results against actual diseases present in the images.

2. Evaluate the speed of disease identification and impact level prediction by measuring the time taken to generate predictions. To evaluate the user interface and make sure the instructions are clear and easy to use, conduct user tests.

3. Quantify the speed of disease identification and risk level determination by measuring the time that is required to generate the forecast.

4. Undergo testing with users to evaluate the interface of the course and to make sure that the instructions are clearly explained.

5. BUDGET AND BUDGET JUSTIFICATION

Expenses	
Requirement	Cost (Rs)
Hardware Infrastructure – (high-performance CPUs and GPUs)	50,000.00
Software Licenses	3,500.00
Travel and Training	15,000.00
Equipment and Supplies	10,000.00
Total	78,500.00

6. REFERENCES

[1] J. Zhang, Y. Rao, C. Man, Z. Jiang, and S. Li, "Identification of cucumber leaf diseases using deep learning and small sample size for agricultural Internet of Things," IEEE Internet Things J., vol. 8, no. 8, pp. 6545-6554, Aug. 2021, doi: 10.1109/JIOT.2021.3071746.

[2] O. Metin, "Automatic Detection and Classification of Mildew Disease in Cucumber Using Deep Learning Algorithms," Journal Title, volume(issue), page range, year.

[3] M. J. Mia, S. K. Maria, S. S. Taki, and A. A. Biswas, "Cucumber disease recognition using machine learning and transfer learning," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 6, pp. 3432-3443, Dec. 2021, doi: 10.11591/eei.v10i6.3096.

[4] J. M. Kable, "Key to Common Problems of Cucumbers," University of Maryland Extension, Feb. 18, 2024. [Online]. Available: https://extension.umd.edu/resource/key-common-problems-cucumbers/. [Accessed: Feb. 20, 2024].

7. APPENDICES