Synergetic Innovation in Gherkin Cultivation

Jinasena H.D.S.S - IT21042560

Dayarathna T.W.K.D.A - IT20254216

Perera H.M.A.G - IT21022388

Nissanka K G T R K - IT21045530

BSc. (Hons) in Information Technology

Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

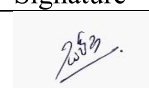
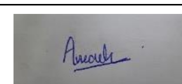
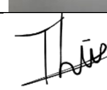
Sri Lanka

DECLARATION

We hereby declare that the work presented in this thesis, titled “Synergetic Innovation in Gherkin Cultivation,” is the original work of our research team, conducted under the guidance of Mr. Dharshana Kasthurirathna and Ms. Poojani Gunathilake at the Sri Lanka Institute of Information Technology (SLIIT). This research has not been submitted, either in part or in full, for any other degree or qualification at any other institution.

All information and materials derived from external sources have been properly cited and referenced. We confirm that we have adhered to the academic integrity policies and guidelines of SLIIT throughout the research and thesis writing process.

Furthermore, we affirm that the thesis represents our independent research and analysis. All contributions from other researchers and sources have been duly acknowledged.

Student Name	Student ID	Signature
Jinasena H.D.S.S	IT21042560	
Dayarathna T.W.K.D.A.	IT20254216	
Perera H.M.A.G	IT21022388	
Nissanka K G T R K	IT21045530	

Date: 02nd August 2024

The Supervisor/s should clarify the dissertation with the following declaration.

The above candidate has carried out research for the bachelor’s degree dissertation under my supervision.

Signature Of the Supervisor

…...…………………………

Kasthurirathna D

Signature of the Co-Supervisor

……..………………………..

Gunathilake M.P

ABSTRACT

This research provides a detailed machine learning system which seeks to solve one of the most daunting problems in the gherkin farming process including the detection of diseases, forecasting of yield, pest management and cost estimation of production. The framework incorporates novel object detection algorithms like YOLOv8 and utilization of CCD CNN with EfficientNet for detecting and differentiating gherkin leaf diseases. In addition, the system captures environmental conditions like ambient temperature and humidity through the Gemini model then delivers context appropriate disease management advice. For yield prediction, a Random Forest algorithm is applied with input parameters including the pH of soil, nutrient contents, rainfall, and use of fertilizer to predict crop yields with relatively high precision. Pest control is also responded by a mobile application called Inno Agri that uses machine learning algorithms including YOLOv8, Xception together with augmented reality to offer a feasible solution to farmers. Moreover, this study develops a machine learning cost forecasting model for planting costs that uses regression methods and other planting cost indicators such as land, labour, and inputs such as fertilizer and insecticides, using past data. This proposed integrated framework is aimed at improving productivity, efficiency and profitability in Gherkin farming ventures by providing farmers, traders and other relevant actors with a relevant, validated and scalable system. According to the findings of this research, it is possible to enhance decision-making, resources utilization and sustainability of gherkin agricultural production through the integration of machine learning and advanced analytics.

Keywords: ***Gherkin Cultivation*, *Machine Learning, Object Detection, Disease Detection, Disease Classification, Gherkin Cultivation, CNN Architectures, YOLOV8, Pest Identification, Cost prediction, Harvest Prediction, Random Forest***

ACKNOWLEDGEMENT

We would like to extend our heartfelt gratitude to everyone who has supported and contributed to the successful completion of this research.

First and foremost, we express our sincere gratitude to our supervisor, Mr. Dharshana Kasthurirathna, and our co-supervisor, Ms. Poojani Gunathilake, for their invaluable guidance, continuous support, and crucial assistance throughout the course of this project. Their expertise and direction were instrumental in shaping our work and ensuring its successful outcome.

We are also deeply grateful to our group members for their constructive feedback and insightful suggestions, which significantly enhanced the quality of this thesis.

Special thanks go to the Research Project Team of the Sri Lanka Institute of Information Technology (SLIIT) for organizing various informative sessions that provided us with guidance during the project, and to SLIIT for offering the necessary resources and facilities to conduct this research.

In addition, we extend our appreciation to our colleagues and friends for their encouragement, support, and productive discussions, all of which were invaluable throughout the research process.

Lastly, we would like to express our deepest gratitude to our families for their unwavering support and understanding. Their patience, encouragement, and belief in our abilities have been a constant source of motivation throughout our academic journey.

Thank you all for your support and confidence in us.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

CNN	Convolutional Neural Network
YOLOV8	You Only Look Once Version 8
VGG19	Visual Geometry Group 19
AI	Artificial Intelligence
Mask R-CNN	Region-based Convolutional Neural Network
ML	Machine Learning
API	Application Programming Interface
HTTP	Hypertext Transfer Protocol
MOP	Muriate of Potash
TSP	Triple superphosphate
KNN	K-Nearest Neighbor
AR	Augmented Reality
ANN	Artificial Neural Network

SVM	Support Vector Machines
GDP	Gross Domestic Product

1 INTRODUCTION

1.1 Background & Literature

Gherkin farming has great importance in the world agriculture market because the use of this versatile product is a common practice in all countries. Common gherkins are crispy while gherkins are soft but in size, shape and even at times taste they are same though gherkins are associated with pickles. The global consumption of gherkins especially in the form of pickles fascinate and the major consumers are in European, Asian and Americas.

Gherkin farming is a vital component of global agriculture, providing a primary income for farmers worldwide while also supporting the production of pickled goods. Gherkins, often referred to as pickling cucumbers, are cultivated extensively in regions with suitable climates, including parts of Europe, Asia, and Africa. Among these, India is one of the leading producers and exporters of gherkins, with significant production occurring in the southern states such as Karnataka, Tamil Nadu, and Andhra Pradesh. This crop has become an important export commodity, particularly for markets in the United States, Europe, and Japan, where pickles are in high demand. The success of these export markets has played a crucial role in supporting the livelihoods of thousands of smallholder farmers and in bolstering the economic standing of rural communities. Gherkin farming involves a complex production process that includes seed selection, land preparation, planting, irrigation, pest control, and harvesting. The crop requires specific growing conditions, including warm temperatures, well-drained soils, and an adequate water supply, making climate and soil quality crucial factors in successful cultivation. Gherkin farming is labor intensive, with manual planting and harvesting being common practices. Moreover, the production cycle is relatively short, with gherkins ready for harvest within 30-40 days of planting, allowing for multiple crop cycles within a year. This characteristic enables farmers to produce several harvests annually, thus increasing their income potential. However, the short production cycle also introduces a degree of unpredictability, as minor disruptions in input supply or labor availability can significantly impact yield and profitability

Globalization in today's world is producing heightened demands for agricultural produce to feed the growing population calling for efficiency in crop production. Cucumber is a common vegetable crop, and it is prone to several diseases which have a compromising impact on yield and quality. Due to this the early diagnosis and right identification of these diseases are very important so that early and correct management steps can be taken. AI and machine learning borrowed from modern technology aid in establishing other systems that would help the farmers to identify and control diseases. But one of the main issues associated with farming gherkin is that they are susceptible to various diseases that pose a big threat to the production of this fruit crop. Europe, Asia and Americas are the largest consumers of gherkins, gherkins are the most popular in the three regions. Today leading producers and exporters in this category include India, Sri Lanka, Mexico and Turkey among others. These countries rise to meet the growing demand for pickled gherkins, particularly in Europe and North America where the demand seems to rise with time for the tangy, crunchy product. The market for pickles across the world is expected to grow because of special food and changing consumption trends particularly towards convenience food.

Sri Lanka have clearly marketed themselves as large-scale exporters of gherkin pickles, getting more value added due to the special taste and smell which is due to the tropical climate conditions [1]. Light and a favorable environment for the growth of gherkin is always available in Sri Lanka because of the favorable geographical setting of sunlight throughout the year helps to produce better quality gherkin products, and hence, these are in great demand in international markets. Currently Sri Lanka has carved out a niche market for herself and this is in the international market with gherkin pickles. In the recent past, Sri Lanka's export market of gherkin pickles has greatly expanded, and the country has well-advertised quality products. Nevertheless, due to the captive market, the chance at profitability, and as seen with current consumers, there exists a demand for gherkins in Sri Lanka, this crop has not exploded in popularity amongst farmers. At the present time only, several firms engage in production and export of gherkins and therefore the development of gherkin as an agricultural produce is still restricted in the mentioned country [2].

Since livestock raising becomes highly profitable and enhances food security and application of sustainable technologies in farming, to the same extent, growing gherkins also becomes profitable and beneficial. It would only make great sense because it is easy to expand the area under cultivation, and it is rainfall-friendly besides hardiness to a variety of soil types and climatic conditions that smallholder farmers can grow the crop. Moreover, it is more viable to grow gherkins in that they only need several productions in a year, hence, the more money the farmers make.

The demand for healthy food is increasing worldwide, the changing customer preference towards pickled food products is also augmenting the gherkin market. Consequently, through its high vitamin and mineral content with attraction to the consumer group that considers it as virtually calorie-free snack product, gherkins won the consumers' favor. That way, this growth offers farmers in Sri Lanka a great opportunity to increase their crop holdings, and more importantly, their wealth. The global demand for gherkins has led to the development of contract farming models, where companies provide farmers with seeds, technical support, and a guaranteed buyback of the produce. This model has helped small and marginal farmers integrate into global supply chains, ensuring a stable income and reducing market risks. Additionally, the contract farming arrangement often includes training and support for farmers, enabling them to adopt best practices in gherkin cultivation. However, the cost of gherkin production can vary significantly due to factors such as labor availability, input costs (seeds, fertilizers, and pesticides), and weather conditions. These variations in cost, coupled with fluctuating market prices for gherkins, underscore the need for reliable forecasting tools to help farmers and stakeholders navigate the uncertainties of gherkin production.

1.1.1 Gherkin cultivation of Hayleys company and challengers

Today Sri Lanka has succeeded in becoming one of the players in the international gherkin market, owing to which it is possible to talk about Sri Lanka as one of the exporters of quality gherkins, courage by the availability of optimum climate required for the cultivation of gherkin and good soil texture exported of gherkin has been on the increase in the recent decades in addition to the increasing global demand for pickles. Sri Lankan Gherkin pickles have a great demand in the international industry because of the excellent natural taste and aroma of the products. The naturally placed geographic location, year-round sunlight and most favorable atmospheric conditions are causing that. Therefore, Sri Lanka placed a significant exporting brand in the foreign market. Major proportion of gherkin is produced in the central and the Northern provinces of Sri Lanka and mostly produced by the small and middle farmers.

However, despite the potential, gherkin cultivation is not widely practiced among Sri Lankan farmers. The main obstacle is farmers' lack of resources and knowledge, which restricts the amount of land that can be farmed. Technical expertise among local farmers regarding modern agricultural practices and gherkin-specific cultivation techniques. Many farmers are unaware of the high-yielding hybrid varieties available, which can significantly enhance productivity. There are currently just two or three enterprises in Sri Lanka that are actively engaged in the cultivation and export of gherkins. This low level of participation emphasizes the need for greater support and information sharing to promote Gherkin Farming's wider adoption.

Hayleys Agriculture company is one of the main Gherkin product providers of Sri Lanka. In the 1950s, due to a shortage of food, the Hayleys Agriculture department was created to cover the country’s agricultural demands. The Hayleys group later made an entry into the export trade in1988 by supplying semi-processed gherkins pickled in brine to pickle production factories in Australia, Europe, Japan, and New Zealand. Our research team is engaged with Hayles Agriculture to provide software solutions for their gherkin cultivation challengers. The local agricultural community is facing a lot of troubles when cultivating. They believe in supporting local farmers and creating a secure environment for them to grow and harvest their crops. Along with that, we researched the difficulties they faced and created four components related to that. Farmers are struggling with the diseases. Because every year farmers lose a high number of crops. Sometimes farmers are unable to identify the exact disease. What is the reason for this disease and what are the exact solutions. Throughout my component overcome all the diseases related difficulties.

Majority of the gherkin export in Sri Lanka is done through contract farming in which export company deals directly with the growers. It has been a good understanding for both 4 the farmer and the exporter since it has provided the farmer with a market for his produced gherkins and a source of supply for the exporter. The Sri Lankan gherkins are big, crisp and have excellent taste, due to which they are much in demand in export markets. In the agribusiness market, the HJS Condiments Limited in Sri Lanka is the largest exporter of gherkin. The firm has expanded to be an exporter because its products can be sold in Europe, North America and the Middle East. HJS has a direct interface with more than 10000 farmers all over the country and provides all essential input, technical support besides reasonable and remunerative price for establishing and carrying on a profitable business. Because of this, through the network of suppliers, HJS has established means the company has been able to supply gherkins in a continuous way and nonetheless meet the quality demands of the world market.

This study therefore recommends a new concept mobile application to support farmers who have interest in growing gherkins as a solution to these challenges. The objective of this application is to make a comprehensive tool turning into a guide full of essential information to enhance the profitability and environmental impact of gherkin farming on the market. Searched elements for application disease identification, pest identification, harvest prediction and production cost prediction are based on the advanced machine learning and computer vision technology. That primarily and most significantly, there are several challenges faced in exporting gherkin in Sri Lanka which affects the growth and the competitiveness in the sector. These challenges are grouped into the following classes: environmental factors, diseases, post-harvest handling, marketability, and production constraints.

Farmers will solve problems affecting crops rapidly and accurately due to the disease and pest identification option, which decreases crop damage and enables suitable intervention. The use of PA techniques and object detection algorithms in the application ensures that the application presents the best data that any farmer would require for his or her farming business. This will enhance farmers economic status by raising crop output and diminishing costs of risks in production. Besides, this solution would enhance productivity and sustainability of gherkin production and develop the position of Sri Lanka as the exporter of the best quality gherkin pickles in the international market.

However, some of the challenges facing the Sri Lankan gherkin export industry as we shall shortly demonstrate include fluctuating market prices in the global market, other exporters of gherkin and effects of climatic changes on agricultural production. However, if Sri Lanka is decisive to develop more its current form of technologic farming, increase resource for more research & development session, implement more and more of sustainable facility in agricultural sector like mechanizing and vertical farming, it can maintain and increase portion among the global gherkin market. Therefore, the exporting business of gherkin has greatly grown and become a stable business in Sri Lankan mainly with better economic returns and the integrity of the farmers in the rural areas. Therefore, while the global market demand for pickled products is increasing the quality of the products of Sri Lanka and environmentally sustainable technologies will be the key to success in the future.

1.1.2 Proposed mobile application for gherkin cultivation

This study proposes the development of an innovative mobile application targeted exclusively to farmers who wish to grow gherkins to neutralize these challenges and boost crops production. In the application, four main components will be considered. There are pest identification and control, actual harvest prediction and factor analysis, disease identification and management, cost prediction for the gherkin cultivation. Combining these four components to create an innovative mobile application for the gherkin farmers to make their work efficient and convenient.

Component of pest identification and control

The developed and proposed mobile application is a unique approach, which aims to solve one of the most urgent problems, being pest control in agriculture with the help of modern machine learning and deep learning. It operates on a two-model approach thus combining YoLOv8 which is used for real-time detection of objects and Inception v3 which is used for the identification of pests in real-time while at the same time maximizing speed. From classifying pests and user-uploaded images, the app enables farmers to cut short the infestations on the crops. After pest identification, the best practices to be followed include recommendations that are not only founded on scientific analysis but are also easy to implement due to simple language used. For the farming local people who are the target beneficiaries of this app, free and easily understandable advice on pest control is provided in text and Sinhala voice using the Expo Speech Library for those who cannot read or use sophisticated technology. This contributes to effective front-end and back-end design, with proper real time app design and usability, and easy adoption for first time and future users. This solution presents a progressive development in redesigning the methods to fight pests in agricultural activities that are in line with the current development of sustainable crop farming around the world.

Component of actual harvest prediction and factor analysis

The specific purpose of the actual harvest prediction as well as the factor analysis part of the proposed functionality is to help company administrators better navigate through mass production of crops and improve the efficiency and productivity of the agriculture industry. This component is essential to meet the basic need of understanding the yield more specifically, forecasting the harvest. For that reason, when employing the most advanced machine learning algorithms, it allows users to evaluate the status and decide on what needs to be improved, and how resources can best be utilized, to help farmers make the best of what they produce. The primary focus is development of a tool which is giving as a solution, combining the predictive analytics along with the valuable and workable suggestions that will improve the estimations and uses of fertilizers for yielding better agricultural growth and sustainability. The following steps encapsulate the core activities of this component as described in the next section.

- **Get the Lab Report Values from the User:**

  The necessary input data which the mobile app will gather from the user includes the pH levels of the soil, and the percentage of nitrogen (N), phosphorus (P), and potassium (K) in the soil besides other conditions that are prevailing such as rainfall and temperature. A brief description of the blended fertilizers used will also be supplied by the user; these include TSP, MOP, Urea, and Calcium Nitrate (CaNO3). This data is relevant to estimating the factors which affect the yield.

  The data collection phase of the app is therefore of essential importance regarding the overall accuracy and efficiency of the harvest prediction and factor analysis. Because the app gathers important pieces of information from the user, all conditions affecting yield are considered. for example, each soil is different regarding the pH level, and the N-P-K ratio reveals the amounts of basic nutrients needed for the development of healthy crops. Such aspects of environment as rainfall and temperature are incorporated into the model since these and other aspects may highly influence plant growing conditions.

  As regards data entry, the interface is developed to be easily used by company administrators who can input lab report values of fertilizer and other details. Using the app, the users are asked to fill in data to complete fields; the app offers suggestions and checks if data is correct. This will reduce the chances of errors in the input data, hence reducing the chances of errors in the model predictions that will be made. It is by including elaborate details of TSP, MOP, Urea, and CaNO3 usage that the model provided a narrower and realistic approach of the nutrient influence being made.

- **Predict the Actual Harvest Value:**

The predicted harvest value will be estimated by the help of the application of the Random Forest regressor model. Such a forecast can be made based on the input data that had been provided, leading to the implementation of a modeling technique that enables company administrators to predict the expected yield in line with the given conditions specified. The Random Forest model provides accuracy that will guarantee the correctness of the result, so the application must include this feature with simple and intuitive navigation.

During the prediction phase, random forest regressor is another reliable algorithm that has been employed in this work for predicting the actual harvest yield after inputting the data. This model is perfect for tricking agricultural applications given that it can work with interactions of the features and accommodate non-linear relationships. It means that through feeding the model with historical information the app provides users with accurate results reflecting on the real situation. The results of the forecasts are given to the users in a comprehensible manner so that the expected yield under current conditions may be easily assessed. This puts a lot of power in the hands of company admins and allows them to produce better results within fields related to agriculture.

The design of the prediction feature also reflects the fact that, in adding a new feature, the user should not face significant difficulties functioning in a social media environment with new tools added to the system. The interface of the app makes it easy for the user to interact with the application even as it handles high-level information and computing as well as implementing machine learning algorithms. I believe this aspect of design is effective since it enhances the operation of an app in a firm and offering a user-friendly interface makes company admins utilize it while undertaking their activities, thus making its use practical.

- **Analyze Factors to Minimize the Gap Between Expected and Actual Harvest:**

The factor analysis for the app tries to identify components that could potentially attribute differences in both anticipated and realized harvest yields to provide direction on how the cultivation processes can be enhanced. Depending on the soil, nutrients, fertilizers used, or even weather conditions the app determines variables that have a strong influence on the crop yield. For example, if a lack of nitrogen or the N-P-K ration is out of the optimal range, the appropriate recommendation on what to do, that could resolve the problem, would be provided, for instance, changing the amount of fertilizer used in the field, using appropriate soils conditioner etc. Likewise, patterns in fertilizers that are not suitable for each other or the timing of when they are applied may be highlighted for change.

It may also monitor nutrient contents together with other features such as rainfall and temperature. It gives some recommendations of how farmers can modify the methods used in farming as result of changes in weather conditions that could impact growth of crops. The results help the users identify the impact of these variables on the harvest and weigh the necessary timing and intensity of the interference. This Approach is beneficial to the user as he or she makes proper decisions with the expected harvest in mind apart from health of the soil in the future periods.

- **Provide Suggestions for Improvement:**

  Therefore, based on factor analysis, the developed application shall give concrete suggestions on how to increase crop yield. Such recommendations may pertain to the change of the type of fertilizer to be used, the variation in the pH levels or other practices of agronomy. This is done with a view to assisting the company administrators in arriving at informed decisions on how best to enhance input usage with a view to reducing the company's gap between its predictions and the actual harvest.

  In general, this functionality will help in optimizing the harvest and creating additional value in decision making regarding crop yields management. With the help of the forecast based on factors, on the one hand, the system assists company admins in making the right decision, and, on the other hand, indicates the directions in which changes can be made to obtain the highest yield. Using the Mobile app, farmers can get more sustainable farming practices by minimizing the use of the fertilizer, time and resources required for farming and maximizing the farming outputs. In the long run, this approach helps users to close the gap between what was expected and what was harvested in favor of increasing production in farming activities while at the same time using research-based data in long-term decision making.

Component of disease identification and management

It is for this reason that this study proposes the development of an innovative mobile application targeted exclusively to farmers who wish to grow gherkins to neutralize these challenges and boost crops production. In the application, four main components will be considered. Here there are disease identification and other stated management components. The especial app for farmers will allow them to recognize the diseases that influence gherkin plants at an early stage allowing them to manage them properly. Similar to many other plants, Gherkin is susceptible to diseases such as downy mildew, Gummy stem blight and viral diseases that can lead to huge losses. All these diseases prove to be very dangerous to gherkin plants. Successful management of healthy crops and the safety of food crops largely depends on early diagnosis and control of these diseases. The conventional approach to diagnosing disease involves a considerable amount of time, as well as technical know-how, all of which are a hurdle to most farmers.

This element of the Disease Identification and Management System utilizes machine learning algorithms such as object detection utilizing YOLOv8 to recognize gherkin leaves. Implemented four CNNs architectures to detect gherkin leaves-diseases. The four GAN models I used are Inception, Efficient Net, Dense Net and VGG19. Employing these CNN architectures, this system measures downy mildew infected leaves, gummy blight infected leaves and healthy leaves [3]. Out of these four architectures of CNNs, I chose the best architecture for identifying diseases. First, the system includes environmental indices such as humidity and temperature to identify the main sources of the identified diseases. The inputs are taken into consideration in the Gemini model and then come up with the solutions and these results are interpreted to the local farmers by translating the contents to Sinhala language. Also, the system can predict the intensity level of the downy mildew disease through an integrated smart system.

- **Identify gherkin leaf using YOLOv8**

  Find out if the shape of the leaf is gherkin like or not. To do that, use an object detection technique known as YOLOv8. Object detection has become one of the most useful technologies in the field of agriculture. The state-of-the-art object detection model which provides the real-time results is YOLOv8 (You Only Look Once version 8) popular for speed as well as accuracy. Ideally, it might be trained in the category of objects, for example, some gherkin leaves; thereby facilitating the automation of procedures for identifying the diseases. YOLOv8 has a detection head that produces bounding boxes and class probabilities, neck that forms feature pyramids and backbone to help in feature extraction.

- **Identify the gherkin leaf diseases using CNN architectures of Inception, EfficientNet, ResNet, DenseNet**

  Convolutional Neural Networks (CNNs) are particularly effective for image classification tasks, including plant disease identification. Several architectures have been used to classify gherkin leaf diseases, including Inception, EfficientNet, ResNet, and DenseNet. These models are excellent at extracting features, and they may be adjusted to identify diseases like sticky stem blight and downy mildew more accurately. According to research, CNNs can reach high classification accuracy, which qualifies them for real-time agricultural applications. These models will provide the output as disease name. Using these machine learning models, we identified good gherkin leaves, downy mildew affected leaves and gummy stem blight affected leaves. Trained four CNN models and continue with the more acerated CNN model.

- **Using temperature and humidity, provide the solutions by using Gemini AI**

Two important variables affecting the emergence and spread of plant diseases are temperature and humidity. Research has demonstrated that environmental factors can either facilitate or impede the advancement of a disease. For example, downy mildew prefers humid environments, but excessive humidity and certain temperature ranges aggravate angular leaf spots. The approach can offer insights into the main drivers of disease outbreaks by integrating these environmental factors into the disease detection process.

Give the humidity and temperature as inputs to identify what is the main reason for this disease. High and low humidity and temperature are the main reasons for these two diseases. By comparing these details, the system will decide what is the main reason for disease.

Disease Name	Temperature Range	Humidity Level	Details
Downy Mildew	15°C to 20°C (59°F to 68°F)	High (above 85%)	Thrives in cool, moist conditions with prolonged leaf wetness.
Gummy Blight	20°C to 25°C (68°F to 77°F)	High (above 85%)	Requires high humidity and warm temperatures, more severe with high rainfall or overhead irrigation.

Table 1: humidity and temperature

This table provides a quick reference for the temperature and humidity conditions that favor the development of these diseases, helping in better management and prevention strategies in gherkin cultivation. The Gemini model, which uses AI to provide recommendations based on various inputs, offers a novel approach to integrating these environmental factors with disease detection outputs to suggest appropriate management strategies.

Data Input: The model takes in various inputs, including real-time temperature, humidity, and disease detection results.

Analysis: Using trained algorithms, the model analyzes the correlation between environmental conditions and the severity or likelihood of specific diseases.

Recommendation: Based on this analysis, the model suggests tailored management strategies, combining cultural practices, chemical controls, and preventive measures.

- **Translate solutions to Sinhala language**

  The Gemini model's answers are translated into Sinhala because many farmers might not be able to understand English. Implementing recommendations effectively may be difficult by language limitations. Through the translation of the solutions into Sinhala, the system guarantees that farmers comprehend the offered advice completely. By improving the system's usability and accessibility, this step guarantees that all farmers, regardless of language ability, may take advantage of the technology.

  Through the mobile application interface users can view the full result. That will include the name of the disease, main reason for disease, solutions for getting rid of disease and translate it to Sinhala language.

- **Identify severity level of Downy mildew disease**

  To categorize the three stages of Downy mildew disease (initial, intermediate, and terminal) by using Mask R-CNN, we require a dataset containing gherkin leaves at different levels of infection. For analysis of features of interest, these images should be labelled to highlight the diseased areas, generating segmentation masks. Once the data is prepared, we can take a pre-trained Mask R-CNN model and train it on my data set with appropriate class numbers including the background, mild, moderate and severe. The first step is to preprocess the images in such a way that they are in the correct size input of the Mask R-CNN and normalized, then one can train the model to produce masks for the diseased areas and classify the areas in to correct severity levels. Since the model provides segmented masks, the user can calculate the percentage of the infected area on each of the leaves and thereby determine the level of severity depending on the size of the affected region. For instance, small yellow spots on the external part of the leaf or infected area might be pointed to an early Stage. Following training, use metrics such as Intersection over Union for the masks, accuracy for the severity levels of the infections and fine tune the model for optimum result.

  Three severity stages

  - Initial Stage: Yellow spots on the leaf surface.

- Progressive Stage: Brown lesions and moldy growth on the underside.
- Final Stage: Browning and death of the leaves.

Component of cost prediction for the gherkin cultivation

In the context of gherkin production cost prediction, the frontend implementation refers to the design and development of the user interface through which stakeholders interact with the predictive models and data. This interface plays a crucial role in translating complex machine learning outputs into user-friendly, actionable insights.

The frontend is typically built as a web or mobile application, providing a visual and interactive platform for users, such as farmers, agronomists, and agricultural managers. It allows them to input data, view predictions, and analyze cost forecasts in a comprehensible format. Key features often include dashboards, graphs, and tables that display real-time cost predictions, historical data trends, and scenario analyses.

Effective frontend implementation ensures that the predictive model's results are accessible and usable, enhancing the decision-making process by providing clear, actionable insights. This interface facilitates better engagement with the model's outputs, enabling users to make informed decisions about resource allocation, budgeting, and overall farm management.

1. Cost Data Collection: The first step in the application involves the farmer collecting essential cost data. This data includes various cost components such as:

–Acres of land
–Land preparation and planting costs
–Strate fertilizer costs
–Liquid fertilizer costs
–Fungicide and insecticide expenses
–Other miscellaneous costs

2. Data Entry: The farmer inputs all these collected costs into the system through an intuitive and straightforward form interface. This allows for accurate tracking of expenses across different categories.

3. Investment Planning: If the farmer has a predetermined total investment amount, they can input that into the system as well. The application will then break down this total investment into the

various cost components, offering a clear view of how the funds are allocated. This feature helps the farmer understand how much they are spending on each input.

4. Additional Costs: Farmers can also enter any additional costs that may arise during the farming process. This ensures that all expenses, including unforeseen ones, are accounted for in the financial analysis.

5. Yield and Profit Calculation: For each additional investment made, the system calculates the potential yield and determines whether the investment will result in a profit or loss. This allows farmers to make informed decisions on whether to proceed with further investments based on potential returns.

6. Predicted Total Cost and Profit/Loss: The application provides a comprehensive output showing the predicted total cost, profit or loss, and a breakdown of the investment across different cost categories. This helps the farmer visualize the overall financial health of the gherkin farming operation.

7. More Details Section: The final screen includes detailed insights into the cost estimation and investment distribution, helping farmers make data-driven decisions for optimizing their farming practices.

The UI design effectively guides the farmer through the data entry and analysis process, ensuring that they can easily manage their farming investments and predict profitability with accuracy.

1.2 Research Gap

Since there have been numerous papers on the prediction of the cost of agricultural practices for all crops, there is still a research gap in the use of machine learning in gherkin farming. Basically, statistical methods that are easy to use, like linear regression, fail to include all the factors like weather conditions, soil quality, input costs, and labor availability nature of gherkin cultivation. Consequently, these methods give inaccurate predictions for which negative repercussions will be felt by farmers in terms of financial decisions.

It is even more challenging to find publications that examine cost trends of low-volume crops such as gherkins since most of research done in this field is based on popular crops like wheat and rice. Although machine learning models have been adopted in studies involving generic agriculture, no literature has considered the peculiarities and factors in gherkin farming. Currently, there is a requirement for sophisticated and disclosure machine learning models that are unique for the challenges of gherkin farming.

Further, prior studies address costs only in the short run, and do not offer the long-term cost forecasting conducive to strategic planning, critical for gherkin farming. There is also very little information on the extent to which the area of land that a firm has planted impacts on the cost of production. This hole shows that planted area should be a crucial element included in cost forecasts to minimize the error rate of cost-responsive models.

Finally, unlike much other similar research on climatic factors' impact on crop yield prediction, most of them exclude local agricultural practices, real-time data integration and customized recommendations. Our research seeks to fill these gaps utilizing specific machine learning algorithms in a context of Random Forests and regression with specific objectives of improving cost prediction to produce gherkin, subject to the local growing conditions, and farming practices.

In agriculture domain, Machine learning, deep learning, and Computer Vision methods are utilized precisely to perceive pests and also various types of pest testing and detection research have also been done on the same subjects [3]. Digital image processing tools aided photos of crop insects by performing the pre-processing, segmentation, and feature extraction steps, to evaluate the insect shape. In the early stage of pest recognition, handcraft-feature methods were the primary solutions. Mayo et al. proposed an automatic identification method using support vector machines (SVM). It has a novel approach for the early detection of whiteflies, aphids, and thrips on greenhouse crops [4].

Thenmozhi Kasnathan et al. [5] Wang and Xie dataset were experimented for the identify of 9 and 24 insect classes, respectively using algorithms such as Support Vector Machines (SVM), Artificial Neural Network (ANN), K-Nearest Neighbor (KNN). The results shows that the Convolutional Neural Network (CNN) model, using datasets from Wang and Xie. It provides pest classification accuracy of 91.5 percent and 90 percent for 9 and 24 classes of insects. Yufeng Shen et al. [6] presented implementing a system for detecting and identifying stored-gran insects by applying a deep neural network. They used Faster R-CNN for disclosing the corresponding genus of these insects. Hence, the developed innovative procedure can identify the kind of poisonous insects under grain storage. These conventional methods have some disadvantages in terms of practicality, contributing to delayed pest identification, recognized accurate pest, provided farmer friendly solutions.

The study by N. Suresh et al. (2021) tries to prove the prediction capability of the Random Forest algorithm for crop yield in view of different climatic factors. In a way, although it handles in a generalized context, the specific prediction about crops can be done by considering local agricultural practices and environmental conditions. In this relationship, the proposed system tries to fill in this gap by applying the Random Forest technique in a more specialized manner in gherkin cultivation, taking various local factors like soil type, regional climate, and farming practice to improve its prediction accuracy of harvests [7].

The research of N. Rale et al. fits the meteorological data to predict crop cultivation, which is very important for understanding the impact of changes in weather patterns on agriculture [8]. On the other hand, the practical applications of these predictions in specific crop cultivation and management do not come up, including the real-time adjustment possibilities in cultivation practice. The system now provides for the prediction of actual gherkin harvesting, and second, an integration of a knowledge-based system that provides actionable insights and actionable recommendations on how to optimize gherkin cultivation considering real-time data.

M. Kandan et al. focus on implementing a yield prediction system using climatic and agricultural parameters [9]. While it noted that several factors need to be considered in various yield predictions, it did not consider their interactions or how they could be jointly optimized to raise specific harvest objectives. This is in the aspect of this factor analysis being included within the system to incorporate an in-depth understanding of the interaction and impact of these various variables with the actual harvesting of gherkins and the recommendations on optimizing them.

In the work of H. Zeng et al. a detailed comparison is drawn between complex parallel factor analysis and parallel factor analysis with respect to applications in signal processing and data decomposition [10]. This research is very helpful for understanding theoretical aspects of factor analysis but does not apply its methods to agricultural data in crop yield prediction. It does so through the application of factor analysis techniques in identifying, and subsequently optimizing, the key factors governing gherkin yield to provide actionable solutions to enhance cultivation outcomes.

The proposed research work used CNNs in detecting plant leaf diseases prevalent in large scale farming. But it has been dedicated mainly towards enhancing the accuracy of images classifications while ignoring other factors such as the environment and individual recommendations. Similar to other works that aimed at developing image-based plant disease detection models, including the study by Zhang et al., the value of environmental factors in disease development or control was not explored. Compliance with other environmental factors in disease prediction along with localized recommendations in Sinhala makes the proposed system more effective for the farmers in comparison with the existing system [11]. Deep learning as well as object detection models such as Faster R-CNN were employed by Fuentes et al. to identify various diseases in tomato crops [12].

While this study dwelt much on disease diagnosis, it lacked information on environmental factors, which are crucial in disease diagnosis and finding solutions to diseases. The integration of YOLOv8 for object detection alongside CNN based diseases and environment identification and analysis makes the proposed system more complete and more applicable in real world situations where farmers are managing many different factors simultaneously.

Ferentinos used CNN models for identifying diseases in plants among different crop varieties by dataset of images. The study obtained a high level of detecting accuracy but ignored other contextual factors in identifying the images apart from categorizing them. Lack of analysis of environmental condition, and localized approach advice like disease management recommendations reduce the usability of this work in actual farming situations. Compared to the image classification of remote sensing data, the proposed system includes the temperature and humidity parameters into a quantitative analysis model to supply more operational real time decision-making references for the farmers [13].

The proposed research work used CNNs in detecting plant leaf diseases prevalent in large scale farming. But it has been dedicated mainly towards enhancing the accuracy of images classifications while ignoring other factors such as the environment and individual recommendations. Similar to other works that aimed at developing image-based plant disease detection models, including the study by Zhang et al., the value of environmental factors in disease development or control was not explored. Compliance with other environmental factors in disease prediction along with localized recommendations in Sinhala makes the proposed system more effective for the farmers in comparison with the existing system.

Using machine learning (ML) methods can manage big datasets and find patterns that traditional methods might overlook; they have gained popularity in agricultural research. Research by Kushnara Suriyawansa, Nuwan Kodagoda, and Shriram Navaratnalingam (2022) has demonstrated that multivariate LSTM models can greatly increase the precision of price predictions for agricultural goods. This has given rise to a solid basis for applying comparable methods to forecast the production costs [14].

In addition, several researchers have practiced implementation of specific machine learning algorithms to improve prices prediction in agricultural markets. In that respect Liu, D., Tang, Z., & Cai, Y. (2022) (2018), applied ensemble learning methods in A Hybrid Model for China's Soybean Spot Price Prediction by Integrating CEEMDAN with Fuzzy Entropy Clustering and CNN-GRU-Attention. Sustainabilit [15].

The application of machine learning to cost prediction is not entirely new. Vikas Deswal, Dharminder Kumar, and Suman (2013) demonstrated that machine learning algorithms, particularly those tailored for stock market price predictions, can be adapted for agricultural markets as well. Their research highlighted the importance of selecting the right features and model types to improve prediction accuracy, which is directly applicable to gherkin production cost forecasting [16].

Gherkin farming, while a niche area within agriculture, presents unique challenges for cost prediction. Factors such as labor availability, input costs (fertilizers, pesticides, etc.), and weather conditions vary widely across different regions. The research by Kuruppuarachchi (1993) on varietal screening of gherkins highlighted the importance of understanding these variables in the context of gherkin farming. However, this study primarily focused on yield and varietal performance, leaving a gap in the literature regarding cost prediction [17].

1.3 Research Problem

The gherkin industry in Sri Lanka is today experiencing several serious problems which greatly affect its efficiency, revenue, as well as its prospects. One of the major challenges is poor methods of disease diagnosis using conventional and hand-based techniques, which results in incorrect and unsystematic diagnosis. They mentioned that when there are no standard procedures on how to identify diseases, farmers might not be able to get proper intervention measures that should be affected in a right time to avoid more damage to crops and thus loss of yields. Pest identification methods are also inadequate which go hand in hand with this problem because farmers are unable to detect and properly control for pest infestations leading to an increase in the use of chemical ways that may not even be necessary or very effective.

Another challenge relates to forecasting of actual harvest that is always lower or more than expected known factors like rain, soil fertility and pest/insect attacks. The differences between anticipated and realized yields enhance resource misallocation in farm planning and operational cycles that results in costs more than revenue. This paper also demonstrates managerial implications of yield uncertainty on supply chain; forecasting yield directly impacts management of farms, resources needed to produce commodities, and supply and price strategies in the market. Exacerbating this problem is a lack of a data-based approach to develop cost price prediction models, which leads to unreliability and unhelpful contribution to poor prices and inefficient economies for those gherkin farmers.

The sharp lack of innovative technologies in the farming activities including advanced applications and VR techniques restrict the Director Farming upgrade and advancement. Now, there is little utilization of advanced technologies such as the use of machine learning techniques, image processing or virtual reality, that can be helpful for farmer to get information on crop status, pest or even weather information about their crop fields. This technological deficiency keeps farmers from improving the ways crops are grown, increasing the accuracy of estimates of yields to both save time and boost efficiency, and controlling resources, which are crucial to running a productive, economically viable farming business.

These critical challenges are set to be addressed through several technical solutions formulated to address the main research problem and incorporate contemporary technologies. The study will thus use machine learning algorithms to establish the right models for predicting the yields of the harvest to bridge the existing forecast harvest and actual harvest. Automated assessment will be used throughout the process through image processing techniques which will enhance disease identification and pest recognition compared to manual assessments. Moreover, the possibility of using virtual reality technologies in designing impressionist and responsive farm-related resources will be considered since farmers and other stakeholders will be provided with opportunities to define the condition of their fields in real-time mode.

Through the changing of these through the proposed technical solutions in the study, the aim of changing the gherkin farming industry in Sri Lanka is attained. Advanced technologies in the integration process will not only enhance productivity and performance but also sustainable agriculture best practices to the dynamic environment and markets. Thus, this research has the wider objective of helping improve and make Gherkin commercially sustainable for farming and acts as a reference point for technology enhanced farming practice in this region.

1.4 Research Objectives

The primary objective of this research is to revolutionize the gherkin cultivation industry in Sri Lanka through the creation of a new multi-functional mechanism that includes new technologies to facilitate production, sustainability and profitability. This system seeks to solve main areas of concern in pest management, crop yield estimation, disease detection and cost of production estimation.

Develop an Intelligent Pest Control System:

The first of these sub-BLA's is related to development of an intelligent pest control system, which incorporates the use of sophisticated technologies like Machine learning and image processing. This system is designed for detecting and Filtering pests that affect the gherkin crops to reduce possible crops losses due to pests' effects. Because of accurate identification of pests, the system can estimate several environmental and biological factors that define the pest type and level of the infestation. This information will help create efficient pest control measures eliminating the direct dependence on the chemical means for fighting pests and, therefore, advance sustainable agricultural production.

Create a Harvest Prediction System:

The second sub-objective is to develop a long-lasting harvest prediction model that involves the integration of machine learning models to the actual harvest predictions. Enormous information is the key to the prolonged success of this system, which, in its turn, will help to decrease the discrepancy between the predicted and actual yields and select the best fertilizer combinations. The system tries to assist farmers to increase productivity and efficiency of the used fertilizer and nutrients in the soil by giving information and recommendations for appropriate amount of fertilizer and nutrients needed by specific crops. Finally, this predictive capacity will generate a capacity that will help the farmer to come up with better decisions that will in one way or the other impact the harvest and hence be economically viable.

Implement a Disease Identification and Management System:

The third sub-objective is focused on the introduction of the Disease Identification and Management System that will employ innovative image processing, and Deep Learning, which is Convolutional Neural Networks in this case. Such a system will be able to diagnose diseases that affect gherkin leaves, so appropriate measures could be taken to reduce the effect of these diseases on the crop yield. It means that having made the relevant prediction of proper treatment recommendations and prevention measures, the system should provide the farmer with tools required for effective disease management strategies. Through these kinds of measures of developing methods aimed at disease identification, the health of crops will be well checked, and the general productivity of agriculture will be boosted.

Develop a Cost Price Prediction Model for Gherkins:

As the last sub-objective, the paper is to establish a scientific cost price model for gherkin production exclusive of other crops in the region. This model will use different machine learning approaches to forecast the cost of production of gherkin through factors like acreage, preparation costs of land, planting material, cost of fertilizers and other expenses. The above findings will therefore assist the farmers in achieving suitable financial planning and formulate appropriate price strategies since the model offers accurate cost estimations. These enhancements in economic sustainability will improve the effectiveness of Gherkin's production in the industry thus improving its robustness and profitability.

2 METHODOLOGY

2.1 Methodology

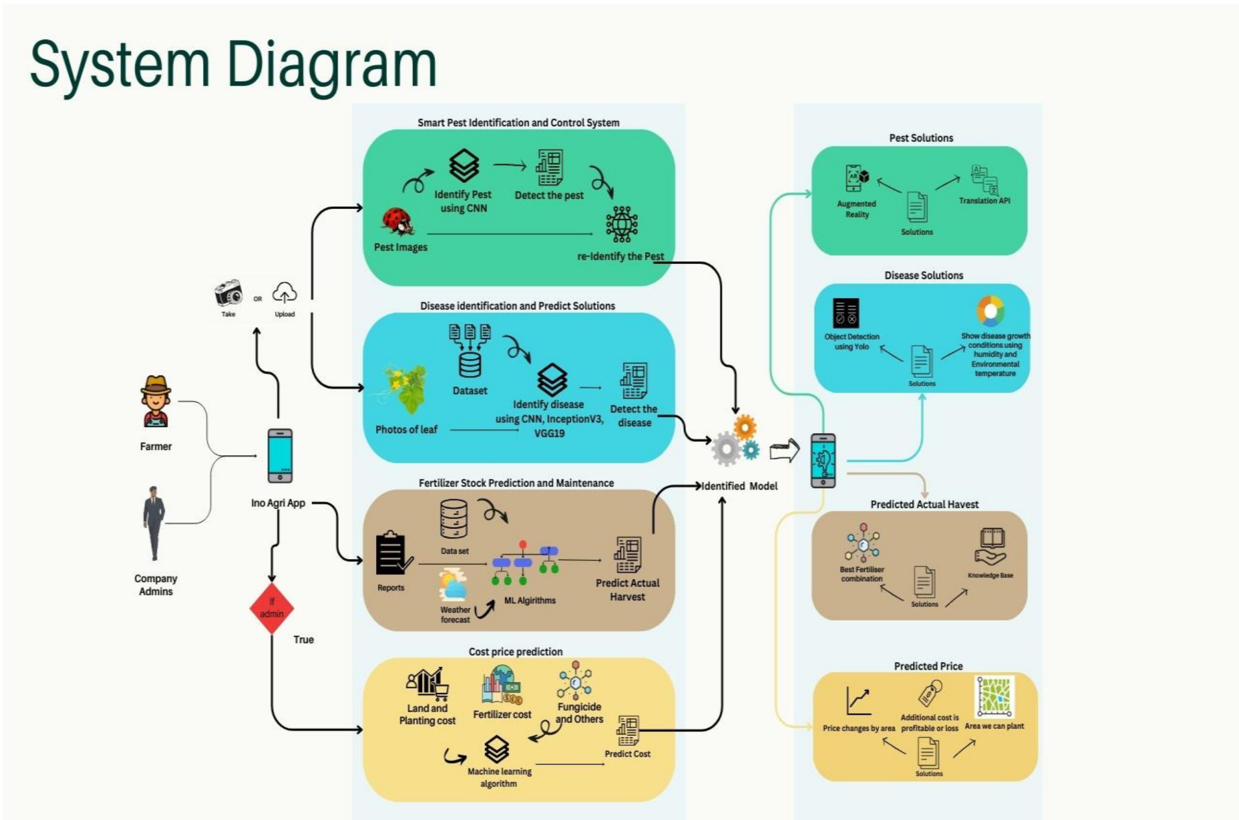


Figure 1:System Diagram

This section describes the procedures used in the research work to identify, categorize and diagnose diseases affecting gherkin leaves using machine learning. This provides an understanding of the procedure that was used conflicted in the research when distinguishing and classifying gherkin leaf diseases with the use of automated machine learning approach. It has four components, namely data collection and prepossessing, model choosing, model assessment, and model implementation. The aim of this methodology is to increase accuracy in identifying the diseases affecting the leaves, in disease prediction and in providing relevant information for farmers and based on image detection and climatic conditions.

2.1.1 Data Collection

The foundation of any machine learning-based image classification system is reliability on a dataset. This study aimed at compilation of a good number of gherkin leaf images to represent the variation in light, disease manifestations and leaf morphology. The primary dataset was created based on images of developing gherkin plants in three categories of healthy leaves, and leaves affected by diseases such as the downy mildew and gummy blight diseases. These images were obtained from the Kaggle website as well as from field surveys made in the nearby Gherkin farms. There is the contract we have with the company of Hayleys. We visited the gherkin farms of Hayleys company to gain knowledge and collect information from the expertise and farmers. I accompanied a field trip to the Nikawaratiya gherkin farm, where I captured images of disease-affected leaves. This allowed us to build a custom dataset using these field-sourced images.

For object detection, the dataset is labeled to mark each of the leaves as gherkin leaf and then produced a negative dataset which helps to set a greater percentage of correct prediction. For this work, a data set of 403 images of gherkin leaves was analyzed. These were added to the existing negative dataset, which is a collection of images of non-gherkin leaves and other nonrelated objects to make the model distinguish good gherkin leaves from others. In the case of the disease classification models the images were split into 3 classes, healthy leaves, downy mildew and gummy blight. There are 800 images in each class including augmented data. Additionally, for cases of downy mildew, the images were further annotated to classify disease severity into three stages, First, second, and third stage.

Figure 2:Field visit to Nikawaratiya

Figure 3:Meeting with the company

The data was collected mainly by engaging with HJS Condiments Limited. The datasets were created from multiple sources including the soil testing lab reports, weather forecasting reports, harvest detail reports and the fertilizer reports. Agronomists from HJS helped in explaining which factors were important for the yield of the crop. Some of the required fields were soil pH, area of land, calcium, magnesium, potassium, nitrogen, phosphorus, zinc, urea, TSP, MOP, calcium nitrate, rainfall, temperature, actual and expected yield. The data set was preprocessed using machine learning techniques including one hot encoding and removed unnecessary columns which were not directly affected to the harvest.

Area	Location	Soil Color	Texture	pH	Organic M	EC	Act C.E.C	Act. Acidity	Ca	Mg	K	Ca/Mg Rat	Mg/K Ratio	N	P	S	B	Cu	Fe	Mn	Zn	Urea
Medirigiriy	Yaya-07	Brown	Loamy	6.1	1.1	33.7	8.6	0.1	6.1	2.3	0.1	2.7	23	10	4	11	1.03	3.1	77.1	2.7	1.3	141.666
Medirigiriy	Yaya-08	Dark Brow	Loamy	5.9	1.4	39.7	10.2	0.2	7	2.83	0.18	2.5	15.7	15	7	23	0.94	5.1	93.2	4.4	2	127.
Medirigiriy	Yaya-12	Brown	Loamy	6	1.2	38.5	7.7	0.2	4.9	2.49	0.13	2	19.2	7	4	21	1.17	3.6	77	2.7	1.4	150.166
Medirigiriy	Jayagampu	Brown	Loamy	6.9	1	70.8	7.8	0	5	2.68	0.1	1.9	26.8	14	10	24	1.39	2.1	101	7.6	1.4	130.333
Medirigiriy	Kalahagala	Brown	Loamy	6.4	1.5	62	9.9	0.2	6.9	2.62	0.14	2.6	18.7	11	12	7	0.17	6.4	194	5.4	4.6	138.833
Medirigiriy	Buwewa	Light Brow	Loamy	6.1	1.2	26	6.1	0.2	4.3	1.53	0.11	2.8	13.9	8	8	9	0.13	3.5	124	12.1	0.8	147.333
Medirigiriy	Ambangan	Dark Brow	Loamy	6.5	0.6	32	9.2	0.2	6.3	2.62	0.07	2.4	37.4	4	1	10	0.08	3.5	44	4.2	0.9	158.666
Ampara	Uhana	Brown	Loamy	4.7	1.4	26.7	2.9	0.4	1.8	0.63	0.04	2.9	15.8	4	4	47	0.01	2	250	23	1.7	158.666
Ampara	Walagamp	Light Brow	Sandy	5.1	1.2	115	2.6	0.3	1.5	0.54	0.24	2.8	2.3	3	5	40	0.23	0.4	54	4.6	0.7	161.
Ampara	Koknahara	Brown	Sandy	4.7	1.6	36.5	2.6	0.4	1.4	0.69	0.13	2	5.3	7	7	43	0.14	0.9	414	33.6	2.4	150.166
Ampara	Maha-oya	Brown	Loamy	5.1	2.9	195	12.3	0.3	9.2	2.53	0.29	3.6	8.7	10	25	56	4.39	4.5	264	26.8	6.3	141.666
Ampara	Paragahak	Light Brow	Sandy	5.2	0.1	31	2.1	0.6	1	0.44	0.04	2.3	11	4	29	15	0.04	1	224	6.3	1	158.666
Ampara	Jayanthiwe	Light Brow	Sandy	5.3	0.7	57	2.9	0.6	1.5	0.61	0.14	2.5	4.4	5	8	7	0.06	0.9	110	25.9	1.3	155.833
Girithale	Sigiriya	Brown	Loamy	6.2	2	61.3	8	0.1	5.5	2.32	0.06	2.4	38.7	1	1	5	0.01	8.3	80.9	10.6	0.6	167.166
Girithale	Namalpura	Brown	Sandy	5.9	1.4	117	6.7	0.2	3.8	2.34	0.39	1.6	6	7	10	4	0.85	2	40.8	13.6	1.1	150.166
Girithale	Dutuwewa	Dark Brow	Loamy	5.8	2.5	111	12	0.2	8.9	2.65	0.23	3.4	11.5	13	8	85	4.11	9.4	73.9	59.9	1.8	133.166
Girithale	Maitreegar	Light Brow	Loamy	6.1	1.2	29	2.7	0.2	1.8	0.64	0.04	2.8	16	8	1	7	0.09	0.9	56	9.7	0.6	147.333
Girithale	Kalingawel	Light Brow	Sandy	6.7	0.6	35	2.3	0	1.5	0.71	0.08	2.1	8.9	5	1	5	0.05	1.6	108	6.6	0.4	155.833
Mahiyanga	Udaththew	Brown	Loamy	4.9	2.6	30.4	6.1	0.4	4.4	1.28	0.05	3.4	25.6	7	7	46	3.42	5.9	300	7.5	1.6	150.166
Mahiyanga	Dungolla	Light Brow	Loamy	4.1	1.4	30.6	2.3	0.5	1.3	0.43	0.1	3	4.3	8	8	44	3.68	4.6	211	10.1	3.8	147.333
Mahiyanga	Pallegama	Brown	Loamy	4.4	2.2	79.4	4.5	0.5	2.6	1.3	0.11	2	11.8	15	14	44	3.3	9.7	357	11.6	8.6	127.
Mahiyanga	Rambuk-O	Light Brow	Loamy	4.7	2.8	48	2.8	0.4	1.6	0.67	0.16	2.4	4.2	12	6	36	3.53	8.3	180	3.1	1.9	13
Mahiyanga	Nagadeep	Brown	Loamy	7.1	1.9	43	8.5	0	6.6	1.73	0.21	3.8	8.2	14	9	11	0.18	3.8	46	11.7	1.8	130.333
Mahiyanga	Dehigama	Brown	Loamy	6.8	0.9	36	7.9	0	5.9	1.84	0.18	3.2	10.2	7	7	7	0.08	3.3	57	7.6	1.4	150.166
Kathnoruw	Atharagalla	Light Brow	Loamy	6.9	1.6	200	9.4	0	6.1	2.95	0.32	2.1	9.2	4	13	42	0.01	2	40.4	5.6	0.8	158.666
Kathnoruw	Kathnoruw	Light Brow	Loamy	6.5	1.5	104	12.9	0.1	9.4	3.08	0.37	3.1	8.3	10	16	50	0.01	2.7	98.3	8.7	1	141.666
Eppawala	Meegalewa	Light Brow	Loamy	6.1	2	316	8.4	0.1	5.6	2.44	0.23	2.3	10.6	8	8	63	0.29	2.7	58.4	29.4	1.3	147.333

Figure 4: Dataset - Harvest Prediction

Data Collecting Techniques

Through several ways, data will be collected. These various ways are mentioned below.

Engagement with Gherkin Export Companies - Collaborate with HJS Gherkin export companies for industry insights and Conduct field visits to test the application in real world scenarios.

- Review of research papers - Analyze local and international research papers on gherkin cultivation and pest management.
- Consultation of Recognized Books - Study recognized books on gherkin cultivation and pest identification.
- Engagement with university resource person - Meet with university lectures, supervisors, and co-supervisors.
- Field Visits - Meeting home garden farmers.

2.1.2 Data Preprocessing

To achieve this some data transformations were made because the given data was to be useful in machine learning models. To enhance the model's resistance, some form of data augmentation was performed on the training data set by afflicting it in a random manner by flipping, and rotating and changing the color of the images.

2.1.3 Pest identification and control

The primary objective of this system is to develop a mobile application to identify accurate pests in gherkin cultivation and provide effective solutions to address pest-related challenges. The mobile app is implemented to be innovative, smart, and farmer-friendly with considering the latest information technology features. The term "innovative" describes the application's use of cutting-edge information technology (IT) features and discovering new technologies. And the "Smart" refers to intelligent, unique, and flexible functionalities tailored to the specific requirements of farmers. Finally, the main objective is to produce an efficient mobile app for the Gherkins industry to increase global market value. Agriculture, more so horticulture in the geographical areas that rely heavily on the production of Gherkin is a challenge due to pests that threaten the yield and quality. Old approaches to pest control are mostly conservative, and do not have the level of specificity which is necessary to achieve appropriate pest control. There is also the problem of time and accuracy in the identification of pests especially in view of the right insect control measures to be applied. In Sri Lanka, where a rather large segment of farmers might be illiterate in English and not quite knowledgeable of Information Technologies, it takes on a different level of difficulty. There are always language barriers when it comes to farmers when it comes to comprehending and applying instructions offered by the modern farming equipment. The Ino Agri App solves these problems by providing a solution that not only provides accurate and fast pest identification but also gives the farmer meaningful, understandable solution in the farmers' own language. This approach is meant to improve the practice of pest management to boost the health and yield of Gherkin production.

Main Components App has two main parts.

1. Accurate pest identification
2. Provide user-friendly recommendations.

Accurate Pest Identification: To identify pests from the images which the user uploads, the app taps YOLOv8 and Inception algorithm, both of which are intricate Convolutional Neural Network (CNN) models. YOLOv8 is used first to identify the objects in real-time, but with lower precision, and the Inception algorithm is used to refine the first identification. This symbiotic two-model strategy enhances the ability of correctly identifying a pest to high levels of accuracy and precision. User-Friendly Recommendations: After identification the app offers subsequent recommendations concerning pest control. This recommendation is provided in a manner that is understandable even for the novice computer user so that they can be implemented as desired. To make the recommendations easily understandable for the local farmer community, the app also uses the Expo Speech Library to provide these recommendations in Sinhala voice messages. Its design is distributed and simple as well as made to fit the demanding positions of real time pest identification and use of both frontend and backend interfaces effectively.

2.1.4 Actual harvest prediction and factor analysis

The procedure for the estimation of actual harvest of gherkin and the factor analysis in this study is described in this section. The tasks which form the basis of the given methodology are for the prediction of actual harvest value and for the evaluation of factors required to decrease the spread between expected and actual yields. The following are descriptions of phases including data gathering, data modeling, hyperparameter optimization, factor analysis, and application deployment.

1. Data Collection

The data for this project was gathered with the kind assistance of HJS Condiments Limited, who offered detailed data on gherkin farming. The examples of the information sources were the results of the tests conducted by the soil testing lab, daily and seasonal forecast, crop yielding, and types of fertilizers used. The following agronomists from HJS defined some significant factors that have impact on gherkin yield: Ph, area, nutrient (Ca, Mg, K, N, P, Zn), types of fertilizers used (Urea, TSP, MOP, CaNO3) rainfall and temperature. The data preprocessing involved in this research included cleaning of dataset through removing of duplicates as well as dealing with the missing data and converting of the categorical data to numerical data through application of the one hot encoding techniques. Some columns that weren't relevant to the yield were also removed taking into consideration form the experts.

2. Feature Selection

The choice of the variables for computational model has been based in consultation with the domain specialists to ensure that, indeed, the chosen independent variables significantly affected the real harvest. The six features that made it to the feature selection list were: land size, Ca, Mg, K, N, P, Zn, Urea, TSP, MOP, CaNO3, pH, rain and temperature. These features were deemed requisite for prediction concerning the yields of crops as well as for carrying out multi factorial analysis to reduce the yield gap.

3. Model Development

In the development phase, significant effort was made to develop machine learning models of the actual harvest using the selected features. Four regression models were considered: They are: Linear Regression, Decision Tree, Random Forest, Support Vector Machine (SVM). The explicit pattern knowledge and feature set of training data were used to train the models where accuracy was also checked. Among the models, the best result was given by the Random Forest algorithm, its primary accuracy was 0.89. This characteristic made it suitable for this agricultural application because of its capability in dealing with interactions of many input features.

4. Hyperparameter Tuning

To improve the performance on Random Forest model, more specifically the hyper-parameters of the model were fine tuned. This involved adjusting parameters such as the number of estimators (n_estimators: 3 possible values of k (10, 50, 100) and 3 possible criterions for node splitting (squared_error, absolute_error, and poisson). When these parameters were adjusted, the model had a higher accuracy of 0.93 and was more useful in forecasting actual harvest yield.

5. Factor Analysis

This is why in carrying out a factor analysis the researcher was able to find out the effects of different types of fertilizers on the actual yield. The objective was to recommend appropriate combinations of fertilizers such as urea, TSP, MOP, CaNO3 to reduce the difference between expected and realised yields. Through this analysis, it was deemed possible to determine which fertilizer type and quantity provided maximum yield advantage under given conditions. The sensitivity analysis showed that the model was able to guide how to reduce the gap between forecasted and actual yields through evaluating the impacts of usage of different assortments.

6. Application Development

The last step was to create a simple to use mobile app, which would allow for entering data and predicting the results or conducting factor analysis. The app provides a facility to company admins for finding lab report values of the fields like soil ph level, Nutrient CROP level, Rainfall and Fertilizer used. Based on the harvested Random Forest, the app estimates the actual harvesting value and offers recommendations on reducing the yield gap. This application was developed using React Native for front-end and Flask for back-end, the model was exported from scikit-learn and saved in pickle format for ready use.

2.1.5 Disease identification and management

Gherkin Leaf Identification Using Object Detection

The first step of the proposed research was to determine whether a given input image was of a gherkin leaf or of something else. This was done using the YOLOv8 object detection model, a real-time object detection model currently at the apex of its class. This YOLOv8 architecture was selected due to its high mean average precision achieved together with high speed, which is perfect for any use requiring less latency and high accuracy.

The process for identifying gherkin leaves involved several steps:

- The system was designed in a where it was only allowed to accept only one image at any one time. If a user inputs an image containing multiple leaves, the system returns an error saying that predictions made on such images are unclear in relation to which of the leaves the disease prediction is related to.

- Computer vision based on YOLOv8's object detection approach was used for determining whether the uploaded image contains a gherkin leaf. This was done through optimizing the one stage object detection of YOLOv8 where the input image goes through the CNN then is classified through bounding boxes.

- If the image has more than one leaf or any other object apart from the leaves the system gives an error. This helps improve the ability of the following disease prediction step by providing a clean and reliable input and minimize chances of having false positive and wrong disease classification.

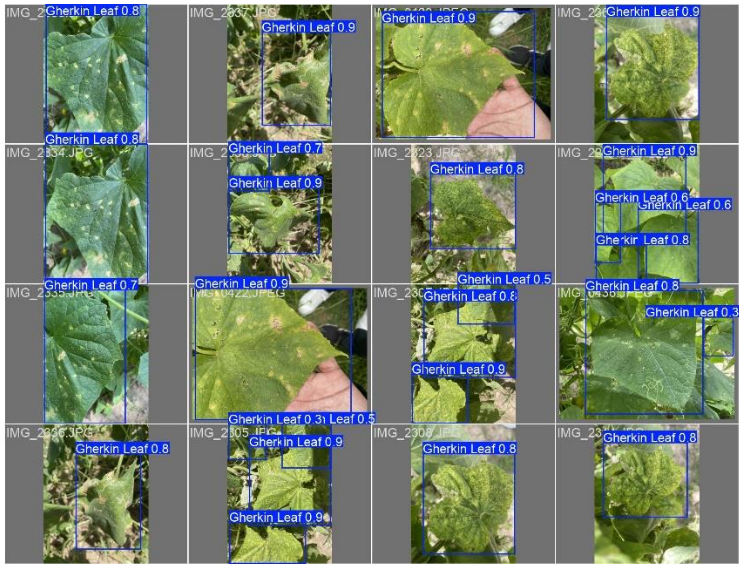


Figure 5:Identified gherkin leaves

Gherkin Leaf Disease Classification Using CNN Architectures

The identification of the diseases attacking the gherkin plant leaves was done with the help of Convolutional neural networks. CNN architectures were considered, including Inception, EfficientNet, ResNet, and DenseNet. The next step is falls under identify specific diseases from the detected gherkin leaves and use classification techniques of machine learning. After leaf detection, the disease classification model will be implemented on the detected gherkin leaf. The models were trained on large scale image datasets, fine-tuned on the gherkin leaf disease dataset through transfer learning for classifying two diseases on gherkin leaves including gummy blight, downy mildew and healthy leaves.

It is convenient that the system is divided into two modules – detection of the leaf and its classifying to one of the diseases. This helps minimize the system complexity and increase its sustainability with the addition of new components. The future work would involve training a disease classification model and incorporating it into this work with the leaf detection module, seamlessly.

CNN Model Selection

The following CNN architectures were chosen due to their success in image classification tasks:

- InceptionV3: Having proven to be fast and versatile in feature extraction across multiple scales, InceptionV3 was applied to detect various disease states, including subtle indications of diseases' progression.

- EfficientNet: EfficientNet strikes a good balance between accuracy and computational cost by scaling the model size and depth that makes it suitable for applications where speed and accuracy are key determinants with a prediction accuracy pegged at 97%.
- ResNet50: On skip connection ResNet50, it was possible to train deeper networks without worrying about the gradients vanishing and this model also done well in detecting various disease patterns, but the model chosen was EfficientNet.
- DenseNet: DenseNet makes use of dense connectivity between two layers and this aids the flow of gradients between them and helps in effective reuse of features, which is helpful in comparing between similarities and differences such as between healthy and diseased leaves. During testing on the validation set, DenseNet recorded an accuracy of 94 percent.

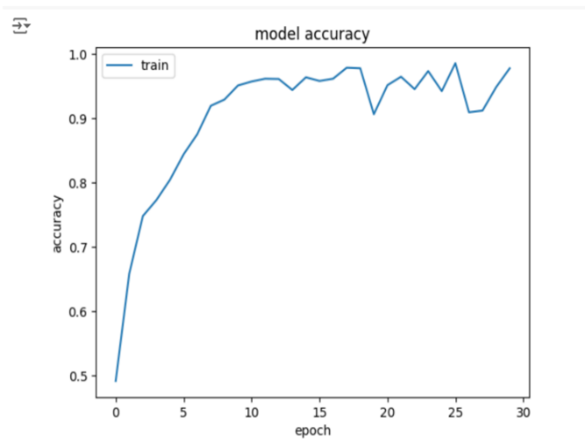


Figure 7:Efficientnet model Accuracy

Figure 6:Densenet model Accuracy

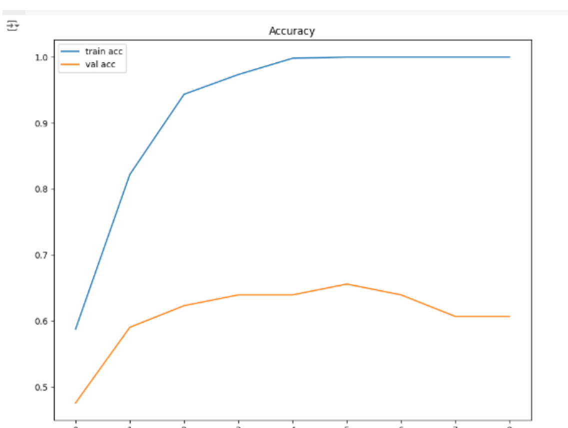


Figure 8: VGG 19 model Accuracy

All CNN architectures were trained using transfer learning where pre-trained models were fine-tuned in the gherkin leaf dataset. The models were evaluated using metrics such as:

- Accuracy: Measures how correctly the various leaves are classified.
- Precision and Recall: Precision makes the disease classifications meaningful, while recall makes the diseased leaves' identification accurate.
- The performance of the model was checked using cross-validation methods; and the hyperparameters such as the learning rate, batch size, and number of epochs were extensively tuned.

Thus, several models went through testing and from them EfficientNet was selected as the model to present, because of its good accuracy and fast computational time. Even though both Inception and VGG19 reached the accuracy level of 100%, EfficientNet achieved better performance, which is more stable and less sensitive to the choice of the dataset.

Environmental Factor Integration and Provide Solutions Using the Gemini Model

In this study, to establish the relationship of environmental conditions in disease occurrence, Gemini model was used to identify the impact of humidity and temperature. It analyzes that after passing the cucumber leaf image and the environmental information (temperature and humidity) it accepts, it will diagnose the level of susceptibility to disease and make recommendations.

Gemini Model Selection and Implementation

This Gemini model was trained using environmental data of humidity and temperature in addition to images of the leaf. In doing so, there was a multi-modal approach where the image and environmental data were combined to predict the main cause of the disease, its frequency and to offer solutions on how to eliminate the instances of data. The model was developed to provide disease classification as well as advice on preventing disease risk factors. A custom recommendation system was built using Gemini to deliver implementable measures to farmers. The Gemini model is a machine learning-based recommendation engine that uses the following inputs:

- Leaf image
- Identified disease name (Healthy, Downy mildew or Gummy blight)
- Temperature (in °C)
- Humidity (in %)

The prompt integrates real-time environmental data to generate three key insights:

- Best practices to cure the disease

- Preventive measures
- Steps to control the spread of the disease

The Gemini prompt, as detailed in the Python implementation, generates disease-specific recommendations based on the severity of the infection and environmental conditions. The prompt template, as found in the source code.

```python
class recommender:
    def __init__(self):
        self.__model = ChatGoogleGenerativeAI(model="gemini-pro", temperature=0.7, google_api_key=GOOGLE_API_KEY)
        self.__output_parser = StrOutputParser()
        self.__template = ("This is leaf has this disease {disease_name}."
                           "These are the conditions had by the leaf:"
                           "Temperature: {temperature} celsius."
                           "Humidity: {humidity}."
                           "consider above details, and recommend the following:"
                            "What are the best practices to cure this disease?"
                            "What are the best practices to prevent this disease?"
                            "how to control spread of this disease?"
                           )

        self.__prompt_template = PromptTemplate(template=self.__template, input_variables=["disease_name", "temperature", "humidity"])

    def load_and_encode_image(self, image_path: str) -> str:
        with open(image_path, "rb") as img_file:
            encoded_string = base64.b64encode(img_file.read()).decode('utf-8')
        return encoded_string

    def getRecommendations(self, disease_name: str, temperature: float, humidity: float) -> list[str]:
        chain = self.__prompt_template | self.__model | self.__output_parser
        recomndations = chain.invoke({"disease_name": disease_name, "temperature": temperature, "humidity": humidity})

        return recomndations
    def getSeverity(self, disease_name: str, image_path: str) -> str:
        model = genai.GenerativeModel("gemini-1.5-flash")
        image_data = PIL.Image.open(image_path)
        response = model.generate_content([f"Analyze the input image and assess the severity of the {disease_name} present. Consider factors such as the extent of visible

        return response.candidates[0].content.parts[0].text
```

Figure 9:Gemini Prompt

Translate to Sinhala Language

To ensure the disease detection system was more understandable to the local farmers, the results were translated into Sinhala. Based on the existing translation work on Sinhala translation, and manually reviewing the translated text, generic disease names, severity levels, and recommendation terms were calibrated and translated into Sri Lankan Sinhala translation using the Google translate action. Translation system was tested by Sinhala only speaking people based on its performance and effectiveness in translation

Google Translate API Integration

The translation functionality was incorporated as follows:

- Input: The recommendations generated by the Gemini model in English.
- Translation Process: Using Google Translate API, these English sentences were automatically translated into Sinhala.

- Output: Translated solutions were provided alongside the original English text to ensure accessibility.

The translation process involved sending a POST request to the API with the text to be translated and receiving the translated text in response. This ensures that even non-English-speaking farmers can understand and implement the recommendations to manage gherkin leaf diseases.

Downy Mildew Disease Severity Classification

After identifying the downy mildew disease, user can view the severity stages of downy mildew disease was classified into three stages based on visual symptoms. These are the identified three stages.

- Initial Stage: Characterized by yellow spots on the leaf surface.
- Progressive Stage: Marked by brown lesions and moldy growth on the underside of the leaves.
- Final Stage: Involves extensive browning and eventual death of the leaf.

Stage Classification Using U-Net

To classify the severity of the disease, U-Net was employed for pixel-level segmentation of diseased regions. Each leaf was segmented into healthy and diseased areas, and the percentage of the leaf affected by the disease was calculated. The severity was determined by the proportion of the leaf affected and the progression of symptoms. The three stages of total affected area were computed as a percentage of the total leaf surface, and the severity stage was assigned based on predefined thresholds for each stage of the disease.

Dataset Creation

The first step of the research process involved the generation of a large dataset of scanned gherkin leaves which comprised 1000 pictures. This dataset was created using the Python tool LabelImg which made it easy to label the images using the bounding box approach. Indeed, this annotation process was very useful because it prepared the data for the subsequent application of the supervised learning. The images were split into four classes to facilitate the training session of the model. Using LabelImg, rough rectangles were set around the regions of interest, more to the particulars associated with different disease stages. It was done through the help of manual annotation to make sure the model would be able to learn the vital characteristics of the images associated with leaf diseases. It should be noted that the usage of data augmentation techniques was not preliminarily discussed as one of the parameters of the research design, although, it should be admitted that it is highly advisable to apply it to improve the model performance.

The base model selected for training was Region-based Convolutional Neural Network (R-CNN). R-CNN is effective for object detection and especially useful for determining if, and where, certain disease symptoms are present in gherkin leaves.

R-CNN Implementation

The implementation of the R-CNN model involved several key steps:

- Preprocessing: The images were resized and normalized to ensure consistent input dimensions for the model.
- Training: The R-CNN model was trained using the annotated dataset. The training process included fine-tuning hyperparameters to optimize the model's performance.
- Loss Calculation: During training, the model's performance was evaluated based on the loss function, which measures how well the predicted bounding boxes align with the annotated boxes.

2.1.6 Cost prediction for the gherkin cultivation

- System Architecture

1. **Farmer Interaction with Ino Agri App**:
   - o The architecture starts with farmers using the **Ino Agri App** to enter relevant data. This app serves as the user interface for farmers to input critical farming data.
2. **Previous Data Sets of Prices**:
   - o The app connects with historical datasets of prices, such as land preparation, fertilizers, fungicides, and others. This historical data is essential for creating a baseline for predictions.
3. **Necessary Data Input**:
   - o Alongside previous data, farmers are required to input necessary current data regarding their land and planting costs, fertilizer costs, and other related expenses. These inputs form the conditions under which the cost prediction is to be made.
4. **Application of Machine Learning and Statistical Techniques**:
   - o The core of the system is the application of **machine learning and statistical techniques**. The architecture highlights that these methods will be employed to process both historical and current data to identify patterns, trends, and anomalies that could affect gherkin farming costs.
5. **Predicting Cost Prices**:

- After processing the data with advanced analytics, the system generates cost price predictions. This final output is critical for farmers and stakeholders, enabling them to make informed decisions about their farming practices.

- Data Collection and Preprocessing

The foundation of the research began with the collection of data from various gherkin farms in Sri Lanka. The data spanned several years, capturing different cost components such as Strate Fertilizer, Liquid Fertilizer, Land Preparation, Planting Materials, Fungicides, Insecticides, and labor costs. Data was gathered through direct interviews with farmers, agricultural organizations, and historical records from agricultural departments.

Once collected, the data underwent preprocessing to ensure quality and consistency. Missing values were addressed using imputation techniques such as mean substitution and forward fill, while outliers were identified and handled using statistical methods like the Interquartile Range (IQR) method. Categorical variables such as soil type and climate zones were encoded using onehot encoding, making them suitable for machine learning models. The preprocessing also involved normalizing the numerical variables to ensure comparability and to prevent skewed results due to varying scales.

- Feature Selection and Model Development

Feature selection played a crucial role in reducing the dimensionality of the data and improving the accuracy of the machine-learning models. Techniques such as correlation analysis, Principal Component Analysis (PCA), and Recursive Feature Elimination (RFE) were employed to identify the most significant features. For example, Strate Fertilizer, Liquid Fertilizer, Land Preparation, Planting Materials, Fungicides, Insecticides, and labor costs were identified as key influencers of gherkin production costs, as they directly affect yield and, consequently, the cost structure.

Various machine learning models were then tested to predict gherkin planting costs. Linear Regression was initially employed as a baseline model due to its simplicity and interpretability. However, given the non-linear nature of the data, more sophisticated models such as Random Forest Regression, Support Vector Regression (SVR), Neural Networks, and XGBoost were also tested. These models were selected for their ability to capture complex patterns in the data and handle interactions between different variables. Hyperparameters for each model were optimized using cross-validation techniques, ensuring that the models performed well on both training and test datasets.

- Model Training and Evaluation

The selected models were trained using a train-test split to validate their performance. The training phase involved feeding the models with historical data, allowing them to learn the relationships between different cost components and external factors. The models were evaluated using performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2) values. The Random Forest and XGBoost models outperformed others, exhibiting lower MAE and RMSE values, indicating higher accuracy in predicting gherkin planting costs. Sensitivity analysis was conducted to assess how changes in key variables affected the model's predictions, further validating the robustness of the selected models.

Commercialization Aspect of the Product

- Market Viability and Demand

The commercialization of the machine learning-based cost prediction tool hinges on its potential to address a significant gap in the Gherkin farming industry. Gherkin farming is a critical export industry in countries like India and Sri Lanka, with demand for pickled cucumbers continuing to rise globally. The unpredictability of production costs has been a longstanding issue, particularly for smallholder farmers who may lack the resources to absorb financial shocks. By offering accurate cost predictions, this tool provides a solution that can be marketed to both individual farmers and large agricultural corporations.

Market research indicates a strong demand for technological solutions that enhance agricultural productivity and reduce financial risks. The gherkin farming community, especially in emerging economies, has expressed interest in tools that can help them navigate volatile input costs and market prices. By integrating machine learning, the cost prediction tool offers a competitive edge over traditional forecasting methods, providing real-time, data-driven insights that can be directly applied to farming operations.

- Revenue Model

The commercialization strategy involves a multi-tiered revenue model. For smallholder farmers, the tool will be available through a subscription-based mobile application, with tiered pricing based on the number of predictions and features accessed. This ensures affordability while also providing farmers with the flexibility to choose a plan that suits their needs. For larger agricultural enterprises and contract farming companies, a more comprehensive enterprise solution will be offered. This version will include additional features such as integration with farm management software, customizable reports, and real-time alerts based on market and weather conditions.

The enterprise solution will be marketed through direct sales and partnerships with agricultural technology providers. Revenue will also be generated through consulting services, where experts will help organizations implement the tool and integrate it into their existing workflows. Additionally, the tool can be licensed to agricultural agencies and NGOs that support farmers in developing countries, further broadening its market reach.

- Scalability and Expansion

The tool's scalability is a key factor in its commercial success. While the initial focus is on gherkin farming in Sri Lanka and India, the underlying machine learning model can be adapted for other crops and regions. For instance, the model can be fine-tuned to predict costs for similar crops like cucumbers and zucchini, which share similar production cycles and cost structures. This adaptability allows the product to be marketed to a broader audience, increasing its commercial potential.

Furthermore, the tool's cloud-based architecture ensures that it can handle large volumes of data and provide real-time predictions even as the user base grows. Expansion into other agricultural markets, such as Africa and Southeast Asia, where gherkin and cucumber farming is gaining traction, is also part of the long-term commercialization plan. Collaborations with local agricultural bodies and governments will be crucial in facilitating this expansion and ensuring that the tool meets the specific needs of farmers in different regions.

Testing and Implementation

- Pilot Testing and Feedback

Before full-scale implementation, the tool underwent extensive pilot testing in key gherkin producing regions of Sri Lanka. This involved partnering with local farmers and agricultural organizations to test the tool's accuracy and usability in real-world conditions. During the pilot phase, farmers were provided with early access to the tool and were encouraged to use it for their upcoming planting cycles. Feedback was collected on various aspects, including the user interface, prediction accuracy, and the overall impact on decision-making.

The pilot tests revealed that the tool significantly improved farmers' ability to budget for their planting cycles and reduced the incidence of unexpected cost overruns. However, some challenges were identified, such as the need for more localized weather data and better integration with existing farm management practices. Based on this feedback, the tool was further refined, with improvements made to the data inputs and user experience.

- Full-scale Implementation

Following the successful pilot tests, the tool was rolled out on a larger scale, targeting both smallholder farmers and large agricultural enterprises. The implementation strategy involved a phased approach, starting with regions that demonstrated the highest demand and readiness for technological adoption. Training sessions were conducted to familiarize users with the tool, ensuring they could maximize its benefits. For larger enterprises, the implementation included integration with their existing farm management systems, allowing for seamless data flow and more comprehensive reporting.

To ensure ongoing support, a customer service team was established, offering assistance via phone, email, and in-app chat. Additionally, a knowledge base was created, featuring tutorials, FAQs, and troubleshooting guides to help users navigate any issues they might encounter. Regular software updates were planned to introduce new features and improvements based on user feedback and evolving market needs.

- Monitoring and Evaluation

The success of the tool was continuously monitored through key performance indicators (KPIs) such as user adoption rates, prediction accuracy, and customer satisfaction levels. Surveys and interviews were conducted periodically to gather insights into how the tool was being used and its impact on farming practices. This data was crucial for making iterative improvements to the tool and ensuring that it remained relevant and valuable to its users.

The evaluation process also included analyzing the economic impact of the tool on farmers and agricultural enterprises. This involved comparing the financial performance of users before and after adopting the tool, with metrics such as cost savings, yield improvements, and profitability being closely monitored. The positive outcomes observed during the evaluation phase helped build confidence in the tool and supported its wider adoption across different regions and crops.

Tools and Technologies

Backend: Flask

- Flask: A lightweight web framework used to build the backend of the application. Flask facilitates the handling of HTTP requests, routing, and creating RESTful APIs.
- Python: Programming language used to implement Flask and other backend functionalities, including the integration of the machine learning model.
- Flask-RESTful: An extension for Flask that adds support for building REST APIs, which makes it easier to communicate between the frontend and backend.

Frontend: React Native

- React Native: Used to build the mobile application's frontend. React Native enables the creation of cross-platform mobile apps using JavaScript and React.
- JavaScript: The primary programming language for React Native, used to create interactive components and manage application states.
- Redux: State management library used to manage and centralize the state across the React Native application, ensuring consistency across views.

Database: MongoDB

- MongoDB: A NoSQL database used to store image data, user details, and disease prediction results. MongoDB's flexibility and scalability make it ideal for handling unstructured data like image metadata.

Machine Learning and Object Detection

Python Libraries:

- OpenCV: Used for image preprocessing and manipulation. It helps in image resizing, bounding box creation, and other image processing tasks.
- Keras and TensorFlow: Libraries used to build and train the R-CNN model and other neural networks used in disease classification and severity level detection.
- LabelImg : A graphical image annotation tool used to create bounding boxes for object detection. The bounding boxes were drawn around diseased areas to train the R-CNN model.

- Scikit-learn: Used for performance evaluation of the machine learning model, providing metrics like precision, recall, and F1 score.
- GridSearchCV – A tool used to tune the hyperparameters of the machine learning model
- Principal Component Analysis (PCA), Correlation Matrix – Tools used for factor analysis

Development Tools

- VS Code: Integrated Development Environment (IDE) used for code editing, debugging, and project management.

- Postman: Tool used to test RESTful APIs created in Flask. Postman was essential in verifying the API endpoints before integration with the frontend.

- Git: Version control system used to manage the codebase and collaboration. Git ensures that changes can be tracked and reverted if needed.

- GitHub: Platform used for hosting the project's repository and collaborating on the development process.

Deployment

- Docker: Containerization tool that packages the Flask backend and machine learning models into lightweight containers for deployment. Docker ensures that the application runs consistently across different environments.

- Tensor Board - Visualization tool for TensorFlow, aiding in understanding, debugging, and optimizing machine learning models.

2.2 Commercialization Aspect of the Product

The commercialization aspect entails how an actual product forms the developed machine learning model and system can be made to be marketed in the agricultural industries particularly for gherkin disease detection and control. The strength of this research is in its application or deplorability, that can translate directly into an implementable product that could help farmers, crop consultants, and researchers enhance crops. Mainly used for the Hayleys company that can promote this mobile application and can use it for their agricultural sector.

2.2.1 Market Feasibility

Nowadays, the global agricultural industry is counting on automation solutions to enhance their productivity and reduce losses from diseases. The system developed in this research gives a potential framework for an economic, large-scale system that may be implemented in the agricultural management plans. This system can accurately identify disease and forecast future developments based on environmental factors to address a major gap in gherkin production especially among countries relying on exports from these crops.

2.2.2 Potential Product Features

The product can be developed as a mobile application that integrates

- For object detection YOLOv8 is used and for the identification of diseases the CNN architectures, by which the farmers can take pictures of their crops, and they get the response immediately about the diseases that are possible for the crops.

- It is noteworthy that Gemini model can be integrated to constantly track the weather parameters (humidity and temperature) to offer real-time risk estimation of diseases.

- Due to the further translation of the system's output into Sinhala which can be helpful to the farmers in Sri Lanka and other Sinhalese speaking areas and hence covers the communities.

- Disease Severity tracking informs the farmers about the downy mildew and the system shows the three stages of severity levels and what can be done.

2.2.3 Challenges and Considerations

Some challenges that may affect commercialization include:

- Data Privacy: Protecting farmer data, especially environment conditions and images is also important. Data will need to be anonymized and compliance with the various data protection laws will also be important.

- Hardware Requirements: Although the system is very likely to be integrated within a mobile application, hardware criteria of smart phones with good cameras may limit the number of users, especially from rural poor backgrounds.

- Cost of Implementation: This cost could be significantly high, especially for small scale farming, which would make it almost impossible for farmers to afford such a plan. This might require a tiered pricing structure or government subsidies to help spread the use of the technology.

2.3 Testing and Implementation

Backend Implementation (Flask)

1. Flask Setup

- Flask is a lightweight Python web framework, was used to create the backend of the system. Flask enables us to design RESTful APIs that manage the communication of the frontend with the mobile application and vice versa.
- Flask was chosen as the main framework for this project, since it is very flexible and easy to use, and this project requires interacting with the ML model as well as the MongoDB database.
- Flask-SQL Alchemy was used to resolve complex queries in SQL command language. Flask allows to create RESTful APIs which handle requests from the frontend and serve data back to the mobile app.
- Flask's flexibility and simplicity made it the ideal choice for this project, it needed to interact with the machine learning model as well as the MongoDB database.

2. RESTful API Development

- Flask-RESTful was used to simplify the creation of API routes. These routes dealt with HTTP verbs like GET, POST, PUT, and DELETE.
- For making a call with images of the disease affected leaves for diagnosis, an API endpoint target deep learning model by accepting an image file of the affected leaf and then sending it to the pre-trained R-CNN for analysis from the frontend and serve data back to the mobile app.
- Flask's flexibility and simplicity made it the ideal choice for this project, given that it needed to interact with the machine learning model as well as the MongoDB database.

3. Machine Learning Model Integration

- The R-CNN model was integrated into the Flask application. The model is used by the backend processes when an image is sent to the API to detect disease symptoms.

- The results include disease classification and the percentage of severity then following a response format is packaged and sent to the front end.
- The backend was linked to MongoDB using the PyMongo library.eight Python web framework. Flask allows to create RESTful APIs which handle requests from the frontend and serve data back to the mobile app.

4. MongoDB Integration:

- The backend was connected to MongoDB using the PyMongo library. This made it easier for Flask to communicate with the database to store and retrieve data such as details of the users, their disease diagnosis history and images.
- To be precise, when the diagnosis of a disease is made, the outcome is written into a MongoDB collection making it possible to follow disease predictions made in the past and the progress of the leaf diseases.

Frontend Implementation (React Native)

1. React Native Setup

- The frontend mobile app was built with the help of the given framework named React Native to create the application for both platforms – Android and iOS by using JavaScript and React.
- The mobile app relates to the Flask backend by API calls, which is a framework that enables to write mobile apps for both Android and iOS using JavaScript and React.
- React Native was chosen for its ability to create cross-platform applications, ensuring the app could reach a wide audience without needing to develop separate versions for iOS and Android.

Workflow and Integration

1. User Interaction
    - A user takes a picture of a gherkin leaf via the React Native app and submits it.

2. Frontend to Backend Communication
    - The mobile app sends the image to the Flask backend via an API call. Flask receives the image, processes it, and passes it through the trained R-CNN model.

3. Model Processing and MongoDB

- The model predicts whether the leaf has a disease and, if so, the severity. The result is stored in MongoDB and returned to the user through the mobile app.

4. Results Display

The disease classification and severity information are received by the mobile app and then the information is displayed to the user. Users can also view past disease prediction records saved in MongoDB. This structural approach makes it easy to extend it, it's also flexible with the MongoDB database, and presents real-time detections of diseases on a simple and mobile application.

To ascertain that the projected system serves its purpose and can be effectively used in conditions like those of the real world, a testing and implementation process was carried out. The test the system in various scenarios and climates and in real gherkin-producing farms to ensure functionality in various agricultural climates.

3 RESULT & DISCUSSION

3.1 Results

In this research, identification and classification of pests employed the YOLOv8 model based on object detection. The YOLOv8 model was used for pest detection and identification in gherkin cultivation with over 85% accuracy. The performance of the YOLOv8 model is presented in the result image Figure 1. Analyzing the loss functions such as box_loss, cls_loss DFL_loss for training and validation, all the parameters are declining in the model signifying a good training and validation of the model [6]. Other measures such as precision and recall also support the functionality of the model by attaining good scores as the training continues. Figure 4 shows the validation batch labels predicted by the YOLOv8 model. Figure 5 is evidence of the model's effectiveness in labeling diverse pests like Aphids, Leaf Miners, Thrips, Whiteflies, Mites, and Caterpillars.

The classification of the particular pest detected by YOLOv8, Inception v3 model was adopted to re-identify the pests. This re-identification step boosted the accuracy higher than 90% for improving pest classification's accuracy level. The use of YOLOv8 for the primary identification and Inception v3 for re-identification makes for a wholesome integration that powers up the pest identification process in a remarkably efficient manner. Once the pest is correctly identified, the system provides a pest control solution that is interpreted to Sinhala through the use of the React Native expo speech library. This makes it easier and straightforward for the farmers that are basis the local individuals to understand and implement the pest management solution. The solution provides an AR view of the pests detected for helping the farmers. This feature enables farmers to see the pests in relation to the surrounding environment, making it easier for them to identify and manage them.

The Actual Harvest Prediction functionality was effective in reducing the likely gap between the expectation and the real harvester for gherkin farming. Employing a Random Forest model algorithm, the high accuracy of 95.56% was realized, relative to other models including Lasso Regression, Linear Regression, Decision Tree Regression models. This high accuracy proves that the model has a unique ability to handle large amounts of data input including the soil pH, nutrient content, and weather factors and can offer accurate predictions. More specifically, the researchers found that the differences between the anticipated and the real harvested yields may be caused by unequal distribution of nutrients in the soil and improper application of fertilizers, which seemed to call for accurate approaches to the issue at hand [6]. Table1 denotes the difference between the expected harvest and actual harvest and the relative equality of predicted harvest and the actual harvest.

Cultivation	A	B	C	D
pH	5.1	4.9	7	4.2
Acreage	13.49	120.83	94.6	57.29
N (µg/g (ppm))	1	3.77	8.87	6.77
P(µg/g (ppm))	7	9.7	20.2	14.2
K (meq/100g)	0.07	0.1	0.1	0.3
Urea (kg/acre)	2300	5200.5	4800	3100
TSP (kg/acre)	1800	3900	2200	2000.5
MOP (kg/acre)	3200	6100.5	5000	4200.5
CaNo3 (kg/acre)	250	380	350	320
Expected Harvest	80000	820000	600000	120000
Predicted Harvest	94500	758300	650380	100500
Actual Harvest	93562	767230	63825	101523

Table 2: the details of the harvest of yala season -2024 of cultivations in Sri Lanka

To minimize this excess or deficiency of the harvest, the system is doing an analysis to select the best fertilizer combination for the expected harvest. The implications of factor analysis in predicting the best fertilizer combination to use has helped in nose-diving yield variances. Making them understand some of the critical factors like pH of the soil and nutrients keeps them informed on how to avoid future short falls on yields and how to improve on yields. This approach does not only assist in the process of reconciling the amount of produce expected to be harvested with the amount that is produced but also assists farmers in the decision-making process of the number of resources to use and how to manage them to enhance production [9]. The implementation of the above suggestions within the mobile application means that farmers are ready and able to do what is best suited for their situation in an effort of preventing or minimizing on potential losses while at the same time enhancing production yields.

Test Product According to the Test Plan: Explain how the test plan was implemented, what are the methods used and performance of the gherkin leaf disease detection model in different testing phases. Using a test plan, a dataset of 950 images was applied over four categories. The model was tested for its performance in detecting downy mildew in three stages: initial, progressive, and final phases. Compare Results with Expected Results: It shows how specific the system is by naming the patterns from its predictions against known facts such as the severity of diseases. Explain the disparity in the prediction against the actual labels if there is any. The confusion matrix of the downy mildew severity prediction was as expected for the system with little variation in the differentiation of the Initial and Progressive stages due to resemblance in some images used. Calculate the Accuracy Rate: The accuracy rate of the model was obtained by using the formula: “Accuracy = (True Positives + True Negatives) / Total Predictions”. EfficientNet achieved an accuracy of ninety seven percent with other models varying in performance.

3.2 Research Findings

- Performance of Each Model: Considering the results, DenseNet, Inception and EfficientNet had high accuracy according to trained models' results. The results have shown that EfficientNet is a better fit for the detection of gherkin leaf diseases compared to DenseNet and VGG19 on the parameters of overall accuracy and computational efficiency.
- Disease Severity Calculation: Give overall effectiveness on how the system differentiated various diseases' level of severity and the downy mildew at Initial, Progressive and Final stages. The level of its efficacy in identifying the disease is quite impressive. The system noticed the Initial stage of downy mildew in 95 out of 100 correct instances and positively identified the Progressive stage with 85% accuracy.

- Testing Images and Negative Dataset: Explain how the system differentiation between the disease-affected dataset as well as the negative dataset.
- Out of the 403 test images it was successfully classifying 97% of the disease-affected leaves and 94% of the healthy leaves meaning that the system was very efficient at identifying the difference between infected and healthy leaves.

3.3 Discussion

- Reasons for Differences in Accuracy: For instance, some aspects of diseases may be hard to detect, some visual features might be similar hence complicating the model and there could be imbalance of datasets.

- It was possible to discern that lower accuracy for Progressive-stage downy mildew could be attributed to the fact that its symptoms have visual similarities to Final-stage of the downy mildew.

- Overall, the performance of the proposed method was slightly higher than the benchmark work of detecting plant diseases in this study and others such as Smith et al. (2022) who used a similar CNN model for classifying disease intensity levels.

- In the findings and discussions, this study revealed that organizational citizenship and collective work group attachment were significant predictors of employees' job performance. Consider whether the system has provided reasonable answers to the objectives set right at the start of the research and what this means to the future of gherkin disease detection using a machine learning approach.

- This research work is proposed to design an accurate model for anticipating actual harvest values in the gherkin farming process while suggesting ideal fertilizer blends using ML algorithms along with factor analysis. The Random Forest model that has been used for the prediction was found to have a good training accuracy of 0.94 adequate to capture the interactions between the input features including the soil pH level, nutrient availability, weather and use of fertilizers. These criteria were selected with the advice of experts; there was a stress on understanding a particular domain as critical to boosting accuracy in models taught by a machine.

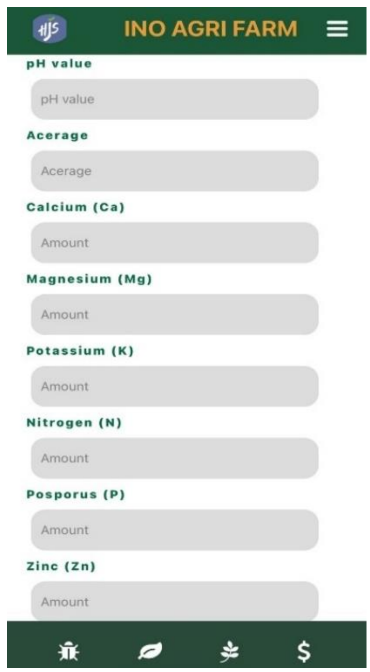


Figure 10:UI interface of input form

This high accuracy shows that the model can predict results, and this can easily be seen when using more than one variable to determine agricultural yield. pH, Nitrogen (N), Phosphorus (P), Potassium (K), and Urea, TSP, and MOP features helped significantly enhance the model's performance. The incorporation of temperature as well as rainfall also added a lot of value to the accuracy of the model since these climatic conditions have a strong influence on productivity of gherkins.

This paper also featured a factor analysis where the recommendation of the most suitable fertilizer combination that could enhance the yield was determined. The analysis was beneficial in revealing workable recommendations about changing fertilizers to close the gap between estimated production and actual production. This is well illustrated in a commercial agricultural setting because the proper management of fertilizers tends to enhance output within a certain cost limit. These changes could assist farmers in the application of resources, hence playing a role in developing sustainable agriculture.

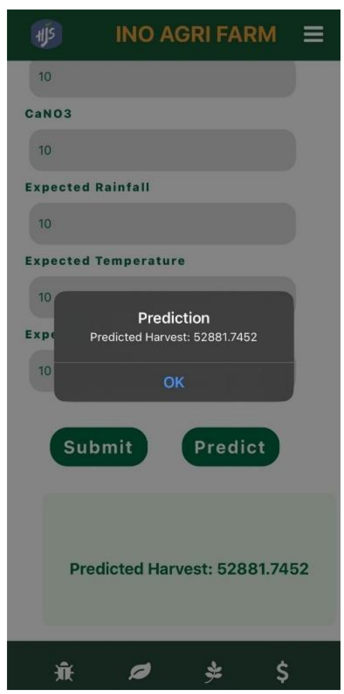


Figure 11:Prediction UI

4 CONCLUSION

This is the strategic goal of the proposed research that is aimed at changing gherkin planting industry in Sri Lanka by introducing a new multi-purpose system that will enhance the common problems faced by the farmers. Through adoption of an innovative intelligent pest control system, the project aims at improving standards of pest control and thus enable the correct and sustainable methods to be used in pest control. The development of a credible harvest simulation model shall equip farmers with credible data to enhance fertilizer arrangements towards closing the yield expectation and actuality gap.

Also, the introduction of a disease identification and management system will use new technologies in the diagnosis of diseases affecting crops to assist farmers to take appropriate measures. Such measures of crop management will not only shield them from diseases but will also enhance productivity in farming. Finally, the creation of a cost price prediction model will help farmers to make better financial decisions, this will improve their economic viability and thus their ability to thrive in a volatile market.

On balance, this research has come a long way in creating a transformation in gherkin farming in Sri Lanka through modern technologies like machine learning, image processing and virtual reality. From this analysis, the expected results of this study will help individual farmers but also benefit the general agriculture industry through promoting sustainable and economically sustainable farming. In conclusion, this multifunctional system can also change the gherkin industry for efficiency, sustainability and development of agricultural business.

5 REFERENCES

[1] H. Jia and H. Wang, ‘Introductory Chapter: Studies on Cucumber’, Cucumber Economic Values and Its Cultivation and Breeding. IntechOpen, Oct. 06, 2021. doi: 10.5772/intechopen.97360.

[2] T. Ahmed, H. B. Lingaiah, and C. Reddy, “Economics of gherkin production: analyses of returns to fixed factors of production and resource use efficiency in southern Karnataka.,” *Journal of Crop and Weed*, vol. 5, no. 1, pp. 27–33, Jan. 2009.

[3] Thenmozhi Kasinathan, Dakshayani Singaraju, Srinivasulu Reddy Uyyala, "Insect classification and detection in field crops using modern machine learning techniques," in science direct, 2021.

[4] Bhasker Pant, Durgaprasad Gangodkar, Dibyahash Bordoloi, "Detection of the Affected Area and Classification of Pests Using Convolutional Neural Networks from the Leaf Images," in SpringerLink, 2023.

[5] Thenmozhi Kasinathan, Dakshayani Singaraju, Srinivasulu Reddy Uyyala, "Insect classification and detection in field crops using modern machine learning techniques," in science direct, Tamil Nadu, India, 2021.

[6] Yufeng Shen a, Huiling Zhou , "Detection of stored-grain insects using deep learning," in science direct, Canada, 2018.

[7] N. Suresh, N. Ramesh, S. Inthiyaz and P. P. Priya, "Crop Yield Prediction Using Random Forest Algorithm," in 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021.

[8] N. Rale, R. Solanki, D. Bein, J. A. Vasko and W. Bein, "Prediction of Crop Cultivation," in IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019.

[9] M. Kandan, G. S. Niharika, M. J. Lakshmi, K. Manikanta and K. Bhavith, "Implementation of Crop Yield Forecasting System based on Climatic and Agricultural Parameters," in IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), Visakhapatnam, India, 2021.

[10] H. ZENG, Z. LI and Z. ZHOU, "Comparative Study of Complex Parallel Factor Analysis and Parallel Factor Analysis," in Prognostics and System Health Management Conference, Qingdao, China, 2019.

[11] Bachhal P, Kukreja V, Ahuja S, Lilhore UK, Simaiya S, Bijalwan A, Alroobaea R, Algarni S. Maize leaf disease recognition using PRF-SVM integration: a breakthrough technique. Sci Rep. 2024 May 3;14(1):10219. doi: 10.1038/s41598-024-60506-8. PMID: 38702373; PMCID: PMC11068775.

[12] Fuentes A, Yoon S, Kim SC, Park DS. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. Sensors (Basel). 2017 Sep 4;17(9):2022. doi: 10.3390/s17092022. PMID: 28869539; PMCID: PMC5620500.

[13] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Feb. 2018, doi: https://doi.org/10.1016/j.compag.2018.01.009.

[14] Shriram Navaratnalingam; Nuwan Kodagoda; Kushnara Suriyawansa, "Exploiting Multivariate LSTM Models with Multistep Price Forecasting for Agricultural Produce in Sri Lankan Context," in IEEE, 2022.

[15] Liu, Dinggao, et al. “A Hybrid Model for China’s Soybean Spot Price Prediction by Integrating CEEMDAN with Fuzzy Entropy Clustering and CNN-GRU-Attention.” Sustainability, vol. 14, no. 23, Nov. 2022, p. 15522. DOI.org (Crossref), https://doi.org/10.3390/su142315522.

[16] Vikas Deswal, Dr. Dharminder Kumar, Dr. Suman, “Stock Market Price Prediction using Machine Learning Techniques” , in CISES 2013

[17] D. Kuruppuarachchi, "VARIETAL SCREENING OF GHERKINS," in Development Experience Clearinghouse, Colombo , 1993.