International Current Journal of Engineering and Science (ICJES)
Volume 3, Issue 11, pp 9-15, November-2024
DOI: https://doi.org/10.47001/ICJES/2024.311002
ISSN (Online): 2583-7257

Prediction of the Diseases in Cucumis Sativus and Soil Health Using Machine Learning Algorithms

[1]Catherine Mary, [2]John Okello Okuto, [3]Simon Osoro

[1,2,3]School of Agriculture and Environmental Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

***Abstract* - Cucumber (*Cucumis sativus*) is an economically important crop worldwide, but its main challenges are various diseases affecting its leaves and fruits. This paper presents a comprehensive review of cucumber leaf diseases, cucumber fruit diseases, impact on marketing strategies and Land recommendation using soil fertility and weather prediction, common diseases like downy mildew, powdery mildew, belly rot and furthermore, using machine learning algorithms to develop soil prediction models to monitor soil health indicators and identify disease causative agents will be coming. The findings reveal significant relationships between soil type, environmental factors, and incidence of cucumber diseases. This information enables stakeholders to implement targeted interventions such as soil amendments, crop rotations and disease-resistant varieties to reduce disease pressure and improve crop yields and quality. Furthermore, a comprehensive analysis of cucumber marketing strategies and consumer preferences provides insights into effective market segmentation and promotional strategies.**

Keywords: cucumber diseases, VGG16, YOLOv5, weather prediction, prediction method, AI, Image processing.

I. INTRODUCTION

The reduction in agricultural yields resulting from crop destruction represents a significant contemporary concern. Particularly within the realm of perishable crops, the loss incurred from the destruction of cucumber crops poses a substantial economic setback for the nation. The work of Shampa Chakraborty, Sadhana Rayalu [1] gives about beneficial effects of cucumber. The cucumber holds considerable economic potential through exportation. The cultivation of cucumber offers multifaceted advantages to the nation. Hence, the detriment caused by its destruction extends to the broader interests of the entire country. Crop destruction arises from various issues such as cucumber fruit diseases, cucumber leaf diseases, and inappropriate selection of land for cucumber cultivation. The primary causative factor behind these challenges lies in the insufficient knowledge among cucumber farmers to mitigate these issues effectively. The consequential destruction of extensive cucumber plantations due to such factors constitutes a significant and pressing concern. Hence, the objective of this research is to propose remedies for the challenges associated with cucumber fruit diseases, cucumber leaf diseases, inadequate land selection, and favorable weather conditions for cucumber cultivation, as well as to offer insights into assessing the marketing performance of cucumbers for the benefit of farmers.

This research introduces an automated solution aimed at assisting farmers in cucumber (Cucumis sativus L.) crop management. It encompasses tools for controlling fruit diseases, leaf diseases, conducting market analyses of cucumbers, and recommending suitable land and weather conditions for cucumber cultivation. Through this mobile application, farmers can effectively identify cucumber fruits, diagnose Belly Rot disease, and detect cucumber pests using image recognition technology, subsequently accessing relevant remedies. Additionally, the application facilitates the differentiation between healthy cucumber leaves and those afflicted with downy mildew or powdery mildew, enabling farmers to discern disease stages and access corresponding remedies. Furthermore, it provides insights into optimal harvesting months and identifies locations where cucumbers fetch the highest prices based on the farmer's geographical location. Moreover, the application assists farmers in determining land suitability for cucumber cultivation and provides weather forecasts aligned with the farmer's specified planting and harvesting timelines. This paper underscores the potential for future research and development in this domain.

II. LITRATURE REVIEW

The fungus Rhizoctonia solani Kühn [3], which causes belly rot, is a major global danger to cucumber (Cucumis sativus) production. Effective management measures are crucial for this disease, since it can cause severe losses up to 80% in individual fields and cause annual crop losses ranging from 5% to 10%. Fungicides are the current method of choice; however, it is not as inexpensive or sustainable due to the absence of resistant cultivars. Since genetic resistance is a desirable substitute, attempts have been made to find cucumber germplasm that is innately resistant to belly rot [4].

International Current Journal of Engineering and Science (ICJES)
Volume 3, Issue 11, pp 9-15, November-2024
DOI: https://doi.org/10.47001/ICJES/2024.311002
ISSN (Online): 2583-7257

The discovery of these resistant cultivators lays the groundwork for additional study and the creation of cucumber cultivars with increased resistance to belly rot. Relying only on fungicides is not as sustainable or economical as considering the possibility of genetic resistance. To introduce resistance into commercially viable cucumber varieties, it is imperative to examine the mechanisms of resistance in these cultivars, comprehend the genetic basis of the trait, and investigate breeding tactics.

Furthermore, since this study was carried out in the early 1990s, it is imperative to examine more recent research to evaluate any developments in our knowledge of cucumber belly rot resistance, such as molecular methods, marker-assisted breeding, and the creation of resistant cultivars that may have arisen since then [3].

Cucumis sativus crops are vulnerable to a range of diseases and pests that, if left unchecked, can result in severe crop loss or output decline. Effective management strategies depend on the early and precise detection of pests and diseases affecting cucumber leaves. Traditional manual detection techniques, however, are slow and don't work in real time. Furthermore, the development of reliable automated detection algorithms has been hampered by the lack of publicly accessible datasets for photos of cucumber leaf disease in real-world situations.

Difficulties in Cucumber Leaf Disease Detection: The literature that has already been written stresses the difficulties in using manual detection techniques and the demand for automated alternatives. The need for cutting- edge technology to improve cucumber leaf disease detection is highlighted by the drawbacks of labor- intensive methods and their lack of real-time performance [5].

Dataset Creation and Gathering: To fill the dataset gap, this project gathers and creates a new dataset on pests and diseases that affect cucumber leaves in a practical setting. The literature emphasizes the importance of building these datasets and how they may be used to train and assess machine learning models for crop disease and pest detection [6].

YOLO-based Models in Agriculture: A common trend in literature is the application of YOLO (You Only Look Once) model architecture in crop disease detection and agriculture. The efficacy of YOLO models for object detection has been demonstrated in earlier studies, which qualifies them for quick and precise identification of pests and illnesses on cucumber leaves [5] [7].

Comparative Analysis with the YOLOv5 Model: The literature analysis recognizes the importance of the original YOLOv5 model for object detection applications. In terms of mAP, precision, recall, and model size, the suggested enhanced model performs better than the original YOLOv5 model, according to a comparative analysis provided in this paper [5].

The oomycete disease Pseudoperonospora cubensis, which causes downy mildew, is a danger to cucumber (Cucumis sativus) crops all over the world. Cucumber leaf metabolism is altered because of complex interactions between the pathogen and the host plant throughout the disease's development. Changes in transpiration rate have been found to be one of these factors that significantly affect how downy mildew spreads.

Transpiration Rate and Disease Progression: Downy mildew development is negatively correlated with transpiration rate, an important physiological characteristic in plants. Research has investigated how the pathogen modifies the transpiration system of the host to help colonies and spread. Gaining knowledge of these dynamics can help one better understand the host-pathogen relationship and the variables affecting the severity of the disease [8] [9] [10].

Early Detection and Discrimination under Controlled Conditions: The ability to distinguish between healthy and infected areas in thermograms is made possible by temperature variations seen in infected leaves. The promise of thermography as a proactive tool for disease treatment is highlighted by this early detection, which happens even before obvious symptoms show. Scholars investigate the usefulness of this method in controlled settings [11][10] [12].

Cucumis sativus L. crops are seriously threatened by powdery mildew (PM), which can result in serious foliar diseases [13]. Developing resistant cucumber cultivars has been hampered by the difficulty in figuring out how PM resistance is inherited, despite the economic significance of the trait. Through the discovery of defining-related genes and signaling pathways, recent research has made significant progress in understanding the molecular mechanisms behind PM resistance. Furthermore, the functional validation and identification of PM resistance genes advance our understanding of the interactions between powdery mildew and cucumbers.

The Value of Good Soil for Growing Cucumbers: As a horticulture product, cucumbers (Cucumis sativus) have significant economic potential and can be grown for both local and foreign markets. The cultivation of cucumbers is highly dependent on the fertility and conditions of the soil, which emphasizes the importance of using the best soil management techniques to improve agricultural growth and productivity.

Volume 3, Issue 11, pp 9-15, November-2024
DOI: https://doi.org/10.47001/ICJES/2024.311002
ISSN (Online): 2583-7257

Agriculture's Use of Control and Monitoring Systems: Providing real-time data on critical parameters, control and monitoring systems are essential to contemporary agriculture. This review of the literature examines current techniques for tracking soil conditions in different crops, with an emphasis on how well they work to optimize farming operations.

Applications of the Internet of Things (IoT) in Precision Agriculture [15]: Precision farming has been transformed by the incorporation of IoT principles. The literature on IoT applications in agriculture is reviewed in this section, with a particular emphasis on the use of IoT platforms to monitor and manage soil conditions during crop production [16]. Sensor technology for Soil Monitoring: The ability to monitor soil has been significantly improved by advances in sensor technology. This section of the literature review examines several soil sensor types, emphasizing their value in improving agricultural operations. These types of sensors include pH, temperature, and soil moisture sensors [17].

Testing and Validation of Soil Sensors: Determining the accuracy and dependability of soil sensors requires their validation [16]. This subsection compares the performance of soil temperature, moisture, and pH sensors by reviewing the literature on testing techniques and presenting findings from related studies.

The production of cucumbers is hampered by harsh weather and other environmental factors. Researchers have studied and modelled the effects of low temperature (LT) and poor light (PL) pressures on cucumber growth to address these problems. This review of the literature examines significant discoveries in this field, with an emphasis on stress parameters, experimental methods, and the creation of models to better understand and forecast the reactions of cucumber plants.

Creation of Models for Cucumber Growth Prediction: Based on temperature and light characteristics, researchers have created models to improve comprehension and forecasting abilities, by simulating above-ground dry weight, these models hope to shed light on the intricate interactions that occur between LT, PL stress, and cucumber growth. In comparison to the GDD and TEP models, the LTE model shows potential accuracy with a smaller root mean square error (RMSE), according to comparative analyses.

Vegetable prices are very unpredictable and volatile, which presents serious problems for growers, consumers, and governments alike. In the vegetable market, regulating these changes and putting in place efficient early-warning systems depend heavily on forecasting techniques. Examining previous studies on time series forecasting, this literature review highlights the necessity for precise models that can represent both linear and nonlinear trends in vegetable prices [18].

The significance of time series forecasting in agricultural economics has been previously established, especially regarding the prediction of vegetable prices [19]. Prices fluctuate in both linear and nonlinear ways frequently, thus accurate forecasts require sophisticated modelling methods.

Difficulties of Linear Forecasting Models: When handling the nonlinear correlations included in vegetable price data, linear forecasting models—which are frequently utilized in time series analysis—face difficulties [19].

These models might not adequately represent the complex dynamics impacted by a range of variables, including exogenous shocks, demand-supply imbalances, and seasonal patterns.

Difficulties with Neural Network Modelling: Although neural networks are highly effective at capturing intricate patterns, they can become unresponsive when faced with simultaneous linear and nonlinear components. Forecasting accuracy may suffer from traditional neural network models' inability to identify the unique properties included in the time series data [18].

Experimental Evaluation of Vegetable Prices: The experiment focuses on the price of vegetables, namely cucumbers. Conventional techniques such backpropagation artificial neural network (BP-ANN) techniques and autoregressive integrated moving average (ARIMA) are compared. Evaluation measures demonstrate how well the suggested hybrid model predicts cucumber prices.

III. METHODOLOGY

The "Cucumber Farmy" mobile application is specifically designed to cater to agriculture professionals and farmers, helping in the identification and management of cucumber leaf diseases, cucumber fruit diseases, marketing analysis, and land recommendations for cucumber cultivation. Leveraging image processing, machine learning, and IoT devices, the application processes leaf and fruit images, as well as numerical data, for the identification and evaluation of various issues. As depicted in Figure 1, the Overall System Diagram illustrates the components of the system. Primarily, users can capture images of cucumber leaves using their mobile devices, and the system promptly analyzes them, providing real-time insights into common diseases such as Powdery mildew, downy mildew, and belly rot affecting cucumber leaves and fruits. Furthermore, by inputting the anticipated cultivation month,

International Current Journal of Engineering and Science (ICJES)
Volume 3, Issue 11, pp 9-15, November-2024
DOI: https://doi.org/10.47001/ICJES/2024.311002
ISSN (Online): 2583-7257

the application displays the expected harvest time and cucumber prices in the respective regions during that period.

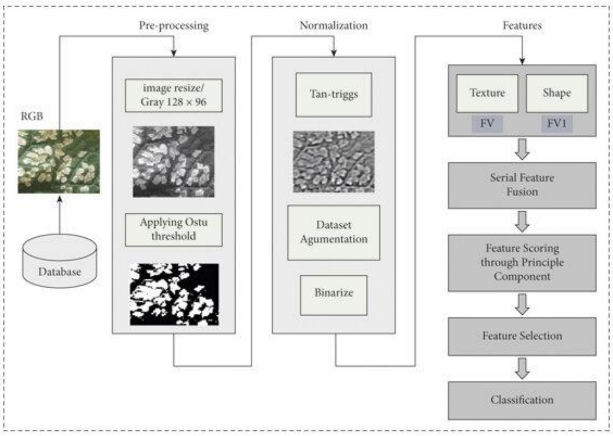

Figure 1: Overall System Diagram

a) Data Collection

In the process of gathering data for this research, a comprehensive collection of various types of images was undertaken. This encompassed acquiring high-quality images of healthy cucumber fruits, cucumber fruits affected by belly rot, as well as images depicting healthy cucumber leaves and different types of leaf abnormalities. Additionally, images portraying cucumber leaves afflicted with powdery mildew and downy mildew were included in the dataset. Furthermore, images capturing various cucumber pests were obtained. Moreover, data pertaining to monthly average retail prices, monthly average wholesale prices, Sri Lankan weather data, soil data, and other relevant parameters were meticulously compiled.

b) Methods

In this research, data augmentation serves as a pivotal strategy to enhance the training of machine learning (ML) models or transfer learning models by supplementing the process of cucumber pests' identification, have opted to train a modified YOLOv5 model and subsequently evaluate its performance using the validation set. To optimize the model's performance, we have fine-tuned hyperparameters such as learning rate, batch size, and number of epochs. These adjustments are aimed at maximizing the model's effectiveness in accurately identifying pests in cucumber leaves. The evaluation of the trained model's performance will be based on key metrics including mean average precision (mAP), precision, and recall. These metrics serve as indicators of the model's predictability, correctness, and comprehensiveness in pest detection tasks. To further enhance detection accuracy and focus on relevant features, we have integrated the Convolutional Block Attention Module (CBAM) into the updated YOLOv7 model. This addition aims to improve the model's ability to highlight important spatial and channel-wise features, thereby enhancing its overall performance. Furthermore, we will assess the trained model's capacity for generalization by testing it on data that it hasn't been exposed to, utilizing the testing set. This step ensures that the model can effectively detect pests in cucumber leaves across a variety of scenarios. Upon satisfactory validation of the model's performance, we will proceed to deploy it for real-time pest detection in cucumber leaves. This practical application will demonstrate the model's efficacy in real-world scenarios and validate its utility in agricultural pest management.

In analyzing the cucumber marketing performance Employs various data processing and analysis techniques to forecast cucumber prices and optimize cultivation decisions. Initially, machine learning models are loaded using "joblib" facilitating predictions based on historical data. "Pandas" is utilized to manage and manipulate datasets, incorporating datetime for temporal analysis. Functions are defined to predict prices for upcoming months and determine optimal cultivation start dates. Additionally, the script calculates potential incomes based on user-input parameters such as expected harvest amount and retail ratio.

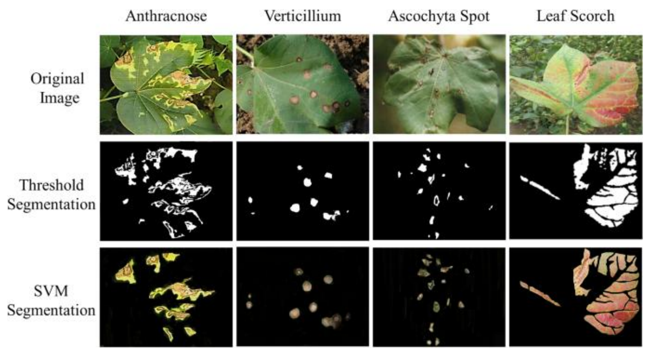

Figure 2: Identification of leaf deceases using image processing

These provide the techniques and actionable insights into price trends and cultivation strategies and aid decision-making processes for cucumber farmers. Predicting the harvesting duration of cucumbers and suggesting neighboring locations for selling involves data collection on cucumber growth stages and historical harvest durations. Statistical analysis and machine learning algorithms, such as regression models or decision trees, are then applied to predict harvesting durations based on factors like climate, soil conditions, and cultivation practices. Additionally, spatial analysis techniques, like geographic information systems (GIS), are utilized to identify neighboring locations with suitable market conditions and demand for cucumber sales. The methodology also includes

International Current Journal of Engineering and Science (ICJES)
Volume 3, Issue 11, pp 9-15, November-2024
DOI: https://doi.org/10.47001/ICJES/2024.311002
ISSN (Online): 2583-7257

validation through field trials to refine predictions and recommendations.

In analyzing the cucumber marketing performance data collection and processing techniques using a Python script for web scraping. The script utilizes the Requests library to fetch HTML content and 'BeautifulSoup' to parse and extract relevant data. Data cleaning procedures are implemented to format the extracted data, followed by organization into a Pandas Data Frame. The resulting dataset is saved to a CSV file for further analysis.

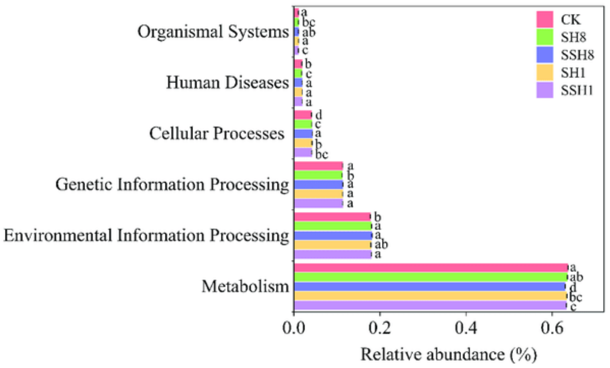

Figure 3: Monthly Average Wholesale Prices of cucumber

Training detection models involve initial data preprocessing to handle missing values and normalize features. Evaluation metrics such as train and test accuracies are calculated for both models. In the SVR model, a train accuracy of 53.18% and a test accuracy of 43.87% were achieved, based on these results, the Random Forest model is selected as the best architecture for its higher accuracy and generalization performance. Figure 4 shows Residuals for SVR model and Random Forest model.

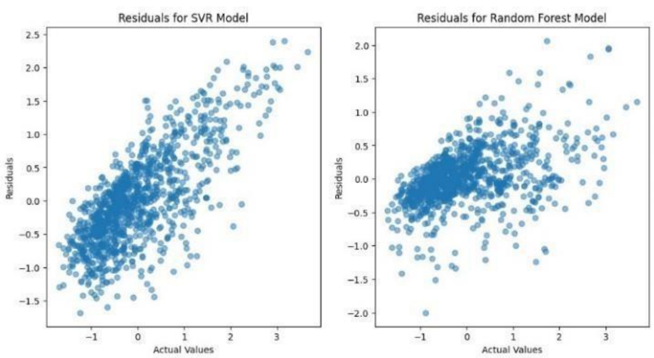

Figure 4: Residuals for SVR model and Random Forest model

This study revolves around the utilization of the Neural Prophet algorithm to develop a predictive model for weather parameters crucial to cucumber cultivation in Sri Lanka. Leveraging a dataset spanning from 2010 to 2023, consisting of 120,000 data points, the algorithm is trained to forecast weather conditions with a remarkable accuracy of 90%. This comprehensive analysis encompasses various weather parameters essential for optimal cucumber growth, including temperature, humidity, precipitation, and sunlight exposure. By employing advanced machine learning techniques and extensive historical weather data, this research aims to empower farmers with actionable insights for improved cultivation practices and enhanced crop yield in the dynamic Sri Lankan climate The dataset is meticulously curated to include thirteen attributes such as location, longitude, latitude, temperature, humidity, and area sown.

In the realm of land recommendation utilizing soil fertility and weather forecasting, IoT systems prove invaluable for gathering real-time data through sensors. This data is then leveraged by trained deep learning algorithms such as Artificial Neural Networks for predictive analysis. This predictive analysis aids in determining the most suitable crop for cultivation in each area. The Horticultural Crops Research and Development Institute (HORDI). It encompasses various parameters pertinent to crop cultivation.

Crops considered within this analysis include cucumbers. The preprocessing phase involves cleansing and organizing the dataset to ensure its suitability for training the deep neural network. In the IoT design, sensors are strategically deployed to collect real-time data on environmental conditions crucial for crop growth. This data is then fed into the trained deep neural network, which utilizes sophisticated algorithms to make informed predictions regarding crop suitability based on the gathered parameters.

IV. RESULTS AND DISCUSSION

The smartphone app that was created specifically for growing cucumbers is an all-inclusive feature set that provides growers with insightful information. To deliver precise market price predictions for cucumbers, the application makes use of advanced forecasting methods, such as linear Hodrick-Prescott filters and hybrid neural networks. This gives farmers the critical information they need to make smart decisions about when to plant and where to enter markets.

Moreover, the program uses cutting-edge picture recognition technology to identify pests and illnesses that affect cucumber fruit, providing a quick and accurate diagnostic tool. The method uses machine learning algorithms to quickly analyze photos of cucumber fruits, leaves, and general plant health. This helps farmers quickly identify and address possible problems. By managing pests and diseases proactively, crop resilience is increased, and yield is maximized.

International Current Journal of Engineering and Science (ICJES)
Volume 3, Issue 11, pp 9-15, November-2024
DOI: https://doi.org/10.47001/ICJES/2024.311002
ISSN (Online): 2583-7257

Apart from identifying diseases and pests, the mobile application also incorporates the ability to monitor soil moisture levels. Farmers may make precise irrigation plans by utilizing sensor data and real-time readings to acquire insights into the soil's moisture levels. This reduces the danger of illnesses and stress linked to water by providing cucumbers with an ideal growth environment while also conserving water resources.

The application's usefulness is not limited to elements related to crops; it also encompasses weather conditions. Through the provision of current weather forecasts customized to the farm's geographic location, farmers can plan and modify crop operations in response to predicted weather patterns. Farmers can predict and lessen the impact of unfavorable weather events because of this weather integration, which increases the overall resilience of cucumber agriculture.

To sum up, the comprehensive mobile application serves as a comprehensive tool for growers of cucumbers, effectively combining weather forecasts, disease and pest diagnosis, market data, and soil moisture monitoring. This comprehensive strategy helps to make well-informed decisions, use sustainable farming methods, and eventually increase cucumber crop productivity.

V. CONCLUSION

Our research concludes with significant findings that have implications for the cucumber industry. Using machine learning algorithms, in this research successfully identified common cucumber diseases such as downy mildew, powdery mildew, and belly rot with high accuracies, leveraging a dataset of 1300 healthy cucumber leaves, 3198 leaves with disease identification, and 3500 instances of belly rot. Additionally, the incorporation of soil prediction models based on soil fertility and weather predictions, supported by a dataset of 120,000 weather predictions and comprehensive soil health indicators, allowed for precise monitoring of disease causative agents. In terms of marketing analysis, we achieved a remarkable accuracy of 97% using a dataset of 3000 instances, providing valuable insights into consumer preferences and effective promotional strategies. These findings suggest that targeted interventions such as soil amendments, crop rotations, and disease-resistant varieties can mitigate disease pressure and enhance crop yields and quality. Furthermore, optimizing marketing strategies based on market segmentation insights can lead to improved performance and profitability in the cucumber industry. Moving forward, continued research in machine learning algorithms for disease identification, soil prediction models, and advanced marketing analysis techniques will be essential for sustainable cucumber production and industry success.

REFERENCES

[1] F. Bassanezi RB, Gas exchange and emission of chlorophyll fluorescence during the monocycle of rust, angular leaf spot and anthracnose on bean leaves as a function of their trophic characteristics, 2002.
[2] Sriboonchitta M.S., H. T. Nguyen, A. Wiboonpongse, J.Liu, "Modeling volatility and dependency of agricultural priceand production indices of Thailand: static versus time- varyingcopulas," International Journal of Approximate Reasoning.
[3] B. Oancea, C. Ciucu, "Time series forecasting using neuralnetworks," 2013.
[4] K. Alok Kumar, " “Plant Disease Detection using VGG16,” International Journal of Creative Research Thoughts, Volume 11, 1 January 2023.," [Online]. Available: https://ijcrt.org/papers/IJCRT2301347.pdf.
[5] Z. C. e. al, "Plant disease recognition model based on improved , vol. 12, p. 365, 2022.," [Online].
[6] B. T, Regulation of the production of stress ethylene and its significance.
[7] Caeneghem WV, Messens E, Lamber H, van Montagu M, van der Straeten D, "Presymptomatic visualization of plant–virus interactions by thermography," Nature Biotechnology.
[8] Boever F, van Montagu M, van der Straeten D, "Thermographic visualization of cell death in tobacco and Arabidopsis," Plant, Cell and Environment, 2001.
[9] Hagenbeek D, De Bruyne E, Valcke R, van der Straeten D, "Thermal and chlorophyll-fluorescence imaging distinguish plant–pathogen interactions at an early stage," Plant Cell Physiology, 2004.
[10] HORDI, "HORDI," [Online]. Available: https://doa.gov.lk/hordi-crop-cucumber/.
[11] S. R. Shampa Chakraborty, Health Beneficial Effects of Cucumber,” Studies on Cucumber (Cucumis sativus L., 2012.
[12] T. C. Michael S, Resistance to Belly Rot in Cucumber Identified through Field and Detached-fruit Evaluations.
[13] Z. Chen, Plant disease recognition model based on improved, 2022.
[14] M. Yang, Fruit target detection based on BCo-YOLOv5 model.
[15] Akshatkary H, Benhamou N, Brodeur J, "Ultrastructural and cytochemical investigations of the antagonistic effect of Verticillium lecanii on cucumber powdery mildew".

International Current Journal of Engineering and Science (ICJES)
Volume 3, Issue 11, pp 9-15, November-2024
DOI: https://doi.org/10.47001/ICJES/2024.311002
ISSN (Online): 2583-7257

Citation of this Article:

Catherine Mary, John Okello Okuto, & Simon Osoro. (2024). Prediction of the Diseases in Cucumis Sativus and Soil Health Using Machine Learning Algorithms. *International Current Journal of Engineering and Science - ICJES*, 3(11), 9-15. Article DOI: https://doi.org/10.47001/ICJES/2024.311002

****