Smart Greenhouses Decision Support System for Tomato Cultivation

24-25J-064

Project Proposal Report

Najas MNM

BSc (Hons) in Information Technology Specialized in Information Technology

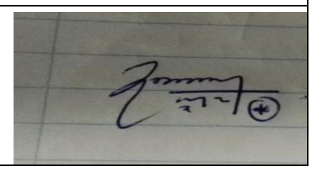
Department of Information Technology

Sri Lanka Institute of Information Technology Sri Lanka

August – 2024

DECLARATION

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

Name	**Student ID**	**Signature**
Najas MNM.	IT21186592	

Signature of the Supervisor: Date
(Mrs. Geethanjali Wimalaratne)

……………………………….. ……………………

ABSTRACT

Tomato plants are very vulnerable to different diseases, and if they are affected it will highly affect the yields and the quality of the crops. In the recent past, traditional disease diagnostics involve inspection which is a tiresome process and carries along with it errors from man. As a decision of this research, the following detailed decision support system is suggested where image analysis and machine learning would help detect the diseases that affect the tomato plants and suggest the right treatment for the same. To that end, they employ annotated images of tomato plants with a Convolutional Neural Network (CNN) model for the classification and identification of common tomato diseases. These are image resizing, normalization, and augmentation which are fed into the system to improve the model and the accuracy of the system. After the detection of the disease, the system produces treatment suggestions that are as appropriate to the disease severity level and type. The proposed system is to be an improved version of the present manual inspection with less crop damage and over-application of pesticides. The performance of the model is very assessed and measured using parameters such as accuracy, precision, and recollection. It was believed that the integration of this system in smart greenhouses would make disease management practices improve hence enhancing productivity and sustainability in the practice of tomato farming.

Keywords: Tomato Disease Detection, Machine Learning, Convolutional Neural Networks (CNN), Image Analysis, Disease Classification, Treatment Recommendation

Contents

LIST OF FIGURES

1. INTRODUCTION

Tomato plants are one of the most important crops worldwide, yet they are very fragile and vulnerable to several diseases that have magnified effects on the yields as well as quality. Existing approaches for disease identification involve a visual examination, which takes a lot of time, requires a lot of physical effort, and is rather inexact. Such limitations often cause delayed action and thus many crops end up severely affected hence leading to higher application of pesticides.

Fortunately, developments in recent innovative tools and approaches such as image analysis and machine learning significantly provide the solutions to these challenges. The proposed research is under the broad theme of a decision support system in which CNNs will be used to diagnose diseases affecting tomatoes and suggest the right treatment. To reduce the disease diagnosis burdens of farmers and prevent crop loss by giving appropriate and timely detection of diseases, this system has been designed.

The images will be collected from the sites displaying tomato plants' photographs with annotations; The operation of preprocessing will be applied to)prepare the images for entering the system through the scale of resizing, normalization, and augmentation to improve such parameters as accuracy and loss. The CNN model will then learn how to look at those images for specific diseases with an indication of the severity of the disease. From these results, the system will recommend the best treatments to be applied to the diseases enabling farmers to make the right decisions on how to manage the diseases.

Adding this system to smart greenhouses enables the monitoring of the greenhouses in real-time and also alerts of any problem encountered. The goal of this research is to alleviate the deficiencies of conventional practices to enhance disease control and thus promote more favorable conditions for tomato production.

Finally, this project presents a multi-faceted system where disease identification and treatment suggestions would be done by a machine learning system to revolutionize the tomato farming process. fields in tomato cultivation.

2. BACKGROUND & LITERATURE SURVEY

2.1 Background

Tomato plants are very prone to diseases that result in low yields and poor quality of the crop. The conventional system of detection is done through visual examination which is costly in time, labor, and manpower and often gives inaccurate results hence early interventions are not done thus increasing crop losses. The opportunities for the improvement of speed and accuracy in the detection of diseases are fair due to the progress made in the area of image analysis and machine learning, especially Convolutional Neural Networks (CNNs). This research will focus on designing an automated system that uses CNNs to diagnose tomato diseases, as well as suggest appropriate treatment, thus improving the disease control measures and reducing the reliance on foliar checks.

Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, we train a deep convolutional neural network to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale(‘Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016)..pdf’, no date).

Plants are crucial for human survival, contributing to food production, water conservation, and maintaining atmospheric balance through photosynthesis. However, plant diseases can significantly impact the quality and quantity of agricultural products, leading to economic losses and food insecurity. Historical instances, such as the 1943 rice helminthosporiose outbreak in northeastern India, which led to significant food shortages and deaths, underscore the critical need for effective plant disease management(Zhang, Shang and Wang, 2015).

Approximately 80% to 90% of plant diseases manifest on leaves, making leaf-based disease recognition a key area of research. Although numerous methods exist for leaf disease recognition, challenges remain due to the complexity of leaf color and shape(Zhang, Shang and Wang, 2015).

Digital image processing techniques are increasingly being utilized in agricultural applications, particularly for the detection and diagnosis of plant diseases. This is a critical area of research as it helps in accurately assessing disease severity, which is essential for effective crop management. The detection of plant disease symptoms on leaves has been a focus of research for several decades, as noted by studies dating back to the 1980s(Barbedo, 2016).

One of the primary challenges in detecting plant diseases through image processing is the accurate segmentation of diseased and healthy tissue. Traditional RGB color space often fails to provide sufficient information for precise segmentation, prompting researchers to explore alternative color spaces such as HSV (Hue-Saturation-Value) and L*ab** (Lightness and two color channels representing opposing color dimensions)(Barbedo, 2016).

The paper by Konstantinos P. Ferentinos focuses on the development of deep learning models, particularly Convolutional Neural Networks (CNNs), for the detection and diagnosis of plant diseases. The complexity of diagnosing plant diseases based on visual symptoms, especially through the optical observation of plant leaves, is highlighted as a significant challenge. The need for an automated system that could aid agronomists and even farmers in diagnosing plant diseases is presented as a critical motivation for this research(Ferentinos, 2018)

2.2 Literature Survey

Early CNNs: CNNs have been pivotal in advancing the field of computer vision. Early architectures like LeNet and AlexNet laid the groundwork for deep learning by demonstrating the effectiveness of convolutional layers for image recognition tasks. AlexNet, in particular, brought CNNs into the mainstream by winning the ILSVRC 2012 with a significant margin, proving that deep networks could be trained efficiently with the help of GPUs(Zeng *et al.*, 2016).

Traditional Image Processing: Before the advent of deep learning, plant disease detection relied on image processing techniques that involved feature extraction and classification using algorithms like Support Vector Machines (SVMs). These methods were limited by their inability to generalize well across different environmental conditions and required significant manual effort in feature design(Saleem, Potgieter and Arif, 2019).

Inception and ResNet: Architectures like Inception (used in GoogLeNet) and ResNet introduced deeper networks that could learn more complex features while mitigating issues like vanishing gradients. These models achieved state-of-the-art performance in image classification tasks, including plant disease detection(Saleem, Potgieter and Arif, 2019).

3. RESEARCH GAP

Diagnosis and control of diseases in tomatoes is very important to achieve maximum yields and quality of the produce. As for the traditional methods, they involve manual inspection which is very slow, error-prone, and demands huge experience. Some new approaches to this problem have emerged in the recent past with the help of the latest developments in the fields of machine learning and image analysis. Nevertheless, there are still some open questions and challenges that can be considered in the development of the current technologies and their application in practice for agriculture.

The study by Mohanty et al. (2016) focuses on the application of deep learning techniques, particularly Convolutional Neural Networks (CNNs), for the detection of plant diseases using image-based data. While their research demonstrates the effectiveness of CNNs in accurately identifying plant diseases, it predominantly addresses the detection aspect without extending into the practical implementation of disease management or treatment recommendations based on the detected diseases. Furthermore, the study is conducted in a controlled environment with a limited dataset, which may not fully capture the variability of real-world conditions, such as different environmental factors or varying disease stages(‘Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016)..pdf’, no date).

The study conducted by Zhang et al. (2015) focuses on the recognition of plant diseases through image-processing techniques applied to plant leaf images. The research introduces a method that involves leaf image segmentation, feature extraction, and classification using a K-nearest-neighbor (KNN) classifier. While the approach shows effectiveness in recognizing plant diseases, particularly in controlled environments, several gaps remain that are critical for further exploration(Zhang, Shang and Wang, 2015).

Figure 1

The study by Too et al. (2019) focuses on the fine-tuning and comparative evaluation of various deep learning models, such as VGG 16, Inception V4, ResNet, and DenseNets, for the task of plant disease identification. While the paper successfully demonstrates the effectiveness of these models in identifying diseases using images of plant leaves, it highlights several areas that could be expanded upon or improved in future research(Too *et al.*, 2019).

The study by Singh and Misra (2017) explores the use of image segmentation techniques combined with soft computing methods, such as genetic algorithms, for the detection and classification of plant leaf diseases. While the research makes significant contributions to the field by proposing an automatic detection system that leverages advanced image processing techniques, several gaps remain that could be addressed in future research(Singh and Misra, 2017).

Existing Products / Research	**Predictive Analytics**				
	Image Analysis Tools	Manual Inspection	Disease Specific Detection	Automated Treatment Recommendation	Predictive Analytics
Current Methods	✗	✓	✓	✗	✗
Basic Automated Tools	✓	✓	✗	✗	✗
Proposed System	✓	✗	✓	✓	✓

Figure 1

3.1 Proposed System

The proposed system is a more intelligent decision support system developed to improve the identification of diseases affecting tomatoes and the correct treatment to be given. This system employs a Convolutional Neural Network (CNN) that has been trained on a large set of tomato plant images that have been appropriately labeled. To improve the performance of the model, the images go through some of the preprocessing steps such as resizing, normalization, and augmentation.

The CNN model is incorporated into a surveillance system that constantly searches for tomato plants in greenhouses or the open farmland. When the system identifies a specific disease, then it provides an immediate categorization of the type and the extent of the infection; it also provides treatment recommendations. These recommendations consider the peculiarities of the identified disease, to avoid excessive use of pesticides and to improve the crop.

Furthermore, the system has an interactive part that can be accessed through a WebApp interface through which farmers can check the state of plant health, receive notifications, and take swift action. This system uses a combination of advanced image analysis with machine learning to offer a more efficient solution, which can be used on a large scale and can be applied in the modern agricultural industry since it eliminates the inefficiencies of manual inspections.

4. RESEARCH PROBLEM

Tomato plants are one of the most important crops in the world, yet they are prone to several diseases, which is a potential danger to food security and production. Usually, the recognition of diseases involves physical examination, which is very slow, requires many people, and is highly susceptible to errors. Such limitations lead to early diagnosis and management that escalate crop losses, and excessive application of chemical pesticides, which affects the environment and human beings.

While contemporary peri-sylvan frameworks are largely dependent on agricultural technology, the current intricate arrangements are mainly related to disease identification rather than encompassing treatment strategies. Furthermore, most of the current models are derived using data collected in a controlled environment, and may not generalize well to the diverse, practical agricultural scenarios. These systems also lack friendly interfaces, which can restrict the application of various technologies by users who are not IT-savvy, for instance, farmers.

The research problem, therefore, focuses on designing a more effective, efficient, and easy-to-use system for the detection of tomato diseases using machine learning techniques like CNN's and makes a corresponding recommendation on the type of treatment needed. Solving this issue is critical for advancing crop management techniques, decreasing the reliance on visual observation, and supporting sustainable farming.

How can we develop a decision support system for tomato greenhouses that detects diseases and recommends appropriate treatments based on image analysis and machine learning?

The use of a decision support system in the context of tomato greenhouses requires the use of image processing and machine vision to ensure the accurate identification of diseases and the right course of action to take. The system allows various diseases and their severity to be detected and diagnosed by training Convolutional Neural Networks (CNNs) for a wide range of annotated tomato plant images. Real-time monitoring integration enables constant screening of plants and the system offers a recommended treatment depending on the noted disease. This approach not only improves disease management practices but also minimizes the dependency on manual inspection for early and accurate interventions in the greenhous

5. OBJECTIVES

The research aims to develop a comprehensive decision support system that uses Convolutional Neural Networks (CNNs) to detect tomato diseases and recommend appropriate treatments.

5.1 Main Objective

Develop an advanced system that accurately detects tomato diseases and provides tailored treatment recommendations using image analysis and machine learning, ultimately improving crop health and maximizing yield.

5.2 Specific Objectives

1. Select Data from Kaggle:.

- The first step is to gather a dataset from Kaggle, which is one of the most famous platforms that contains many datasets available for the public. For this research, the focus is on the identification of a dataset that would include images of tomato plants infected by various diseases. This dataset is used to train and test the machine learning model.

2. Label the Images with Corresponding Disease Types:

- After the dataset is chosen, every picture in the dataset has to be properly labeled with the right disease type. Labeling is important as it offers the needed reference information that the model will be trained with during the model training process. The labels assist the model in determining which features in the images are related to specific diseases.

3. Preprocess the Images (Resize, Normalize, Augment) for Model Training:

- Image preprocessing is also very important in the preparation process to enhance the efficiency of training the model.
- **Resize:** Every image is also scaled to a fixed size so the model takes inputs of similar size an shape.
- **Normalize:** The pixel values of the images are normalized to a standard scale (most of the time 0-1) to better train the model and acquire a faster and better convergence.
- **Augment:** data augmentation, the images are transformed in a random way such that the set of training images meanly and then rotated, flipped, zoomed, and so on. On the same note, this makes the model generalize well and increases its ability to perform well on unknown data.

4. Develop a CNN Model Architecture Suitable for Image Classification:

- This step involves creating an architecture for CNN that would be suitable for the current task of image classification. The CNN has layers that identify features of images that are useful in differentiating between healthy plants and diseased ones; these are the edges, textures, and patterns.

5. Train the CNN Model on the Labeled Dataset and Evaluate its Performance:

- However, the CNN architecture is now established & next, the model should be trained by using the labeled data. In the course of training, the model acquires the capability to map certain features in the images with the disease labels. The trained model can be assessed in terms of accuracy, precision, and recall among others, to judge the ability of the model to categorize new images which it was not trained to distinguish. The evaluation assists in improving the model since the model is modified for higher efficiency and real-world use in subsequent iterations.

6. METHODOLOGY

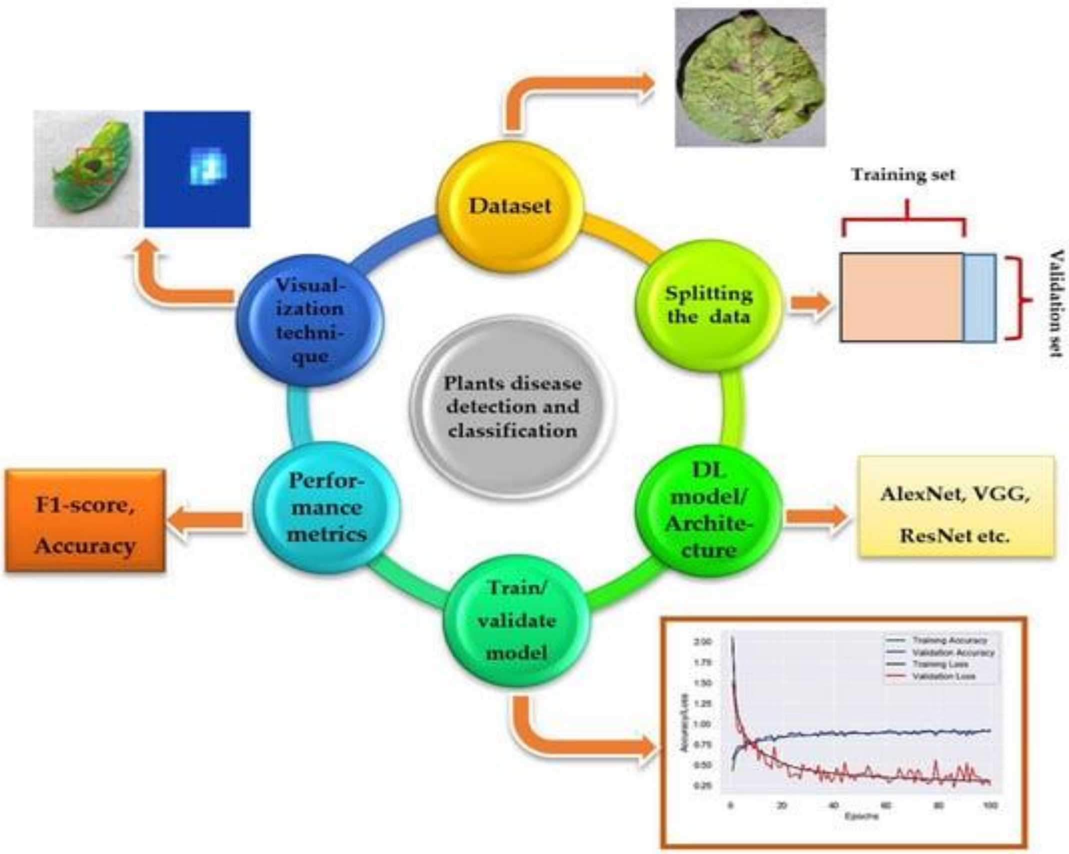


Figure 2(Saleem, Potgieter and Arif, 2019)

The methodology involves a systematic approach to developing a robust system for detecting tomato diseases and recommending treatments. First, images of tomato plants are selected and annotated from Kaggle to create a labeled dataset. The images then undergo preprocessing, including resizing, normalization, and augmentation, to optimize them for model training. A Convolutional Neural Network (CNN) is designed and trained on this processed dataset to classify different diseases. The model's performance is evaluated using metrics such as accuracy, precision, and recall. Finally, the system is implemented to provide tailored treatment recommendations based on the detected diseases.

- **Data Collection and Labeling**: Select and annotate images from Kaggle.
- **Image Preprocessing**: Resize, normalize, and augment images.
- **CNN Model Development**: Design and train a CNN model for disease classification.
- **Performance Evaluation**: Evaluate the model's performance using metrics such as accuracy, precision, and recall.
- **Implementation of Recommendation System**: Provide treatment recommendations based on disease detection.

Technologies and Techniques

- Python, TensorFlow, Keras.
- Image preprocessing techniques.
- CNN model design and training.
- Data visualization tools.

6.1 Software Solution

The software solution under consideration is an IDS which is aimed to facilitate the detection of diseases in tomatoes and offer accurate recommendations on their treatment. The system is based on the CNN model that is trained on the selected dataset containing the images of tomato plants with annotations obtained from Kaggle. These images are preprocessed to fit the model by resizing, normalizing, and augmenting the images for better performance.

supplying live data to the CNN model to diagnose the disease on the spot. After a disease is diagnosed, the system assesses the nature and intensity of the infection using a predefined knowledge base to produce treatment plans.

It has a web and mobile application where farmers can check the health of the plants, notifications, and comprehensive treatment plans. The interface of the system has a clear layout and is developed in a way so that it can be used by people who do not have a technical background. Furthermore, the system is easily scalable to accommodate the size of greenhouses or larger farming operations. This solution covers multiple aspects of crop production with the primary goal of minimizing the reliance on manual inspection, improving crop management, and advancing the principles of sustainable agriculture

6.1.1 Requirements Gathering

In this research, the functional and the non-functional requirements are established for the requirements gathering. **Functional requirements** include obtaining annotated images from Kaggle, building a powerful CNN for disease diagnosis, and creating a mechanism that prescribes treatment depending on disease severity. **Non-functional requirements** are concerned with how the system will work effectively and meet the quality attributes such as scalability and reliability among others. The interface should be user-friendly and not require much technical skill; it should have real-time monitoring and alerting. Further, the system may need to be flexible to the prevailing conditions of agriculture and deployable to small and big-scale farming.

6.1.2 Testing and Quality Assurance

The steps for testing and quality assurance (QA) for this research entail checks and balances at every level. A validation dataset is also used on the CNN model to assess the accuracy, precision, recall, and F1 score of the model. Cross-validation is used to check the model's stability and performance across different subsets of data. After the deployment, the functionality of the system is evaluated on real-life data to determine its ability to diagnose diseases and suggest correct treatments. When the design is complete, user acceptance testing (UAT) checks whether the interface would be easily usable and satisfactory to the users. To resolve any problems, maintenance is provided, and constant modifications are made to make the system more efficient and flexible in numerous agricultural contexts.

7. PROJECT REQUIREMENTS

7.1 Functional Requirements

1. Data Collection and Annotation –The system should be able to download and load high-quality labeled images of tomato plants from sites such as Kaggle. Every picture should be properly marked with the type of disease, which will be used as a ground truth for training.
2. Image Preprocessing – Some of the elements to be incorporated in the system should be to crop the images to ensure that they are all of the same size, change the pixel values in the images for the ease of training the model, and also perform data augmentation to ensure that the images used for training are as diverse as possible. This step is needed to get the images in the format that can be properly classified by the CNN model.
3. CNN Model Development- The system is required to grow and build a CNN model specific to distinguishing this and that type of tomato disease. The model should be able to capture the images that have gone through the preprocessing and give a probable diagnosis regarding diseases seen based on the features learned during training.
4. Treatment Recommendation System – After a disease has been diagnosed; the system should then give recommendations on the treatment best suited depending on the type and grade of the disease. This feature is crucial in reaching essentially nurses and farmers and assisting them in applying the most suitable treatments immediately.
5. User Interface – It should have an easy-to-use customer interface that is freely accessible via web applications.

7.2 User Requirements

1. Ease of Use The system should have an easily understandable and simple user interface that can be mastered by both a technical and a non-technical person. It should be simple enough to understand that the farmers and agricultural workers do not need to undergo a full-blown training regime to use the system.
2. Accessibility - The system should be clickable on different devices such as mobile, tablets, and on the computer. This makes it possible for the users to keep track of plant health and be able to get notifications either when they are in the field, in the greenhouse, or even when they are at the office.
3. Real-Time Monitoring and Alerts - The users need the system to monitor the tomato plants in real-time and notify them the moment a disease is identified. Users of the system should be alerted in real-time in the instance that there is a problem to enable them to prevent crop loss.

7.3 System Requirements

System requirements define the hardware and software components necessary to implement the project. They include,

7.3.1 Software Requirements

1. Python -for backend development and machine learning model implementation.
2. Flask - for web server framework to build the web application.
3. Arduino IDE - for programming the ESP32 microcontroller.
4. IntelliJ IDEA - for software development.
5. Firebase - for cloud-based database management.
6. Web Technologies - including HTML, CSS, and JavaScript for the UIs.

7.4 Non-functional Requirements

1. Performance - For it to achieve its objective, the system should be able to process and analyze images as this gives real-time results of the diseases to be treated and the treatment to be administered.
2. Reliability - The system should thus be designed to be very efficient; the amount of time that should be taken to maintain the system should be very limited. It must be functional as intended under varying environmental conditions and contain processes of updating and maintenance to make it work with maximum efficiency and accuracy in the future.
3. Scalability - The system has to be able to provide support capacity since the number of users would be unpredictable, and the size of the farms would vary from hobby farms to large commercial farms. It should respond to increased data size, processing demands, and users in a fast manner without compromising on performance.
4. Security - Security requirements of the system include particularly the protection of users' data especially if it is stored in the cloud. To avoid exposure of sensitive information to unauthorized persons or organizations and cyber thefts, their system must incorporate; strong authentication methods, data encryption and security checks, and balancing.

8. BUDGET AND BUDGET JUSTIFICATION

8.1 Budget Justification

- Other Charges - All required software tools are open-source, minimizing the cloud hosting is necessary for the web-based application, and a contingency budget is set aside for any unexpected costs.

9. REFERENCES

Barbedo, J.G.A. (2016) *A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing*, *Tropical Plant Pathology*. Available at: https://doi.org/10.1007/s40858-016-0090-8.

Ferentinos, K.P. (2018) ‘Deep learning models for plant disease detection and diagnosis’, *Computers and Electronics in Agriculture*, 145, pp. 311–318. Available at: https://doi.org/10.1016/J.COMPAG.2018.01.009.

‘Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016)..pdf’ (no date).

Saleem, M.H., Potgieter, J. and Arif, K.M. (2019) ‘Plant Disease Detection and Classification by Deep Learning’, *Plants 2019, Vol. 8, Page 468*, 8(11), p. 468. Available at: https://doi.org/10.3390/PLANTS8110468.

Singh, V. and Misra, A.K. (2017) ‘Detection of plant leaf diseases using image segmentation and soft computing techniques’, *Information Processing in Agriculture*, 4(1), pp. 41–49. Available at: https://doi.org/10.1016/j.inpa.2016.10.005.

Too, E.C. *et al.* (2019) ‘A comparative study of fine-tuning deep learning models for plant disease identification’, *Computers and Electronics in Agriculture*, 161(October 2017), pp. 272–279. Available at: https://doi.org/10.1016/j.compag.2018.03.032.

Zeng, G. *et al.* (2016) ‘Preparation of novel high copper ions removal membranes by embedding organosilane-functionalized multi-walled carbon nanotube’, *Journal of Chemical Technology and Biotechnology*, 91(8), pp. 2322–2330. Available at: https://doi.org/10.1002/jctb.4820.

Zhang, S.W., Shang, Y.J. and Wang, L. (2015) ‘Plant disease recognition based on plant leaf image’, *Journal of Animal and Plant Sciences*, 25(3), pp. 42–45.