CCDC: Centre for Crop Disease Control

Project Proposal for HackWknd – Agrofest 2019

Sophia Chulif, Heng Kiat Jing, Danish Ezwan, Chang Yang Loong

Neuon AI

1.0 Project Objective

The objective of this project is to adopt computer vision and machine learning in crop disease detection. The aim is to reduce crop loss, hence increase quality crop yield by leveraging efficient crop disease recognition.

2.0 Introduction

Crop disease identification plays an essential role in the agricultural production sector. In order to increase crop production and reduce loss caused by crop diseases, proper diagnosis is required before necessary control and mitigation can be suggested. Proper diagnosis of crop disease is extremely important not only in curative measures but in preventing the spread of disease to other crops as well as recurring outbreaks in the future. One of the main challenges faced by farmers is to identify the type of diseases contracted by their crops. Without accurate identification of the respective disease, efforts and resources spent on control measures can be wasted.

At present, crop disease detection is done with direct and indirect methods. Direct disease detection consists of molecular and serological methods to identify the disease-causing pathogens such as fungi, viruses and bacteria. On the other hand, indirect detection identifies the crop diseases through observations of the crop morphological changes or compound released on its parts. This may include signs of abnormality in any of its parts such as the leaves, fruits, stems, shoots, roots or the whole plant itself.

In recent years, there has been an increase of study on the effects of climate change on diseases which cause crop loss. As extreme weather such as drought, heat waves and sudden rainfall patterns are becoming more common, findings have predicted that plant diseases are likely to become more frequent and severe as it allows some pathogens to thrive and spread. Due to international crop trade, outbreaks of crop diseases will likely spread into new areas and impose a threat on food security globally.

One of the common existing problems faced by farmers in detecting crop disease is the lack or unavailability of human expertise e.g. plant pathologist in the farms. This is especially in the case of rural farming as resources are limited and the agricultural extension support is costly. Smallholders as such therefore suffer a huge loss when crops are affected and crop yield is low.

With regards to this, the rapid technological change happening has open doors to address epidemic diseases such as mentioned. Computer vision and artificial intelligence has been in the works of creating automated systems to diagnose crop diseases. The use of machine learning, more specifically deep-learning convolutional neural networks (CNN) has been applied to classify plant diseases with a high degree of accuracy. For instance, the mobile application namely, “Nuru” that was built by PlantVillage (a research and development unit of Penn State University). It was made to assist smallholder farmers in diagnosing multiple plant diseases which include the fall armyworm infection, potato disease and wheat disease in Cassava. This AI assistant is based on extensive research comparing the accuracy of machine learning models to human experts and extension work.

Crop disease recognition by human perception alone may be challenging especially to those without the expertise. Today, the accuracy of modern computers with CNN has shown to match and even exceed the accuracy of human vision. This proves its significant potential in crop disease recognition. With the help of artificial intelligence, early detection of crop diseases can be carried out, allowing rapid response and control over the disease. The inability to detect and manage crop disease will incur a huge loss on farmers, especially smallholder farmers whose income depend solely on healthy crop production. Since the large majority of crop producers in Sarawak are made up of smallholder farmers, it is vital that we aid them in producing the best quality of crops.

3.0 Problem statement

1. How can we create a sustainable farming solution in preparation for the 2050 food supply deficiency?
2. How can we empower rural farmers to utilize data mining to improve their farm's efficiency while at the same time lowering their farm operating cost?

4.0 Proposed solution

For this project, we propose a crop disease recognition system using computer vision and machine learning that allows farmers to diagnose their crops efficiently. This application can be highly useful for small-scale farmers who do not have the research infrastructure or agricultural extension systems for support. It can be an initiative to encourage local farmers to make use of current AI technologies to reduce their operating cost. At the same time, a collaborative platform can be introduced where farmers and researchers are able to work together to identify crop diseases, employ organic solutions and mitigate the occurrence of crop loss.

By helping farmers to diagnose crop disease and suggest control measures, crop loss can be reduced enabling the increase of healthier crop yield thereby accommodating the crop demand locally and internationally. Not only would this application benefit individual farmers but also farmers as a whole who would experience similar crop disease. The crop-infected images captured by the farmers serve as a data warehouse for data mining which allows the improvement of detection and discovery of new crop diseases, specifically within the local community which may be unknown in other countries. In a long run, the data and information collected by its users can be sustained for research purposes and future generations.

5.0 Methodology

5.1 Acquire and process data (images) for training

In building a machine learning model for the detection of crop disease, sample images of the infected crop have to be first fed into the convolutional neural network (in other words, the computer or machine) for training. The neural network requires a set of training images to learn and recognise from so that it can make predictions on the respective disease when tested or "seen". As a start, the sample images can be acquired from publicly available datasets or images provided by collaborating farmers or researches.

5.2 Train network

The Tensorflow-Slim image classification model library will be used in the implementation of the network. It is open-sourced and provides a high-level API for defining, training and evaluating complex models. The framework of the network applies supervised-learning algorithms whereby the network will learn to classify the test images based on the given training images.

5.3 Test accuracy of model

After training the desired network for the crop disease detection model, the accuracy of the model has to be tested out to ensure its liability. A certain degree of accuracy has to be achieved before deploying the model. If it does not reach the desired accuracy, the network is retrained.

5.4 Deploy disease detection model

Once the model is ready for deployment, it is packaged and deployed into the Android mobile and Windows PC platform for use.

5.5 Continuously update and train network

The network needs to be continuously updated and trained with new training data in order to achieve significant results corresponding to real-world problems. With the help of researchers or plant experts, the network can be further improved to achieve better prediction of crop diseases. This will be a continual research for crop disease identification among farmers and experts alike.

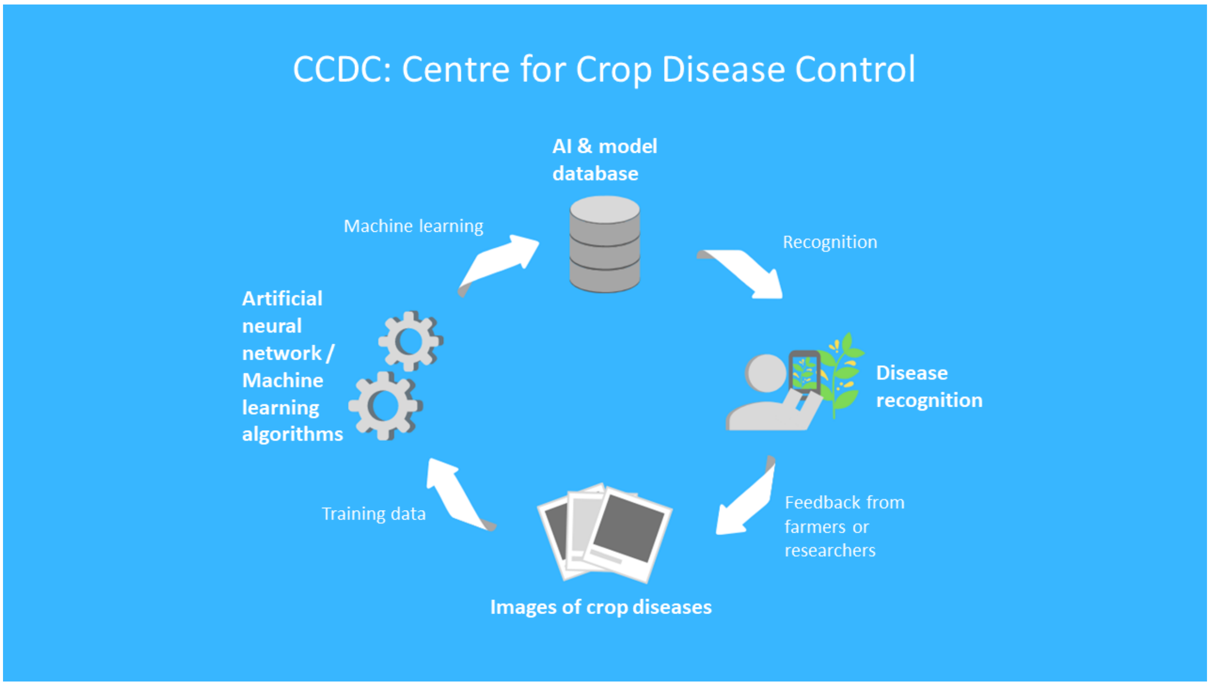


Figure 1: Workflow of proposed solution.

Figure 1 illustrates the workflow of our proposed solution. The farmers and users would participate in data collection via App created by CCDC which is complete with labelling UI for meta information. These data will later be processed by the expert to either label or validate the images. Then, these annotated images are used to train and update an existing model of crop disease. Finally, the AI model is deployed/updated in the form of App for the farmer to recognize various diseases. The cycle continues as the farmers continuously submit their detected data and disease queries via the same App to enrich the crop disease data to produce a better AI model.

6.0 Proof-of-concept / Prototype:

6.1 Overview

In realising our idea, our team has created a proof-of-concept project / prototype which focuses on the automated detection of tomato plant diseases. The project utilises the advantages of convolutional neural networks in learning significant visual features which are unique to a particular tomato plant disease. With the aid of computer vision and machine learning, we hope that mobile devices or smart phones can come in handy in helping smallholder farmers as well as large-scaled farms to efficiently identify the diseases that infect their crops.

6.2 Data Collection

The training data used for creating our prototype is sourced from PlantVillage and IPM Images. The PlantVillage project is a non-profit initiative that gathers leaf-scan images infected with crop diseases which aims to help smallholder farmers increase their crop yield. The dataset contains 54,305 colour images of diseased and healthy plant leaves from 38 categories. From the dataset, we only extracted images of healthy and infected tomato plant leaf images which fits our case. This reduces the dataset to 18,160 images with 10 categories.

Since the PlantVillage dataset only consists of leaf scans, another 436 images are sourced from IPM Images to diversify the types of diseased tomato plant images used in training the networks. These include the plant stems, fruits, buds, entire view, microscopic view and far view. The images are assigned with the same labels with the training images from PlantVillage. However, there are no images of tomato plant infected with target spot from IPM Images. Although the number of tomato plant images from IPM Images is incomparable to the PlantVillage dataset, our preliminary trainings show that adding tomato plant disease images from IPM Images during training will allow the networks to generalise better and predict better on both PlantVillage and IPM dataset images. This prove that those images do provide important information on additional visual features that are caused by crop diseases other than those features that exist on tomato leaves.

The 10 categories output by the tomato plant disease classification model are:

a. Bacterial Spot
b. Early Blight
c. Late Blight
d. Leaf Mold
e. Septoria Leaf Spot
f. Two Spotted Spider Mite
g. Target Spot
h. Tomato Mosaic Virus
i. Tomato Yellow Leaf Curl Virus
j. Healthy

6.3 Training Process

The framework for identifying tomato plant diseases in our prototype is a type of supervised learning method, namely deep learning or convolutional neural network. To achieve this, every input or training image has to be annotated with its ground truth label so that the network can learn the appropriate weights to map the output label to the input image correctly. To enable weight updating, the softmax cross-entropy loss between the true label and the predicted label will be computed. The optimiser will then update the network weights following the gradient-descent algorithm in minimising the softmax cross-entropy loss.

During network training, random cropping, flipping, rotation, colour distortion and contrast are applied to the training images before feeding the image into the network. This allows the network to learn features that are invariant to their locations in the images and various transforms.

There are two models trained for this tomato disease identification project, which include transfer learning on Mobile Net v2 and Inception Resnet v2. The former contains less network parameters (which makes it predict faster) and is therefore adopted on mobile devices; while the latter is able to extract more complex features from the input image which is more accurate and necessitates the use of devices with higher computational power such as a Windows PC.

6.4 Results

The following table summarises the network training and validation accuracy achieved by both Model 1 and Model 2.

Table 1: Comparison of network performance between Model 1 and 2.

Accuracy	**Model**	
	1 (Mobile Net v2)	**2 (Inception Resnet v2)**
Training (Top 1)	88.13 %	96.88 %
Validation (Leaf Scan - Top 1)	97.94 %	99.75 %
Validation (Non-leaf Scan - Top 1)	64.44 %	78.89 %
Validation (Leaf Scan - Top 3)	99.92 %	100.00 %
Validation (Non-leaf Scan - Top 3)	93.33 %	91.11 %

From Table 1, it can be seen that Model 2 in general achieves a better performance than Model 1. This is judged through five parameters, which are the Top 1 training accuracy, as well as Top 1 and Top 3 validation accuracy on both leaf scans (PlantVillage dataset) and non-leaf scans (IPM Images).

While Model 2 is able to predict correctly on 99.75% and 78.89% of the validation set for leaf scan and non-leaf scan respectively, the Top 3 prediction accuracy on both validation dataset has exceeded 90%. This would simply mean that the first three disease predictions by Model 2 from our AI solution will always match the actual disease or ground-truth for more than 90% of the cases.

On the other hand, Model 1 has achieved a satisfactory performance comparable to Model 2, while requiring a lesser amount of computational power and being able to run individually on Android smart phones. Achieving an accuracy of 88.13% during training, Model 1 is able to predict correctly for 97.94 % of the leaf scans and 64.44 % of the non-leaf scans. Likewise, the Top 3 prediction accuracy on both the leaf-scan and non-leaf scan validation dataset by Model 1 is always more than 90%, which is similar to Model 2.

Note that the lower prediction accuracy on the validation set for non-leaf scans are due to the significantly smaller size of the non-leaf scan training dataset. However, it is observed that the networks are still able to learn the disease-related visual features on those non-leaf scans. This opens up the possibility of using non-leaf scan tomato plant disease images for predicting tomato plant diseases, and we believe that our solution will perform even better if our training dataset is continuously enriched with more images, especially non-leaf scan images with random background.

At the same time, confusion matrix is used to tabulate the tomato disease classification results for both Model 1 and Model 2. This allows a better understanding on how well our tomato disease classification models predict and what are the exceptions.

Predicted

Actual	Bacterial Spot	Early Blight	Healthy	Late Blight	Leaf Mold	Septoria Leaf Spot	Two Spotted Spider Mite	Target Spot	Tomato Mosaic Virus	Tomato Yellow Leaf Curl Virus
Bacterial Spot	421	1	0	7	1	2	0	3	0	1
Early Blight	1	199	0	9	1	3	0	0	0	1
Healthy	0	0	336	2	1	0	0	0	0	1
Late Blight	1	3	2	389	4	0	0	2	0	0
Leaf Mold	0	3	0	0	198	0	0	1	0	0
Septoria Leaf Spot	0	11	0	2	3	338	0	5	2	0
Two Spotted Spider Mite	0	1	0	0	0	0	330	5	0	0
Target Spot	0	1	2	0	0	0	2	276	0	0
Tomato Mosaic Virus	0	0	0	1	1	0	0	0	75	0
Tomato Yellow Leaf Curl Virus	12	0	0	5	1	0	3	0	0	1058

Figure 2: Confusion matrix for Model 1 predictions on the validation set.

From Figure 2, it is observed that Model 1 is able to achieve a high precision and recall on the validation dataset for all the 10 categories as it is able to identify the ground-truth disease correctly (highlighted in red).

Predicted

Actual	Bacterial Spot	Early Blight	Healthy	Late Blight	Leaf Mold	Septoria Leaf Spot	Two Spotted Spider Mite	Target Spot	Tomato Mosaic Virus	Tomato Yellow Leaf Curl Virus
Bacterial Spot	430	2	0	1	1	4	0	0	0	2
Early Blight	0	209	0	5	0	0	0	0	0	0
Healthy	0	1	338	0	0	0	0	0	0	1
Late Blight	0	0	2	398	1	0	0	0	0	0
Leaf Mold	1	1	0	0	200	0	0	0	0	0
Septoria Leaf Spot	2	2	0	0	0	357	0	0	0	0
Two Spotted Spider Mite	0	0	0	0	0	0	336	0	0	0
Target Spot	0	1	0	0	0	0	1	279	0	0
Tomato Mosaic Virus	0	1	0	0	0	0	0	0	76	0
Tomato Yellow Leaf Curl Virus	1	0	2	0	0	0	0	0	0	1076

Figure 3: Confusion matrix for Model 2 predictions on the validation set.

As for Model 2, it is able to achieve an even higher precision and recall according to Figure 3. As compared to Model 1, Model 2 has a smaller number of false predictions, which makes it a better candidate for the tomato plant disease classification model if the computational power permits.

6.5 Mobile & PC Application

There are two platforms for running our automated tomato disease identification model, namely Windows PC and Android. The identification model that runs on Android devices is a lighter-weight version (Model 1), which allows the farmers to identify the possible plant diseases on the go just by capturing the picture of the infected tomato leaf. On the other hand, the identification model that runs on a Windows PC (Model 2) is able to detect tomato plant diseases with a higher accuracy. This allows the farmer to take more samples or pictures of the infected tomato plant leaves and feed into the model on Windows PC for a more detailed analysis in the lab.

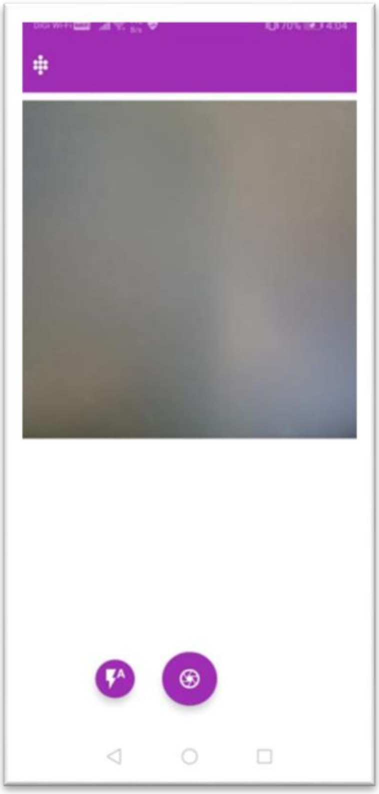
For future plan, the images captured by farmers can be stored in the system servers. Together with the model prediction and feedbacks from farmers and experts alike, the images collected will be utilised to improve the tomato disease classification model from time to time.

The mobile and PC applications built are both able to work as a standalone without being connected to the internet. This allows the farmers who farm in rural area without internet connection (which is most of the cases) to identify the possible diseases that affect their tomato plants. With the decision made by our offline AI solution which will otherwise require expert level knowledge, the farmers can carry out appropriate corrective measures to prevent the spread of tomato plant diseases in a timely manner.

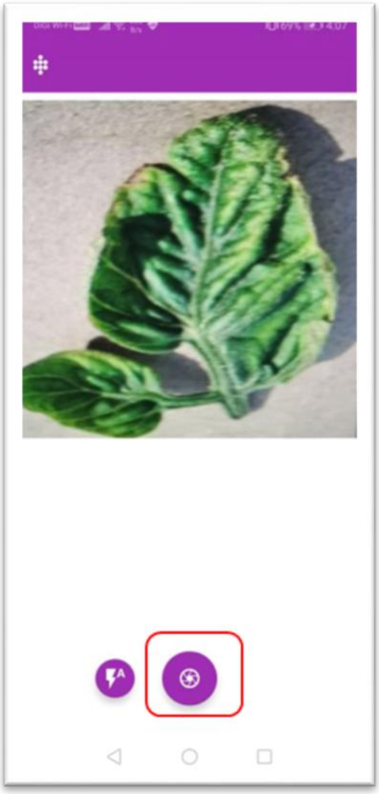
6.6 User interface of the application

<u>Mobile application</u>

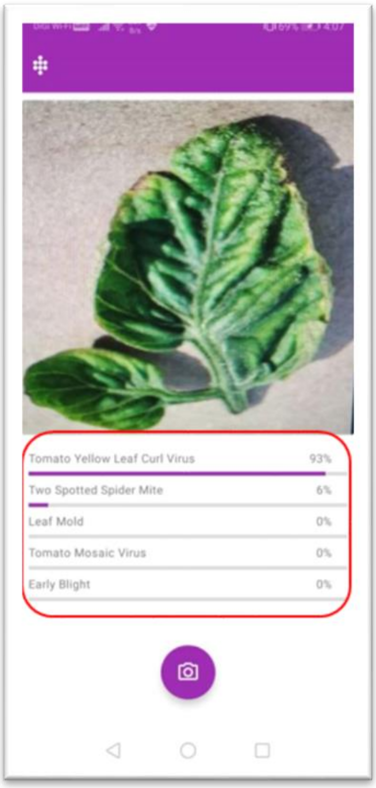
1. Launch the Android application and the following screen will show up.



2. Point the phone camera on the target (tomato plant organ) and hit the camera button.

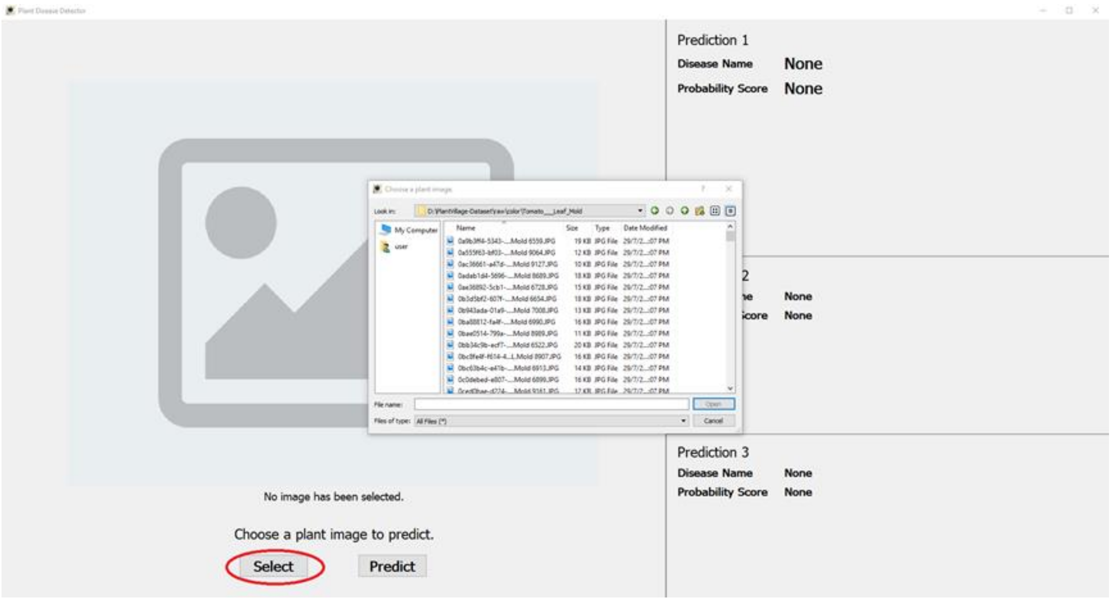


3. The top 5 disease predictions will be displayed on screen.

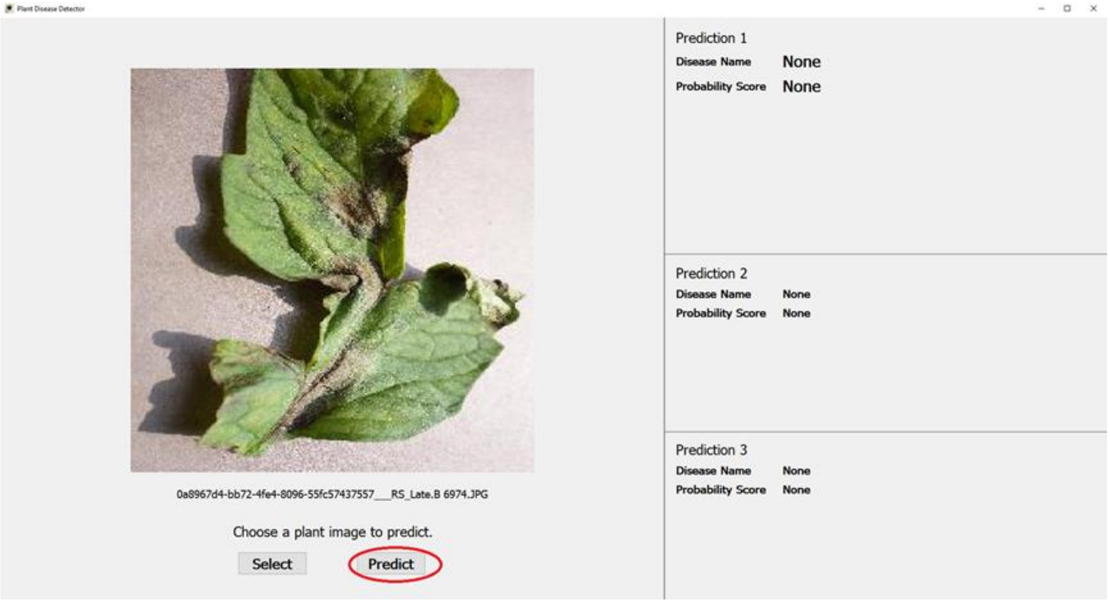


Windows PC application

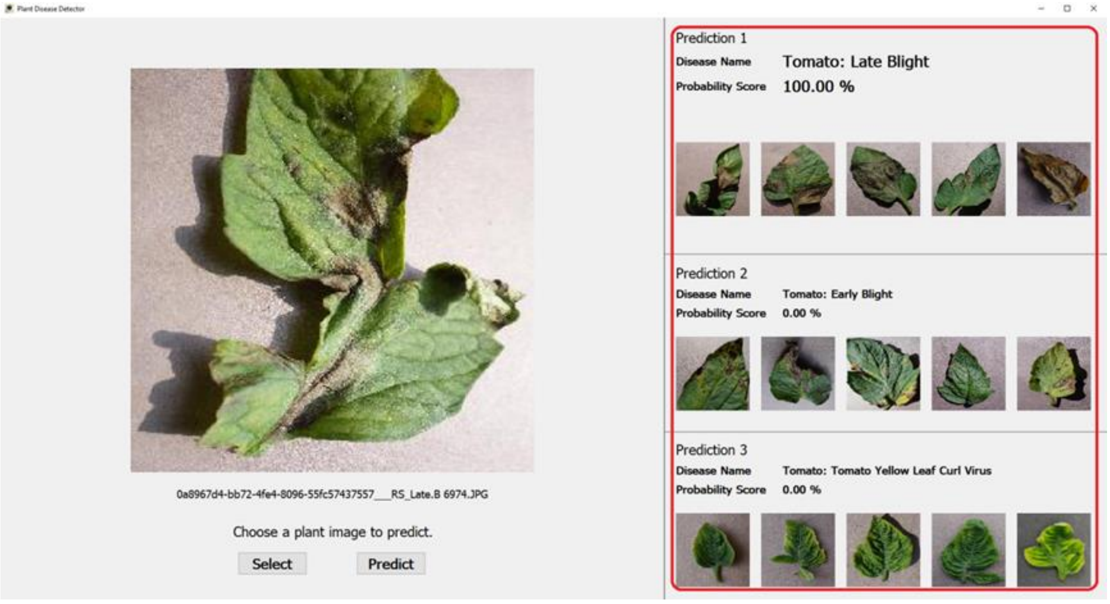
1. Click on the "Select" button to choose the desired image.



2. Click on the “Predict” button after choosing the image.



3. The top 3 disease predictions will now appear on the screen, along with their probabilities. The database images for the predicted classes will also be shown on the screen for the user’s reference.



7.0 Conclusion

In conclusion, our team proposed a crop disease recognition system that uses computer vision and machine learning to allow the identification of crop diseases. In executing this project concept, we have built two tomato plant disease recognition model which is able to identify whether or not the tomato plant is healthy or infected from nine different tomato diseases trained. The first model is run on Android mobile while the other is run on Windows PC. Both work as a standalone and does not require an internet connection. In comparison of both models, the second model which runs on Windows PC achieves a higher accuracy however it requires a higher computing power.

Our tomato disease recognition system shows that the application of artificial intelligence in crop disease detection is practical. With the decision made by our offline AI solution which will otherwise require expert level knowledge, the farmers can carry out appropriate corrective measures to prevent the spread of tomato plant diseases in a timely manner. This can be helpful for farmers in increasing quality crop yield, hence accommodating the need of food supply locally and abroad.

Moreover, this crop disease recognition system creates a platform for farmers, researchers and plant experts alike to collaborate and work together to identify and mitigate diseases found within our community. Data on existing-known crop disease as well as new crop disease can be sourced from the images captured by the farmers allowing the improvement of detection and discovery of new crop diseases. This contributes significantly to the recognition of local crop diseases which may be unknown in other countries. In addition, the knowledge harvested on crop disease within our community can be sustained for existing research purposes and future generations.