SOFTWARE REQUIREMENT SPECIFICATION

PREPARED FOR

Mini Project Proposal (IS 31230)

Department Of Computing & Information Systems

Faculty Of Applied Sciences

Sabaragamuwa University Of Sri Lanka

PREPARED BY

A.I. Hewarathna

17APC3081

Introduction

Purpose

In the agricultural industry for success need continuous monitoring and finding solutions for unexpected situations throughout the various process. Traditionally these actions are done by humans with various experience levels and so many of them participate. But in modern agriculture one of the major problems is the insufficiency of labour to do some things like they did in past. Most of the time farmers need to continuously look for decease and find some time solution. otherwise, it will be too late to do something to rescue their crops and maybe they will put other cultivators in the same danger. Because of this situation, they need help from machines. But how do machines identify diseases? Then computer vision comes to our help. This solution need to identify plant then need to understand if that plant affected by and disease then what is that diseases and then what is the level of risk. Then it should propose some possible solutions for that disease and how to implement that. By extending its ability it could find reasons for this situation like what is a suitable environment and what are seasons these crops should cultivate likewise. For a clear understanding of farmers, it will show more details about market details with rich graphical representations.

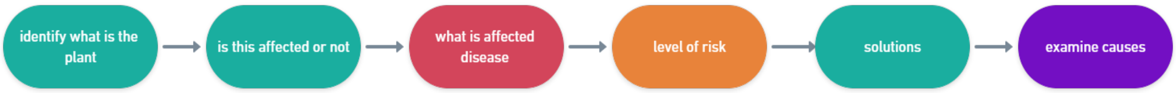

Figure 1 - the core concept

By this proposed system, professional cultivators and novice farmers can get benefits such as automating the very crucial processes and core components of their industry when it comes to large scale they can use drones to gather aerial imagery and input them into this system in real-time and system will investigate for them. In small scale farmers and beginner level, farmers can identify diseases on their own with help of this proposed system. The goal of this proposed system is to give expert knowledge to cultivators in every experience and knowledge level also automate a process of identifying diseases.

Scope

This proposed system can identify 9 major crops types and 30 diseases overall which are common with these plantations. Also, it can distinguish between healthy and affected plants.

plant	diseases
Rice	Bacterial leaf blight
	Brown spot
	Leaf smut
Cassava	Bacterial Blight
	Brown Streak Disease
	Green Mottle
	Mosaic Disease
Apple	Apple scab
	Black rot
	Cedar apple rust
Cherry	(including_sour) Powdery mildew
Corn	(maize)Cercospora leaf spot Grey leaf spot
	(maize) Common rust
	(maize)Northern Leaf Blight
Grape	Black rot
	Esca (Black_Measles)
	Leaf blight (Isariopsis_Leaf_Spot)
Bell pepper	Pepper,_bell Bacterial spot
Potato	Early blight
	Late blight
Strawberry	Leaf scorch

Tomato	Yellow Leaf Curl Virus
	Mosaic virus
	Target Spot
	Spider mites Two-spotted spider mite
	Septoria leaf spot
	Leaf Mould
	Late blight
	Early blight
	Bacterial spot

Table 1 - plants and diseases systems can identify

Whether plantation affected or not more insights will be present to the user of the system as a dashboard. The first thing in this analytics is about the plant, which consists of success level, harvest, nutritions, fertilizers, common diseases, suitable weather conditions, suitable locations in-country, and seasons. Secondly, how the market behaves on a certain product, historical patterns what will be the predicted need, what will be the production amount likewise.

Risk

A major risk of this system is models overfitting to training data and doing worse in real inputs when going to production and operations states. To prevent this situation two approaches can use, one is to use the 70:20:10 ratio in splitting data. 70% of data points were chosen randomly without replacing to training and 20% of rest to validations and final 10% to test the performance of the trained models. One of the side effects of this approach is data leakage from validation data to the model. The second method is to use an 80:20 ratio with k-fold cross-validation. Which 80% of randomly selected data points were used to train and the rest of 20% was used to test the performance of the trained model. But in training cross-validation is implemented. Basically, train data randomly split into k splits and hold out one of them to validation and the rest of the splits (k -1) used to train the model. This routing is done to all possible combinations of the splits.

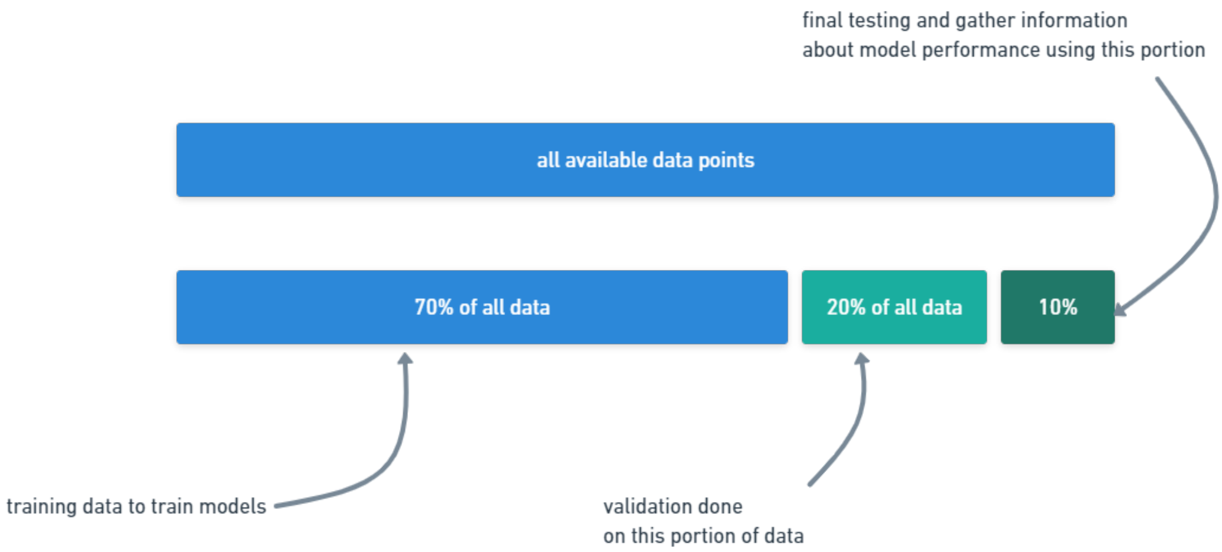

Figure 2 - 70:20:10 method

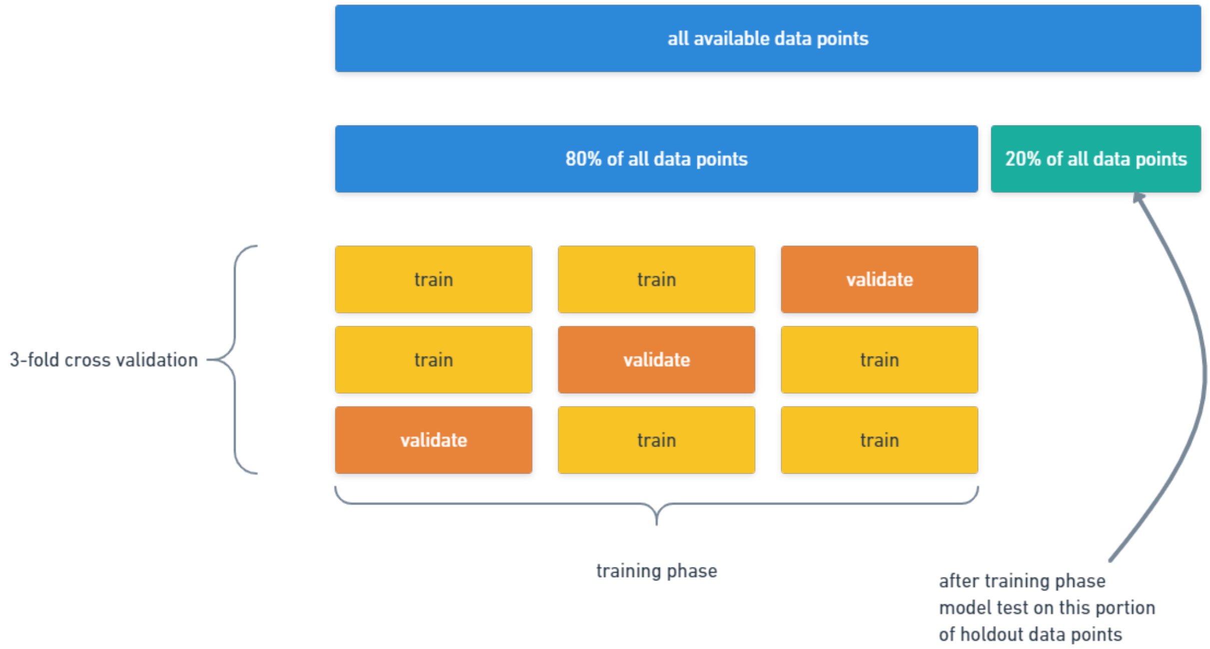

Figure 3 - 80:20 method with K-fold cross-validation

System features and requirements

Functional requirements

Any user can fully access

For wider accessibility, this system should be embedded into a mobile application so anyone who has a smartphone can use this software. But models with this much complexity maybe struggle to properly run on mobile platforms. So its back end should be separate from the application and established in a cloud computing platform. Mainly mobile applications act as an interface to upload information and output results from the machine learning models.

Available as much as plant types

There are hundreds of plants that humans cultivate for their use. But it’s not practical at this level to build a solution that can identify each Train plant and each disease that can be affected to them. So the solution is to narrow down to the most common plants and common diseases that can be affected to them. Plants that are available with this solution are rice, cassava, cherry, corn, grapes, potato, soybeans, strawberry, and tomato. There are 30 overall diseases this solution can identify.

Most common disease types

Diseases this model can identify are reported more frequently than other diseases so farmers can identify those diseases easily using this solution. Because of the high probability of affecting one or more diseases to crops this solution is very useful for large scale to small vegetable farms.

Present possible solutions

After identifying the affected plant and then deciding what is disease affected this plant. Then using a knowledge base connected with the backend retrieve possible solutions. Those solutions can be categorized into three categories which are chemical treatments, non-chemical treatments and others. The category of the solution depends on the disease itself and affected area, seriousness of the disease.

Indicate seriousness level

Seriousness level indicates how much danger affected disease and how much damage it can do to the crop and harvest. it‘s a numerical scale created using historical data observed when certain diseases outbreak and do critical damages to eco-systems and harvests. High values indicate critical and lower values indicate less danger.

Provide market insights

Current market patterns will be important for farmers to decide which time they should harvest and guess approximately what will be the prices for their products. Also, historical patterns are provided with this solution to get educate what happen recent past to detected plant harvest in market. By this method, food waste can be controlled. To create these insights use data gathered daily basis in major harvest collecting centres and major markets throughout the country.

Dynamic dashboards with filters to retrieve much-needed details about plant

Rather than presenting static image type charts proposed to create dynamic dashboards to present analytics prepared based on findings. By providing filters to change detail levels and easily understand patterns user intended to know.

Captured images saved for further model training

Users need to upload or capture source images to model to work on then those sources save in a repository to use future training of the model. Because one requirement for the more accurate model is more data to learn. By collecting ground truth data models can improve continuously to more accurate predictions.

External interface

To get inputs and show model findings to users there is an application and it acts as the only user interface. Through it, users of the solution can upload images of plants to the model located in the cloud platform and then after prediction is done then sen results and details to create dashboards in the user interface. Those uploaded sources are also saved in the cloud platform for future use. There is no user authentication on the user side.

Non-functional requirements

Performance

Fast classification and retrieving data from knowledge bases to send user interface is needed to secure the quality of the solution. The model should be trusted with its outputs while testing and training phases with a stress test to guarantee accuracy.

Accuracy

The accuracy of a multilabel classification task like this accuracy should be at least over 70% to put trust in that model. But over time with more data models can be more accurate. To secure accuracy and recall is crucial when training so regularization techniques need to use. Another hand more than one architecture need to create while constructing the model because the optimal solution among them can be put into a production state. Because different architectures work well in different situations so use a few architectures certainly help to choose optimal architecture.

Minimise FN and FP rate

Another important requirement is to lower false negative and false positive rates when doing predictions. False-negative and false-positive is a concept that explains wrong predictions by models.