How can machine learning be used in determining crop disease in plants?

By David Chen, Altrincham Grammar School for Boys

*Abstract*— Plant diseases are defined as an abnormal growth or dysfunction within a plant. Crop diseases affect plants that are grown on a large scale can result in many different problems for society, ranging from a drastically lower economic output, to a lack of food supply in third world countries. Current plant disease identification is often done manually, which is time consuming and costly. Machine learning has emerged as an effective approach for automated plant disease detection. This aim of this report is to examine how current machine learning approaches can be used for detecting crop disease. We do this by applying different machine learning methods to a publically available dataset to detect plant disease. The computing for these methods is no longer than a couple seconds. The experimental results show that tensorflow is better than the other methods.

Keywords—Image processing, pattern recognition, machine learning, crop disease, deep learning

# Introduction

As the population of the globe increases exponentially, projected at nearly 10 billion by 2020[1], being able to balance converting unused land into habitable land, and farmable land becomes more difficult to manage. Eventually, we will have to be able to maximise the gains from each farm in order to be able to sustain life on the planet. Currently, the most common technique used to identify crop disease is by manual visual identification. The problem with identification using the naked eye is that a lot of people are needed in order to cover a farm area effectively. However this brings in each person’s objective view. As one person may see crop disease in one plant, another may not. These inconsistencies combine resulting in different plants being wrongly flagged. It’s important to recognise crop disease as one outbreak could lead to many people either emigrating away from a country, or dying from starvation. This is hugely harmful to any country. An example would be the disease named “late blight”. This disease is responsible for destroying more than half of the tomato crop in the United States. It is also famous for causing the Great Famine in Ireland in the mid-19th century. This can also happen for many other diseases, as crop diseases are not limited to blight. Pathogens consisting of much fungi, virus, and bacteria can all infect any parts of a plant. These infections and diseases may also be highly contagious, which can devastate an entire crop field incredibly fast. The use of pattern recognition itself when examining the crops are currently being looked at in a number of different settings, like identifying weeds, and discolouration in citrus fruit. There are many viable methods to detect crop disease, and therefore the aim of this project is to highlight the efficiency of Machine Learning, and specifically, pattern recognition, in this application.

# Current research

The research on the concept of using image processing for detecting crop disease has started relatively recently( “Digital image processing techniques for detecting, quantifying and classifying plant diseases”, Jayme Garcia Arnal Barbedo, 2013). When it comes to mobile devices, the research began even more recently. This could be due to the inferior computing capacity that you can have on a mobile cellular device, compared to a desktop computer. However, as the processing power of mobile devices become stronger, the ability to perform complex image processing becomes easier to achieve. Alternatively, a different method that could be used in the future includes having a cloud server for which image taken with cellular devices can send the pictures to. This cloud server would contain a mainframe computer that would do the processing before sending the results back to the user. This alternative method has problems that stem from being able to stay in range of a wireless network, which would require a long time to set up, and is also expensive.

Sagar Patil, Bharati Dixit and Anjali Chandavale (Patil, Chandavale, & Dixit, 2015) conducted a research into automatic detection of plant diseases with which their intensity of colours is the same, however the colours themselves are different. This method takes into account the RGB values of each pixel in the picture to classify whether each picture contains a disease or not. The median filter technique was applied to smoothen the image, and the Otsu thresholding method is used for calculating boundaries between different diseases.

Yi Fang and Ramaraja Ramasamy (Fang & Ramasamy, 2015) listed many ways to detect plant disease, most of which included biological tests such as the Enzyme Linked Immunosorbent Assay or ELISA test. These tests are very accurate, as the antigens from the virus, bacteria or fungi can only bind to one shape of antibody. This assay contains the specific shaped antibody, meaning this method is almost always correct when it comes to determining whether a plant is a carrier of the pathogen which can cause a specific disease. A couple methods that are less manual and rely on imaging, include methods such as thermography. This method takes into account the difference in colour when an infrared picture of a plant is taken via thermographic cameras. The resulting disease is carefully monitored using thermographic imaging, and the amount of water lost is able to be calculated, without changing the surrounding environment.

Another paper (Martinelli, et al., 2014) reviews several methods, based around nucleic acid-based. As some pathogen detection methods are based upon DNA, such as fluorescence in situ hybridization (FISH) and many PCR variants. Others are RNA based, such as reverse transcriptase-PCR, and nucleic acid sequence-based amplification (NASBA). These methods overcome uncertain diagnosis, enabling a fast and accurate detection and quantification of pathogens. This method however is only efficient when running diagnosis one a few specimens. When taking into the account of large areas to test for pathogens, it is not the most efficient method to go around. They also review different methods such as lateral flow microarrays (LFM). These allow fast hybridization based nucleic acid detection by using a colorimetric signal. The arrays are built on lateral flow chromatography nitrocellulose membranes, and have detection limits similar to microarrays, and are able to reduce the need of expensive equipment. However this technology is not widely available, and so not many people are able to utilise this method.

# Proposed method

The following diagram shows the flowchart of the training process, which we will use in order to train the classifier

Testing stage

Classification

Feature extraction

During the input stage, training images that contain pictures of plants with crop disease and without crop disease are inputted in the program.

PSEUDOCODE:

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

DATASET= “dataset/trainingset”  
 Pictures=[]

For picture in DATASET:

Pictures.append(picture)

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~