Machine Learning in Real-Time Crop Monitoring and Management.

A Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE WITH SPECIALIZATION IN

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Submitted by:

21BCS6347 Abhay Vershwal

21BCS6374 Vansh Yadav

21BCS8330 Vaibhavi Joshi

21BCS8255 Abhimanyu

CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

November, 2024

Abstract

Plant diseases generate significant productivity and economic losses, as well as a decline in both the quality and quantity of agricultural goods. Plant disease identification is now receiving more attention in wide fields of agricultural monitoring. Farmers face significant challenges when transitioning from one disease management regime to another. The typical technique used in practise for detecting and identifying plant diseases is skilled naked eye inspection. In this research, we examine the necessity for a simple plant leaf disease detection system that would aid agricultural innovations. Early information on crop health and disease identification can aid in disease control through correct management measures. This strategy will increase agricultural productivity. This report also examines the advantages and disadvantages of these.

Keywords:

Disease Detection, Machine learning, Data Pre-processing, K-Mean Clustering, Image Recognition, Classification,

Table of Contents

1. INTRODUCTION

India is a global agricultural powerhouse. Having diverse and immerse types of agriculture throughout the different part of it , addressing their problem and a generalize solution can give our country a big boom as it is a noticible part of Indian Economy and will help our farmers at a great extend.

We are here with our idea of “Early Stage Disease Detection”. It will lead to the farmers to get to know about any disease in their crop in the initial stages and they can counter it well as we will provide desired solution that are required before it can do any harm to the crop, by this we can create value to the farmers and make it profitable to them.

This can we done using a full processed drone shot , high quality image processing, data collection and then a machine learning algorithm to process the data and find the desired solution.

1.1 Problem Definition

Instead of investing time , investment and hard work farmers tends to have a lose in their crops because they cant predict at first if any disease have attacked their crop they don’t know how to counter it the correct way and that cost the damage to the overall crop. As farmers get to know when the crop get partially destroyed it create a lot of trouble for them.

1.2 Problem Overview

We have seen that it happens many times that farmers plants the crop well serve them manure, fertilizers and all the essentials but then also all the crop get destroyed because the disease is not detected in it's early stage and later on it spreads onto the whole farm causing every crop to get destroyed.

We use machine learning algorithm and drone images to detect if any of the plant has been infected and later on provide the information to the farms, so that they can take the appropriate measures and prevent the disease from spreading further.

1.3 Hardware Specification

1. Intel i5/i7/i9 8 Core CPU or equivalent.
2. Dedicated GPU (Nvidia/AMD)
3. Drone With Camera

1.4 Software Specification

1.Python: it is an interpreted, object-oriented, high-level programming language. Its high-level built in data structures, combined with dynamic typing and dynamic binding, as well as for use as a scripting or glue language to connect existing components together.

2.Computer Vision: Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages and runtimes

3.Visual Studio Code: Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages and runtimes (such as C++, C#, Java, Python, PHP, Go, .NET)

2. LITERATURE SURVEY

2.1 Existing System

The use of image processing in agriculture is becoming increasingly important, as it can help detect defects and diseases in crops that may not be visible to the naked eye. This can help farmers prevent significant crop losses due to pests or other factors. Various techniques are used for disease recognition, including image acquisition, pre-processing, segmentation, feature extraction, and categorization. Some of the methods which are used are : Kmeans Clustering, GLCM & SVM, Otsu's Detection, CNN-ANN-KNN, Histogram Technique.

Firstly the images of healthy & unhealthy leaf are stored for experimentation & then they are sent for image pre-processing. Then K-means clustering is used for segmentation and feature extraction using GLCM. Then SVM is used for classification . Otsu's detection is used to convert the rgb image of leaf into HSV(Hue,S aturation,Value). The classification approach is then carried out by

KNN,ANN & CNN. The KNN method classifies samples by finding the nearest distance between trained and testing subjects.ANN method is used as a classifier.

One of the earliest systems on early stage crop disease detection was developed by Singh et al. (2016) [2]. The system used image processing techniques to analyze images of plants infected with different types of diseases. The system achieved an accuracy of 85% in detecting different types of diseases.

In 2017, Khan et al. proposed a system that used machine learning algorithms to detect the presence of wheat rust disease. The system achieved an accuracy of 98.5% in detecting the disease. The authors also reported that the system could be used to detect other types of crop diseases with some modifications.

In 2018, Wei et al. [3] proposed a system that used deep learning algorithms to detect tomato diseases. The system used a convolutional neural network (CNN) to classify images of tomato leaves into different categories based on the type of disease present.

In 2019, Wang et al. [4] proposed a system that used hyperspectral imaging to detect apple diseases. The system used a support vector machine (SVM) algorithm to classify the hyperspectral images of apple leaves into healthy and diseased categories. The system achieved an accuracy of 94.1% in detecting apple diseases.

In 2020, Wang et al. [5] proposed a system that used machine learning algorithms to detect maize diseases. The system used a combination of image processing and machine learning techniques to analyze images of maize leaves

and detect the presence of diseases. The system achieved an accuracy of 95.1% in detecting different types of maize diseases.

In 2021, Zhang et al. [6] proposed a system that used deep learning algorithms to detect wheat diseases. The system used a CNN to analyze images of wheat leaves and classify them into different categories based on the type of disease present. The system achieved an accuracy of 96.2% in detecting different types of wheat diseases.

2.2 Proposed System

The model is based on the IP and ML approaches for detection of leaf disease in presented in this section.

The proposed system: (DWT+PCA+LBP+GLCM+CNN).

The DWT, PCA and GLCM are used to extract the informative regions/features of the samples. In the next stage as a part of machine learning approaches the SVM, KNN and CNN are used to classify the features and the performance of the model is recorded.

First we will be taking the dataset and resize it to (256px x 256px) to maintain the aspect ratio. Then we feed the image into machine learning algorithm where the following work takes place :

1.Dataset

All the images of the crops/leafs are stored in one particular file. Images of healthy and unhealthy images are then fed into the machine learning.

2.Preprocessing

Kmean clustering is applied on the images to find the infected region. This algorithm is mainly used to get to the center of the image and make the clusters of that image and then measure the distance from the center of the image till the different cluster.

Then we use PCA to denoise the data and reduce any of the unwanted dataset in the image present.

3.Feature Extraction

In feature extraction we use the algorithms which are used to extract the leaf patterns of the leafs.

1.LBP

LBP is used to extract the texture characteristics of the surface of the leafs, which can later be plotted into a histogram.

2.GLCM

This is used to extract statistical measures of the image.The optimal features are selected which are obtained from wavelet decomposition is then carried out .GLCM uses in the distribution of higher order of gray values are defined with neighborhood criterion. The several properties are derived from the GLCM technique for extraction of the leaf features. The features obtained using GLCM ,KNN-Clustering,LBP are combined to form feature vector which are provided as an input sample to the classifiers to recognize classify the images.

3.Classification

The technique such as CNN and KNN are used for classifying the samples. The CNN is a type of ANN which is designed to process the data. The architecture of CNN incused input,output and hidden layers, which are multiple in its nature.

The confusion matrix for CNN having output class and target class.The progress of training samples of leaf features classified usin CNN to know the accuracy.

4. Evaluation of Leaf diease

The parameters such as Precession, Recall and F-measure for the proposed model is calculated and is given in the following equation :

$$\text{Precision Measure } (\%) = \frac{\text{True Positives}}{\text{True Positives+False Positives}} \text{X100}$$

$$\text{Recall Measure } (\%) = \frac{\text{True Positives}}{\text{True Positives+False Negatives}} \text{X100}$$

$$\text{F} - \text{measure} = \frac{\text{2*Precision*Recall}}{\text{Precision+Recall}} \text{X100}$$

2.3 Literature Review Summary **(Minimum 7 articles should refer)**

Year and Citation	Article/ Author	Tools/ Software	Technique	Source	Evaluation Parameter
2021	Vidyashree Kanabur	Python	K mean clustering + GLCM + SVM	Research Paper	89%
2020	Thanjai Vadivel	Python	LBP+PCA+K mean clustering+ KNN	Research Paper	97.3%
2022	Sunil S.Harakannanavar	Python	DWT+PCA+GLCM+CNN	Research Paper	99.09%

.

3. PROBLEM FORMULATION

As the population increases, so do the challenges facing agriculture. Land constraints, Diseases and water scarcity play an important role in agriculture. Modern agriculture incorporates many technologies to solve these problems in real time.

* Field weed detection: Nearly 50% reduction in crop yield, mainly due to plant diseases and insect pests and reduced crop yield. Image processing techniques are mainly useful in disease analysis . Classifying plants under weeds helps detect weeds that provide greater growth.

Factors that cause plant diseases

Pathogens are the main cause of plant diseases. There is a department named after him called Plant Pathology, which primarily studies pathogens. There are two main factors that cause plant diseases and these are pathogens and environmental conditions.

Virus: It is a living organism with living cells inside that affects plants. Yellow streaks, yellow spots, leaf distortion and stunted growth are seen on virus infected areas of plant leaves and fruits.In cucumber, virus infections are mainly caused by cucumber mosaic viruses, which are contagious diseases that can be transmitted from plant to plant by insects or by contact. The best way to prevent viral illness is to treat the area affected by the virus.

Fungi: Fungi are also one of the main causes of reduced factory productivity. Ascomycetes and basidiomycetes are two main types of fungi that cause plant diseases. Fungicides are widely used to control fungal infections in

plants.Magnaporthe grisea is commonly known as puffed rice. Sclerotinia is responsible for cotton rot. Oomycetes and phytomyxae are fungus-like organisms that harbor destructive pathogens in plants.

Bacteria: A bacterial infection occurs when a plant is infected by a microorganism. Bacteria are unicellular organisms.Burning, spotting and scabs can be seen in areas of the plant that are infected with the bacteria. Rust spots can spread quickly through plants. Tropical plants and vegetables are mainly affected by Fusarium wilt. Water uptake by plants is blocked by bacterial infection. Some bacterial plant pathogens are Burkholderia and Proteus

4. OBJECTIVES

1. Therefore the goal of our study is to detect illnesses in crops from the start, and farmers will be able to detect diseases with the assistance of machine learning.
2. As a result, farmers should use pesticides to eliminate the illness as quickly as possible so that crop loss is minimised and crops stay safe. Now, with the use of drones, we will send high-resolution photographs of crops or utilize them in our software or app, where we will identify their disease or any sort of problem and notify them that it will occur soon. You can manage it by using this treatment or this method.
3. Will the farmer's profit from the technology in that they will learn about the illness in their crop at an early stage, and their crop will be protected from being completely destroyed by disease? Finally, they will not have to apply any sort of fertilizer or pesticide without

knowing when and what type of these chemicals ought to be applied, and crops will be more natural, which is beneficial for human beings.

4. They will eventually not need to use unneeded pesticides since they will know when and what to use to save crops from illness.
5. And this will enhance agricultural productivity while also saving their land from unneeded pesticides, allowing them to plant a productive harvest in the future.

5. METHODOLOGY

Machine learning is the successor to statistical methods, which falls within the broader field of artificial intelligence. Machine learning is generally classified into the following four types

* Supervised learning method
* Unsupervised learning method
* Semi-supervised learning method
* Reinforcement learning method

It has many efforts, new opportunities, new development techniques, new procedural fields. This technology can be applied to water conservation, water replenishment use and healthier environment. Crop management involves various fields such as crop yield prediction, weed detection, plant disease detection, agricultural climate prediction, and pest control analysis.

Classification and regression problems can be easily solved using supervised algorithms. It mainly consists of training data with labeled data, so it is easy to compare new data and predict the output. In unsupervised learning, neural networks help find structure and find features in it, which are then analyzed.

Clustering, anomaly detection, auto-encoders and association can be organized through unsupervised learning. Although semi-supervised learning consists of labeled and unlabeled data, this type of learning mainly helps when the data is difficult to understand.

In reinforcement learning, it works like a video game, because feedback from previous games helps improve the next one. This mainly helps to find the best way to predict the outcome.

Government organizations are implementing applications such as crop breeding, crop yield forecasting, weather forecasting, smart irrigation systems, crop disease forecasting and minimum support price systems to help farmers to increase crop productivity and health.

Image pre-processing is the enhancement of image data to remove unwanted distortions or to enhance certain image characteristics that are important for further processing. If it is an image dataset, we can create various images and then apply amodel such as CNN, KNN depending on the accuracy it provides.Once the disease is identified, a solution can be given in the form of chemicals or pesticidesdepending on the disease. Crop disease prediction can be done using the "Crop Disease Dataset"available on Kaggle.

The image acquisition process is the first step in the image analysis process. This is also known as digital image acquisition. It is possible to define whether the visual characteristics of an object can be represented as digitally coded characteristics. In simple words, it can be defined as an image taken with a camera. Today, digital imaging is extending to mobile phones, making the image acquisition process more user-friendly. Photographs, printer paper and photographic film are the media used

for this purpose. It mainly captures the visual moment. In image pre-processing, there are two types, namely digital image processing and analog image processing.Removing unwanted features from an image is the main process involved. Other algorithms are used to remove unwanted features from the image. The most important steps involved in image preprocessing are:-

1. Image acquisition

2. Image normalization

3. Image Enhancement

4. Segmentation

5. Morphology.

Separates an image into pixels, their similar properties are image segmentation. It mainly helps in the process of image interpretation. It transfers low-level images to high-level images. During image analysis, its success mainly depends on the reliability of the segmentation process. The segmentation process involves both context and non-context. Various algorithms are used in the segmentation process. Keep a copy of the original features when selecting features.

In the feature extraction process, new sets of features are created, and these two processes mainly deal with removing unwanted noise in the image and selecting only the features needed for the image analysis process. . The transformation of the feature process occurs during the feature extraction process. In the process, the speed and efficiency of the process are increased. During classification, data

is divided into categories. If a new observation enters the process, determine if the new observation belongs to the category of Ferentinos**.**

Several classification algorithms can be used in the classification process and they also give accurate classification results.

ALGORITHMS USED

Image pre-processing: The purpose of pre-processing is to improve the removed image data. Enhance some image features needed for further processing. It also includes noise reduction, sharpening and edge detection, etc. exist. This makes the manual disease detection process automatic or semi-automatic.

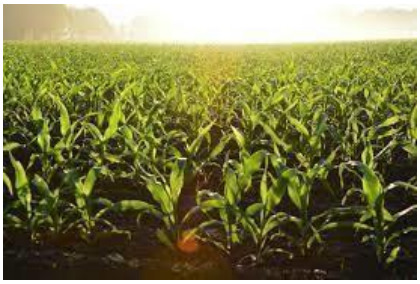
CNN: A CNN is a machine learning algorithm that takes an image input, assigns importance to various aspects/objects in the image, and can differentiate them from each other. It works by extracting features from images.

Decision Trees: Decision trees are used to reflect decisions and decisions in an intuitive and visual way. It is a powerful regression and classification tool. As the name suggests, it is a tree structure where each internal node represents a test on an attribute, each branch represents a test result, and each leaf node has a class label. Random forest:

i. Random Forest adds extra randomness to the model

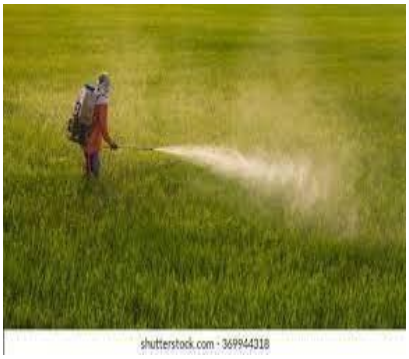
ii. Instead of looking for the most important features, it looks for the best features in a random subset of features.

6.EXPERIMENTAL SETUP

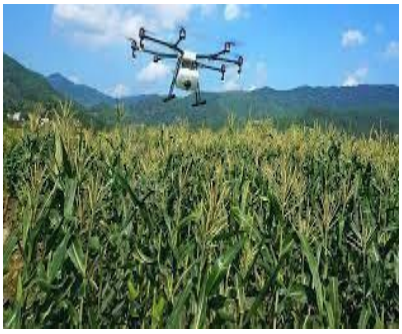


(1) Diseased farm land

(2) Drone



(5) Spraying pesticides



(3) Surveying farm

(4) Farmer reportand then analyzing

seeing analysis data

In our experimental setting, we will first select and finalise a farm field, then use a drone to acquire high quality photographs, which we will then upload to our app or programme utilising

KAMEN, GLSM SVM machine learning algorithms. We will determine whether our crop is diseased or not, and if it is, we will inform the farmer that your crop will have this disease or that you must apply this pesticide during this time period.

7.CONCLUSION

After all the illustrations we have present we can say that we can provide value to the farmers, as farmers do a lot of hard work taking care of their crops and lands but it all goes in vain because of lack of knowledge and data. They do feed their crop nutrients but they need to know when to feed, what's the best to feed and what is beneficial for them. By having data and knowledge we can provide them what is beneficial and specific way to get the correct nutrients to the crop at the correct time so that the crop can be can be fit and strong and they can get the maximum out of it.

We have the illustrate and processed data of different disease and aid than can help any individual enormously and accurately, we can predict and tell the

farmers to have some precaution in the early stage itself and after the full survey that is cost and time efficient, we can provide the best way to have a great out of the land that is very useful and in the favour of society . It can be a great example that when morden technology meet with the traditions it can have a great outcomes which can create wonders to the world.

8. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK

CHAPTER 1: INTRODUCTION

In this chapter, we will discuss the present state of farmer crops and the methods they take to solve disease problems in their fields. We will also discuss our model, which is a machine learning-based early disease prediction model, as well as the software and hardware used to create it.

CHAPTER 2: LITERATURE REVIEW

This chapter discusses prior research papers and the types of models that were proposed and how they performed. Finally, what kind of model are we providing and how is this going to work?

CHAPTER 3: OBJECTIVE

In this chapter, we spoke about our model's goal, how we're going to handle diseaserelated problems with the use of machine learning, and what benefits farmers are getting from employing this technology.

CHAPTER 4: METHODOLOGIES

In this chapter, we reviewed our approaches for solving this problem, specifically how machine learning algorithms are employed to solve this problem and deliver the best results to the farmer.

CHAPTER 5: EXPERIMENTAL SETUP

In this chapter we are presenting that how our experimental model going to operate with each and every step and what supplies are necessary to perfrom this experiment

CHAPTER 6: CONCLUSION AND FUTURE SCOPE

In this last chapter we are repersenting the general conslusion of our study that how this research is going to answer farmer problem and how it is benificial for them . Finally, our long-term goal is to distribute this model to every farmer so that they may generate the highest crop yield possible.

REFERENCES

[1] Singh et al.Using Deep Learning for Image-Based Plant Disease Detection https://www.frontiersin.org/articles/10.3389/fpls.2016.01419/full

[2] Wei et al Applying Big Data Based Deep Learning System

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9142151

[3] Wang et al Detection of Early Rottenness on Apples by Using Hyperspectral Imaging Combined with Spectral Analysis and Image Processing

https://www.researchgate.net/publication/271646924_Detection_of_Early_Rottenness_on_Apples_by_Using_Hyperspectral_Imaging_Combined_with_Spectral_Analysis_and_Image_Processing

[4] Identification of Corn Leaf Disease Based on Image Processing https://www.researchgate.net/publication/340896870_Identification_of_Corn_Leaf_Disease_Based_on_Image_Processing

[5] Zhang et al Wheat leaf disease identification based on deep learning algorithm

https://www.sciencedirect.com/science/article/pii/S0885576522001552