ICSD/79

DETECTION OF DISEASES IN TOMATO PLANTS USING DEEP LEARNING

Rushi Chaudhari
SCOE Pune, Savitribai Phule Pune University, Maharashtra,India
rushic24@gmail.com
Rohan Marathe
SCOE Pune, Savitribai Phule Pune University, Maharashtra,India
rohanmarathe1997@gmail.com
Rohan Rane
SCOE Pune, Savitribai Phule Pune University, Maharashtra,India
rohanrane1998@gmail.com
Vedang Pingle
SCOE Pune, Savitribai Phule Pune University, Maharashtra,India
vedang.pingle@gmail.com

Abstract

Crop disease is increasingly becoming a concern for the farmers due to many factors such as climate change, growing resistance of pathogens, etc. An ill-diagnosed disease, leading to incorrect treatment ultimately affects the quality and quantity of yield, thereby causing financial distress to the farmer. The proposed system aims to serve as a tool for the farmers to identify diseases in tomato plants, using a smartphone application as a means of interface and a Convolutional Neural Network (CNN) for identification of diseases. Processing and storage constraints are eliminated with the use of Cloud platform. We also aim to make the application more comprehensive by suggesting the remedies and preventive measures for the detected disease. The proposed system is highly scalable and aims at achieving a higher accuracy in disease detection by surveying the drawbacks of the existing system.

Keywords: Tomatoes, Disease Detection, CNN, VGG19.

I. Introduction

India is widely considered to be an agrarian country and ranks second in the world in terms of agricultural output. Despite agriculture being practiced almost everywhere throughout the country and playing a crucial role in the nation's economy, its potential remains relatively under fulfilled due to lack of penetration of technology in this sector. One of the prominent issues in agriculture is crop damage caused by various viral, bacterial, and fungal diseases.

Tomato is one of the most popular and widely produced commercial crop. Tomato production is sensitive to temperature; the optimal temperature required for the growth of tomato is 23-27^{0}C. Temperature below 15^{0}C and above 35^{0}C during the day time and above 21^{0}C during the night is detrimental to the fruit setting. Tomato crop suffers through large number of severe diseases caused by fungi, bacteria, viruses, mycoplasma, nematode etc. Many a times, in order to combat these diseases, farmers use pesticides, fungicides and fertilizers excessively which in some cases may sharply increase yield from cultivation, but at the same time has caused widespread ecological damage and negative human health effects. Also if the farmer misrecognizes the disease or fails to detect it in time, it may spread throughout the field and cause irreparable damage.

Researchers have studied and performed disease detection based on image analysis in the past using various image processing and or machine learning techniques. Learning models such as SVMs and K- Means Clustering have shown decent results when it comes to smaller datasets with lesser parameters, but the technique which has been the most promising is Deep Learning. Deep learning is a subset of machine learning. Deep artificial neural networks are a set of algorithms that have set new records in accuracy for many important problems, such as image recognition, sound recognition, recommender systems, etc. Two different deep learning network architectures AlexNet and SqueezeNet have been trained and tested previously on image datasets for disease classification. However, such CNNs are heavy and have high processing demands which are difficult to satisfy using a regular computer. It is therefore preferable to train and store the model on a Cloud platform.

II. Literature Review

Recently, a similar deep learning approach was adopted for the Plant Village dataset by Durmus et al. where AlexNet and SqueezeNet was used for training and validation [1]. Although relatively high accuracy was achieved, there were processing constraints due to the use of an onboard computer. Another method involved the use of K-means clustering for segmentation followed by feature extraction using GLCM [2]. An ANN was employed consisting of Multlayer Perceptrons for learning. This system achieved relatively lesser accuracy of 83-87% and the diseases detected were limited to 4. Hase et al. proposed a system involving an android application which served as an interface for image acquisition and image analysis [3]. Foreground extraction using Grabcut was followed by feature matching using FLANN. This system was not as accurate as previous systems and is difficult to scale.

A combination of VGG16 CNN model and an SVM was used [4], where VGG16 model was used as image feature extractor and the SVM was used as a classifier. Data augmentation techniques were employed to increase the dataset size. The model achieved a moderate accuracy of 89% on the high quality dataset. Islam et al. developed a system for detection of potato diseases, in which they employed a Multiclass Support Vector Machine for classification [5]. Feature extraction was performed using Grey Level Co-occurrence Matrix (GLCM). The system was limited to detection of two diseases only.

III. System Overview

The designed system is an android-based stand-alone system and works independently of any other hardware or software. The android module acts as an interface with the user. The system involves a cloud platform for processing the image and database storage. The android module is connected to this cloud platform securely and data transfer takes place between the two. There is an additional localization module which uses Wi-Fi signal strength of neighbouring routers to determine the location of the user.

Following are the main functionalities provided by the proposed system:

1. Identification of disease based on analysis of input image.
2. Suggestion of solutions including but not limited to a range of fertilizers and fungicides.
3. Performing localization of the diseased crop with a reasonable accuracy.
4. Intuitive and interactive image segmentation via an android application.

The system requires the user to capture the crop image against a monochromatic plain background which is pre-processed and then subsequent functions are performed.

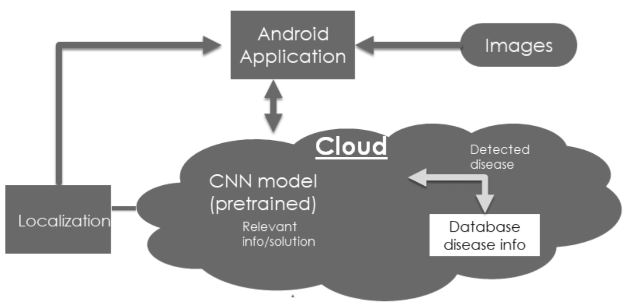

Fig 1. System Overview

An incremental model is used in the development of this project. The incremental build model is a method of software development where the model is designed, implemented and tested incrementally (a little more is added each time) until the product is finished. This model combines the elements of the waterfall model with the iterative philosophy of prototyping. The incremental model applies sequences in a staggered fashion.

The process is repeated following delivery of each increment, until complete product is produced.

During the development of this project, we will first construct a VGG19 CNN model which trained and tested on Plant Village dataset. Once desired maximum accuracy is achieved, a basic camera app to capture picture and process the image would be created. In this way, each add-on would be added in sequence. During this process requirements may change and product so developed would be able to accommodate these change in requirements.

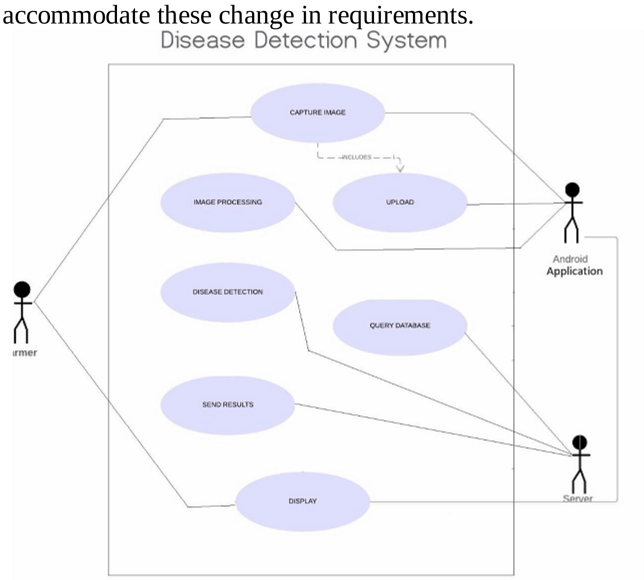

Fig 2. Use Case Diagram

IV. Mathematical Model

1. Let ‘S’ be the system

   S = {……..

2. Identify the inputs as U,Lp

   S = {U,Lp………..

   where U are the user inputs, Lp are the identified time and coordinates

   U = {I}

   I = {i|’i’ is the image of a leaf of the diseased crop}

   Lp = {X,Y,T}

   X = {x|’x’ is the x coordinate of the location}

   Y = {y|’y’ is the y coordinate of the location}

   T = {t|’t’ is the timestamp}

3. Identify the output as O

S = {U,Lp,O…….

O = {O1,O2}

O1 = {disease predicted}

O2 = {Rc|Rc is the method/remedy suggested to overcome the disease}

4. Identify the processes as P

S = {U,Lp,O,P…….

P = {Ic,Dc,Sc}

5. Ic is the Image pre-processing and filtering for relevant feature extraction and compression

Ic = {Ps,Pc}

Ps = {f|'f' is function segmenting the image and extracting foreground}

Pc = {f|'f' is function compressing the image}

6. Dc is processing and analysis of data on cloud

Dc = {Iip,Pi}

Iip = {t|'t' is valid input from the smartphone device}

Pi = {p|'p' is function that analyses the input and provides a decision based on the said input}

Pi(Iip) = Iop

IOp = {m|'m' is the probabilistic output of the classifier }

7. Sc is the remedy suggestion

Sc = {Rip,Rp,Rop}

- Rip = {r|'r' is the input from Dc}
- Rp = {r|'r' is function for association of input & reference dataset}
- Rp(Rip) = Rop
- Rop = {r|'r' is the remedy suggested}

8. Identify failure cases as F'

S = {U,Lp,O,P,F'…….

Failure occurs when

- F' = { ϕ }
- F' = {p|'p'ϵ I,p is inconsistent}U{Fs|Fs $\notin$ Fc for C}

9. Identify success case (terminating case) as e

S = {U,Lp,O,P,F',e…….

Success is defined as-

e = {a|a ϵ O1,a is correct}U{Fs|Fs ϵ Fc for C}

10. Initial conditions as S0

S = S = {U,Lp,O,P,F',e,S0}

S0 = {U = { ϕ },Lp = { ϕ , ϕ , ϕ },O={ ϕ },P={Running}}

V. METHODOLOGY

A. *Leaf image analysis for disease detection*

- The process of taking image as an input, and returning a vector (features)/matrix (image) as an output.
- Algorithm:
  1. Represent Image By a Set Z = {z1, z2, z3, ……..zn} where n is the number of pixels
  2. Zn is a 1-D parameter like gray value
  3. The image is treated like a graph with n nodes, and two special nodes, source and sink, which are connected to each node by a t-link, while the n-nodes are connected to its adjacent nodes by a n link.
  4. A cut is selected such that a minimum cost is incurred from the assigned weights.

B. *Deep Learning (VGG 19 For Classification)*

Deep learning neural networks were inspired by our understanding of the biology of our brains – all those interconnections between the neurons. But, unlike a biological brain where any neuron can connect to any other neuron within a certain physical distance, these artificial neural networks have discrete layers, connections, and directions of data propagation. Each neuron assigns a weighting to its input — how correct or incorrect it is relative to the task being performed. The final output is then determined by the total of those weightings. An artificial neural network deep learning model such as CNN is best suited to this task as it gives high accuracy when the dataset is sufficiently large and also outputs multiple probabilistic outcomes and hence can classify multiple diseases.

VGG-19 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 19 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has input image dimensions of 256 x 256. The input image is classified into 1 of 10 categories (1 healthy and 9 diseased).

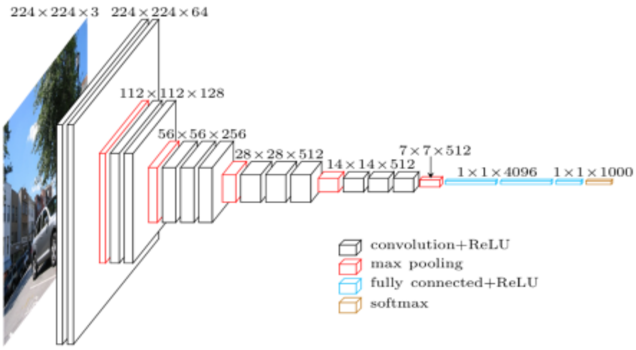

Fig 3. VGG-19 Network Architecture

C. Dataset

- The dataset used for the model training has been acquired from PlantVillage and consists of around 10000 images (1000 images/category approx).
- The images are divided into 10 categories i.e 1 healthy and 9 diseases of the tomato crop.

VI. RESULTS

- The Grabcut segmentation method produces a well-segmented, noise free image as shown in the Fig. 4 and Fig. 5.
- The CNN model trained on the Plant Village Leaf Image Dataset yields 97.5% testing accuracy. Real time accuracy can be improved by retraining further on more data collected in real time.

Fig 4. Leaf Image Fig 5. Segmented Image

FUTURE WORK

The dataset used consists of high quality images which are noise-free. Retraining must be done in future on images captured through smartphone to make the system more robust in real time.

REFERENCES

[1]] H. Durmuş, E. O. Güneş and M. Kırcı, "Disease detection on the leaves of the tomato plants by using deep learning," *2017 6th International Conference on Agro-Geoinformatics*, Fairfax, VA, 2017, pp. 1-5.

[2] [2] M. Dhakate and Ingole A. B., "Diagnosis of pomegranate plant diseases using neural network," *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*

[3] *[3]* A. K. Hase, P. S. Aher and S. K. Hase, "Detection, categorization and suggestion to cure infected plants of tomato and grapes by using OpenCV framework for andriod environment," *2017 2nd International Conference for Convergence in Technology (I2CT)*, Mumbai, 2017,

[4] [4] J. Shijie, J. Peiyi, H. Siping and s. Haibo, "Automatic detection of tomato diseases and pests based on leaf images," *2017 Chinese Automation Congress (CAC)*, Jinan, 2017

[5] [5] M. Islam, Anh Dinh, K. Wahid and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Windsor, ON, 2017

[6] S. P. Mohanty, D. Hughes and M. Salathe, "Using Deep Learning for Image-Based Plant Disease Detection," eprint arXiv:1604.03169, 2016.

[7] [7] Pawar B.T. "Report of Bacterial Diseases of Tomato from Marathwada Region of Maharashtra, India " Research Journal of Recent Sciences ISSN 2277-2502 Vol. 3(ISC-2013), 302-304 (2014)

[8] [8] Karthik & Praburam. N "Detection And Prevention Of Banana Leaf Diseases From Banana Plant Using Embeeded Linux Board " 2016 Online International Conference on Green Engineering and Technologies (IC-GET)

[9] [9] Pranjali B. Padol Prof & Anjali A. Yadav "SVM Classifier Based Grape Leaf Disease Detection" 2016 Conference on Advances in Signal Processing (CASP) Cummins College of Engineering for Women, Pune. Jun 9-11, 2016

[10] [10] Suyash S. Patil & Sandeep A. Thorat "Early Detection of Grapes Diseases Using Machine Learning and IoT" 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP).