INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 462-467

DEEP LEARNING-BASED TOMATO LEAF DISEASE DETECTION AND PESTICIDE SUGGESTION PLATFORM FOR FARMERS

Balaji V[1], Bharath Kumar S[2], Karthick M[3], Manoj k[4], Manoj S K[5], Mr. B. Ganapathy Ram[6]

[1,2,3,4,5]B. E, Ece National Engineering College, Kovilpatti, India.

[6]M.E., Associate Professor (Sg), National Engineering College, Kovilpatti, India.

ABSTRACT

Depending on maximum consumption in daily use, Tomato is considered as the well cultivated and commercial crop. Several viruses spread the diseases on the tomato plants, mainly leaves got infected among the other parts of the plant. There is a wide possibility of spreading of virus to the neighboring plants in the field. This will distress the cultivation of tomato plant. Farmers are seeking a benefitable technology or methodology that predict the type of diseases and the virus that leads to the spread of that virus. And that methodology must provide the instruction or guidelines to farmers about the suitable pesticides to be use to prevent the infection of virus. A web portal that will detect the tomato leaf disease using deep learning method is developed that will suggest the suitable pesticide using database management system. Nearly 10 different diseases infected in leaves was identified and classified using Convolution Neural Network. Around 94% to 98% of accuracy in identifying the diseases is achieved.

1. INTRODUCTION

The tomato is the edible berry of the plant Solanum Lycopersicum, commonly known as the tomato plant. The tomato plant can grow 0.7–2 m (2.3–6.6 ft) in height and as an annual, is harvested after only one growing season. Tomatoes are usually easy to grow, and a few plants provide an adequate harvest for most families. The quality of fruit picked in the garden when fully ripe far surpasses anything available on the market, even in season. Tomato plants fall into one of two types that affect ultimate plant height and cultural requirements. Tomatoes are determined if they eventually form a flower cluster at the terminal growing point, causing the plant to stop growing in height. The first determinate varieties developed had real problems with inadequate foliage cover and taste, but they ripened very early. Newer determinates produce better foliage, may grow taller, and ripen fruit of similar quality to modern indeterminate varieties. They still tend to ripen their fruit over a shorter period, so successive plantings may be desirable with determinates to keep the harvest coming through the entire season.

2. OBJECTIVE

The objective of the project, "Deep Learning-based Tomato Leaf Disease Detection and Pesticide Suggestion Platform for Farmers," is to develop a robust and user-friendly system that utilizes deep learning techniques to accurately detect diseases in tomato plants based on leaf images. The platform will also provide effective pesticide suggestions to help farmers combat the identified diseases. Develop a deep learning model capable of accurately detecting and classifying common diseases that affect tomato plants, such as early blight, late blight, bacterial spot, tomato yellow leaf curl virus, etc. The model will be trained using a large dataset of annotated tomato leaf images representing various disease conditions. Incorporate a pesticide suggestion component into the platform based on the identified disease. This component should provide information about appropriate pesticides, their application methods, and dosage recommendations to effectively treat the detected diseases. The suggestions should consider factors such as disease severity, environmental conditions, and local regulations.

3. EXISTING METHODS

Earlier approaches in tomato leaf disease classification involved different image- based feature extraction techniques that were fed into machine learning-based classifiers.

Support Vector Machine

Support Vector Machines Supervised learning algorithm Support Vector Machine popularly known as SVM is a machine learning algorithm that is used for both classification and regression problems. In this study of the classification of tomato plant leaf diseases, SVM is applied in which each tomato leaf image is plotted as a point in n-dimensional space where n is the properties of the image amidst each value of the attribute being the value of a specific coordinate. These attributes are then segregated by hyperplane to classify the images accurately.

Random Forest

Random Forest is an ensemble model consisting of multiple decision trees. This is a very reliable algorithm that performs random sampling of training data to avoid overfitting and determines the negatively correlated trees to surpass single models.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 462-467

Xtreme Gradient Boost

XGBoost is an open-source library that provides elevated performance in implementing gradient-boosted decision trees. This bundle is extremely powerful and quick to deploy. Unlike decision trees where a single model is trained on the dataset and the results are predicted, the boosting approach is the more effective anditerative technique as it combines multiple models and foresees the output of the final model. Being an ensemble technique, this approach works successively with each other, and the new model is trained to overcome the errors of the earlier modelin such a way that no improvements can be done further, hence in this study extreme gradient boosting approach is applied on the images of tomato plant leaf in which the newer models are trained to correct and reduce the residuals of the previous models in classifying the diseases of the plant accurately.

k-nearest neighbors (KNN)

It is an algorithm that can be used for classification. The idea behind the method is that it assumes that similar things exist nearby. The method takes into considerationthe data points of each image in the dataset using Euclidean distance in other to groupthem. So, when an image is applied to the model, the input image will be convertedto a feature vector. Thereafter, the image will be used to construct a color histogramto classifier the color of the pills and then stored under a class label extracted from the image path

Disadvantages:

- Manually designed features work well in a controlled environment but would yield poor performance in unconstrained settings such as with images captured by mobile devices.
- effective binary and multi-class classification performance on a spectrum sample set of small size.
- The drawback of segmenting the image using K-Means clustering is that the process proposed was semi-automated as the user has to explicitly select the cluster which contains the diseased part. SVM - Even though the obtained accuracy is high, it is not sufficient to predict or differentiate between healthyor diseased leaves.

4. PROPOSED SOLUTION

Plant disease is a major problem in agriculture, that minimizes the quantity of food production and reduces the quality of agricultural products. The heedful disease diagnosis and appropriate handling protect the plants from massive losses Convolutional Neural Networks (CNNs) have revolutionized image processing, especially deep learning methods. Over the last two years, numerous potential autonomous crop disease detection applications have emerged. These models can beused to develop an expert consultation app or a screening app. The proposed systemdiscusses the plant leaf's disease detection by using a method of images. Propose themethod is used to predict the disease accurately. Then it detects all the plant diseases easily. That is used to detect all plant leaf-related diseases and provides more accuracy. Then the detection of plant disease at less computation time. They are easily detected and it is used accurately to identify the diseases in less processing time. The proposed Multilayer Convolutional Neural Network-based ternary classification model is then trained for the detection and classification of Tomato leaves.

This ternary model includes three cases i.e.

1. To classify the given image for Tomato leaf or not,
2. The image is a non-diseased Tomato leaf image, and the image is a diseased Tomato leaf image.
3. Visualization methods to localize the leaf.

Region Proposal Network

This region proposal network takes the convolution feature map that is generated by the backbone layer as input and outputs the anchors generated by sliding window convolution applied on the input feature map.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

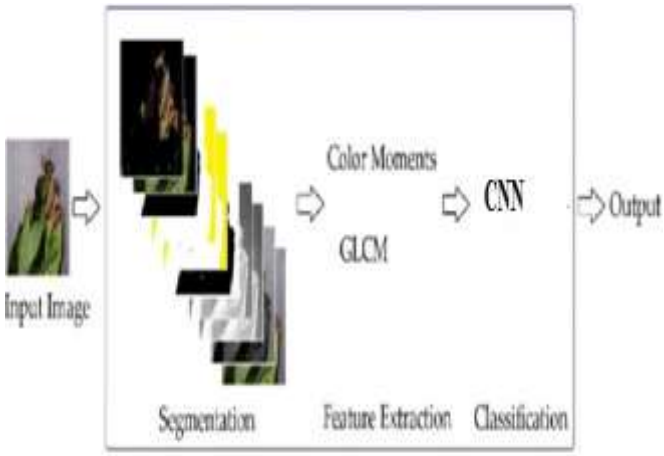
Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 462-467

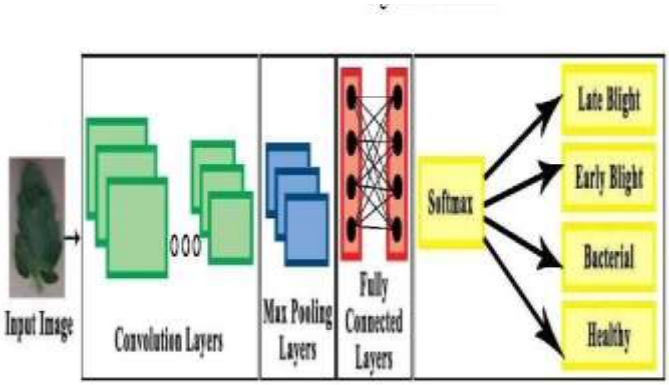
Gray Level Co-occurrence Matrix

Gray Level Co-occurrence Matrix (GLCM) based texture analysis of kidney diseases for parametric variations. GLCM parameters (Energy, Correlation, Contrast, and Homogeneity) were extracted for each color component of the images taken for the investigation. Contrast, correlation, energy, and homogeneity represent the coarseness, linear dependency, textural uniformity, and pixel distribution of the texture, respectively. The analysis of the GLCM parameters and their histograms showed that the said textural features are disease dependent. The approach may be used for the identification of CKD diseases with satisfactory accuracy by employing a suitable deep learning algorithm.



Convolutional Neural Network (CNN)

A CNN is a type of deep learning used to analyze visual scenes. It is characterized by having one or more hidden layers, which extract the attributes in videos or images, and a fully connected layer to produce the desired output. Whereas for the computer, the image is a 3D array (width × height × depth) of values ranging from 0 to 255. It is simply pixels of color, if the number of channels is one, the image is grayscale, black, and white. Besides, the channels are three colors (if the images are RGB). CNN Deep Network has shown outstanding performance in many competitions related to image processing due to its accurate results. CNN is a hierarchical structure that contains several layers. The basic components of the basic convolutional neural networks are the Convolutional Layer, the Activating function, the Pooling Layer, and the Fully connected Layer.



5. PROCEDURE

Pre-processing

Tomato Leaf Image pre-processing are the steps taken to format images before they are used by model training and inference.

The steps to be taken are:

- Read image
- RGB to Grey Scale conversion
- Resize image - All the images collected are modified to be less than 200KB in size with a maximum resolution of 1280 times 720 because the larger the images are the longer it will take to train the algorithms.
- Original size (360, 480, 3) — (width, height, no. RGB channels) Resized (220, 220, 3)
- Remove noise (Denoise) - smooth our image to remove unwanted noise. We do this using gaussian blur.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 462-467

Bi-lateral Filter

Filtering is maybe the most basic operation of image processing and computer version. In the broadest sense "filtering", the estimation of the filtered image at a given location is an element of the estimations of the input image in a small neighborhood of a similar location. For example, Gaussian low-pass filtering calculates each pixel is replaced by a weighted average of its neighbors, in which the weights diminish with distance from the area of focus. The formal and quantitative explanations of this weight difference can be given, the instinct is that images vary gradually over space, so close pixels are probably going to have comparative values, and it is in this manner suitable to average them together.

Binarization

Image binarization is the process of taking a grayscale image and converting it to black-and-white, essentially reducing the information contained within the image from 256 shades of grey to 2: black and white, a binary image.

Split Dataset Training Data

WPA2 A CNN model is developed using 80% of the training data and this percentage may change depending on the needs of the experiment. It's used to train the CNN model, which tries to learn from the training data set. Both the input and the predicted result are included in the training data.

Test Data

The test set is 20% of the original data and is used to evaluate the CNN model on new data. It is used for the model's evaluation process once it has been fully trained.

Tomato Leaf Detection

Region Proposal Networks are a crucial part of the detection model Faster

Take Inception_5b as an example: through the per-network, a shared feature map sizing $7 \times 7 \times 256$ is obtained. Then, the feature map is reshaped to $7 \times 7 \times 512$ using 3×3 convolution kernels. To obtain categories and regression results, the 1×1 convolution process is implemented in the classification layer and regression layer, obtaining feature maps of size $7 \times 7 \times 30$ and $7 \times 7 \times 60$, respectively. Finally, with the arranged anchors, the candidate boxes are gained.

Inspired by Feature Pyramid Networks, an RPN structure is proposed for locating the irregular and multiscale diseased spots Through a deconvolution process, the high semantic information of Inception_5b is integrated with the high resolution of Inception_ResNet-v2. Thus, the proposed detection model can predict diseased spots separately in each feature layer. Furthermore, the bottom-up feature extraction and top-down up sampling method enhance the ability of the model to detect small diseased spots.

Tomato Leaf Feature Extraction

After the Leaf detection, the detected image is given as input to the feature extraction module to find the key features that will be used for classification. With each image, the facial information including shape, size, and pattern is automatically extracted and is then used to calculate the effects of the variation using its relation to the frontal obstacle templates.

GLCM Feature Extraction

GLCM (Gray Level Co-occurrence Matrix) is used to extract the statistical texture features. Attaining the histogram details will only give us details about the texture whereas the GLCM calculates the relative position of the pixels relative to their texture. This statistical approach gives a lot of information about the relative position of the neighboring pixels in an image. GLCM is a spatial domain technique that tabulates the difference in combination of pixel brightness in the image. The features involved in feature estimation are divided into a few steps: First four co-occurrence matrices are calculated from the grayscale image. It considers the distance between the pixels to be 1. So, the co-occurrence matrix is computed at 00. There are four features in every computed matrix namely correlation, contrast, energy, and homogeneity. Hence the feature vector will be of size 16.

Tomato Leaf Disease classification

A Convolutional Neural network is a deep learning algorithm that takes an input image and assigns importance to the entities or objects in an image which helps it to learn and detect. Even the Convolutional Neural Networks are developed based on neurons in a human brain keeping in mind, that the visual cortex is present in humans.

These CNNs are given preference over other neural network architectures because of the following reasons. Convolution Neural Networks require a considerably lower amount of pre-processing when compared to other neural networks. CNNs are efficient to read and capture both temporal and spatial data from an image. The network in CNNs can be trained to understand the important details in the image better.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 09, September 2023, pp : 462-467

Testing Phase: Prediction- After completing the training process, the algorithms have been evaluated on 20 images including the images from the test dataset containing leaves with disease andhealth in various environmental conditions and also including several other aquatic entities.

Prediction Network- The goal of this detection network is to generate a final bounding box by considering inputs from Feature Network and Regional Proposal Network. It consists of 4 fully connected layers which are in turn interconnected to the bounding box regression layer and classification layer that help to generate final detections.

Keras- Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It was developed with a focus on enabling fast experimentation.

- Allows the same code to run on CPU or GPU, seamlessly.
- Built-in support for convolutional networks (for computer vision), recurrent networks (for sequence processing), and any combination of both.
- Supports arbitrary network architectures: multi-input or multi-output models, layer sharing, model sharing, etc. This means that Keras is appropriate for building essentially any deep learning model, from a memory network to a neural Turing machine.

BACKEND

In Backend, we have used Mysql which is a fast, scalable, and easy-to-use databasemanagement system. It is used in conjunction with Php scripts for creating powerfuland dynamic server-side or web-based enterprise applications. we have also used WampServer which allows us to create web applications with the Mysql database. Alongside, PhpMyAdmin allows you to easily manage our database. We have useda python framework named flask normally it is with little to no dependencies on external libraries.

6. SYSTEM IMPLEMENTATION

Problem Description- Our approach proposes a method to detect diseases and pests of tomato plants usingtechnology based on Deep Learning. The system consists of three basic components: a primary diagnosis unit (Bounding Box Generator), a secondary diagnosis unit (CNN filter bank), and an integration unit. For each image and class category, the primary unit generates a set of bounding boxes with scores of a specific class instance and the coordinates that indicate the location of the target. Then, the secondary unit filters the confidence of each box by training CNN classifiersindependently for each class to further Perify their instance. Finally, the integrationunit combines the results from the primary and secondary units.

MODULE DESCRIPTION TOMATO LEAF DISEASE PREDICTOR AND RECOMENDER WEB UI:

The Web UI is developed with Flask Framework using Python and MySQL. Web Interface The complete process of model training and performance is performed in the backend, so for the easy user interaction, we use a frontend application that helps us in analysing the images that are uploaded from the sources like an external camera or downloaded from the internet. These test images are easily analyzed with help of backend processes and are represented using this user interface, as we test these images and represent the result soon after that. It provides the good security in this system. When a server wants to connect to a network resource, the host will send the challenge message. For the received challenge, user will generate the hash value. This value will be sent to the host for identification, comparison will take place. The acceptable user alone gain access to use the network resources.

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies, and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail unacceptably. There are various types of tests. Each test type addresses a specifictesting requirement.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages, and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

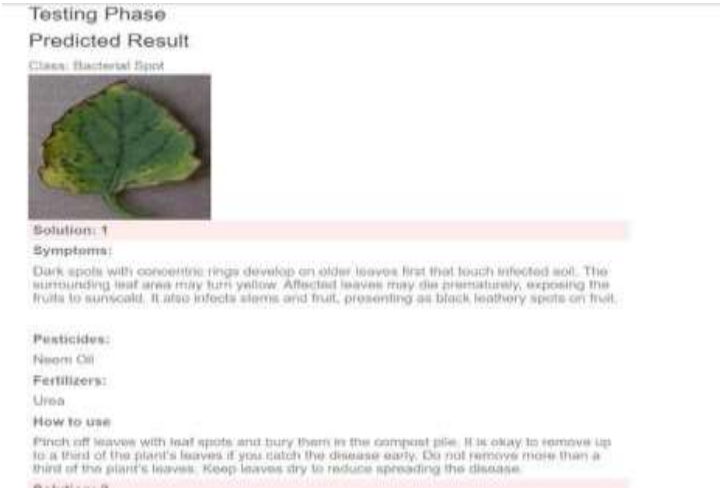
Vol. 03, Issue 09, September 2023, pp : 462-467

7. FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. The organization and preparation of functional tests are focused on requirements, key functions, or special test cases. In addition, systematic coverage about identifying Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effectivevalue of current tests is determined.

Dataset Description- The purpose of this project is, from a photo of a plant leaf, to determine which plantit is, whether the plant is healthy, or which disease the plant is affected by. The datasetis a combination of all the images consisting of 1000 per class for training and 100 per class for the testing dataset which consists of 10000 images divided into 10 classes for training, and 1000 for testing, ranging from healthy to different types of tomato leaf diseases from Kaggle.

8. RESULTS AND CONCLUSION



In this, the dataset is gone into pre-processing which involves the use of a noise filter, and binarization. Then segmentation uses foreground extraction and background extraction which is used for leaf detection. Following the process of segmentation is feature extraction which extracts features like edges, and texture. The processed image which is in the grayscale is provided into the CNN classifier, which detects the type of disease present in the leaf. Then for the training phase, we provided a total set of 10000 images with 1000 per class and for the testing phase, we provided a total set of 1000 images with 100 per class provided from the Kaggle website. From the feature extraction processes, the grayscale image is made as an input to the CNN classifier which classifies the diseases in the tomato leaf. the accuracy of the validation phase is better as we have provided only twenty percent of the dataset,whereas eighty percent of the dataset is used for training, hence while calculating the average, the validation phase will have higher accuracy compared to the training phase.

9. REFERENCES

[1] Y. Yao, X. Wu, Z. Lei, S. Shan, and W. Zuo, “Joint Representation and Truncated Inference Learning for Correlation Filter-based Tracking,” in Proceedings of the European Conference on Computer Vision (ECCV), pp. 1–14, 2018.

[2] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, “Road crack detection using deep convolutional neural network,” in Proceedings of the 23rd IEEE International Conference on Image Processing, ICIP 2016, pp. 3708–3712, Phoenix, AZ, USA, September 2016

[3] Z. Zhou, J. Shin, L. Zhang, S. Gurudu,M. Gotway, and J. Liang, “Fine-tuningconvolutional neural networks for biomedical image analysis: Actively and incrementally,” in Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pp. 4761–4772, USA, July 2017.

[4] W. Ouyang and X. Wang, “Joint deep learning for pedestrian detection,” in Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV ’13), pp. 2056–2063, Sydney, Australia, December 2013.

[5] Yang Wu, Xian Feng, and Guojen Chen, “Plant Leaf Diseases Fine-Grained Categorization Using Convolutional Neural Networks,” IEEE, April 11, 2022.

[6] P. Sharma, P. Hans, and S. C. Gupta, ‘‘Classification of plant leaf diseases using machine learning and image preprocessing techniques,’’ in Proc. 10th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2020.