Cropable - The Crop Disease Detection WebApp

Shashwat Kumar¹, Archisa Kumar¹, Disha Goyal¹, Anannya Chuli², Riddhi Maniktalia¹, Deepa K¹

¹SCOPE Vellore Institute of Technology Vellore, India

Abstract

According to estimates, every year 10% of global production, goes waste due to pests and crop pathogens. For instance, India is a leading producer of many crops, including wheat, rice, lentils, sugarcane, and cotton. But a majority of the farmers are unable to detect whether a crop is infected or not simply by looking at it. As crop pathogens develop greater resistance to fungicides and pesticides, there is an urgent need to find new antifungal compounds to effectively combat them, which over time are rendered useless as the pathogens again develop resistance to these compounds. Thus, the food security of any country is always at risk due to the vulnerability of the current agricultural systems to climate, pests, pathogens, and associated diseases. To solve this problem, we have developed Cropable, The Crop Protection App. In the proposed work, we have used Deep Convolution Neural Networks (CNN) models to detect the disease and further created a web app using flask. Cropable is an Artificially Intelligent Web Application that can help to identify whether the crop is infected or not. We also provide farmers with a treatment for the detected disease, which not only helps them in identifying a disease but also assists them in solving it.

1 INTRODUCTION

Agriculture is the predominant industry in India, and the country ranks second in global agricultural output. Indian farmers cultivate a diverse range of crops that are subject to various factors, such as climate, soil conditions, and diseases, which can affect their yield and production. At present, the identification of plant diseases depends entirely on visual examination without the use of any specialized equipment or technology, which demands additional manpower, adequately outfitted labs, pricey gadgets, etc. Inadequate disease diagnosis can also result in inadequate use of pesticides which may result in the emergence of pathogen resistance in the long run and reduce the crop's natural defense mechanisms.

*Corresponding author: shashwat.kumar2@gmail.com, archisa26@gmail.com, dishagoyal2001@gmail.com, anannya.anannya02@gmail.com, maniktaliariddhi@gmail.com, deepa.k@vit.ac.in

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
Due to increasing global population, political instability, and changing climate conditions, the agricultural industry is seeking new and improved methods to enhance food production.

Many farmers are leaving agriculture for new occupations as a result of industrialisation and low productivity.

Leaf disease poses a significant threat to food security as well. Both the quality and crop output are compromised. It reduces harvest yields and diminishes product quality. Microorganisms such as insects, pests, fungi, bacteria, and viruses, can spread diseases to plant leaves. Feeding of these micro-organisms on both the upper and lower surfaces of the leaves can result in the entire plant being affected. To prevent further agricultural losses, leaf diseases must be identified early. As a result, the economy is boosted by higher food output, which benefits farmers. Finding out how healthy the plant is is essential. Examining the afflicted leaves will reveal the ailment. On the surface of the leaf, irregularly formed black pigment patches appear, and if the area is damp, fungus can grow there. These spots start out little, but over time, they grow until they cover the entire leaf, which leads to deterioration. For the proper diagnosis of leaf diseases, a specific window of time must be given, typically during the early stages of plant development before key processes like pollen movement and fertiliser absorption, are jeopardised.

India’s agriculture system till now is very dependent on natural factors. It’s time we upgrade our systems by enhancing them in order to ease the work for our hardworking farmers. The main aim of our project Cropable- The Crop Protection App is to empower our farmers by providing accurate crop disease detection and treatment measures. This will help them to increase their crop sale and quality of crops. By utilising the modern technology, which enables farmers to boost agricultural output, productivity, and quality at a very low cost, we must expand the use of agriculture. We must keep an eye on and regulate conditions like temperature, humidity, and light in order to produce crops with improved yield and quality.

This paper is focused on Deep Convolution Neural Network (CNN) models to be of service to farmers in terms of disease detection, and our app- Cropable is solution oriented, which gives out treatment measures as well at the same time.

2 Literature Review

Agricultural systems are always vulnerable to climate hazards, pests, and diseases that threaten food security. The health and productivity of plants are essential to humankind, the environment, and the production of food, fiber, energy, and general well-being. Hence the issue of detecting leaf diseases and the various types of diseases affecting crop quality has long been a concern in the agricultural industry. Many researchers have focused on developing computer vision and image processing techniques to create various leaf disease detection systems. Traditional leaf disease recognition depends on visual inspection and human annotation, which involves thorough research, costly equipment, more workforce, laboratories with proper equipment, and constant and continuous monitoring of farms for disease can be time-consuming tasks. Moreover, as many diseased leaves may exhibit similar characteristics, the traditional approach to identification may not always be accurate.

Dr. K. Thangadurai K. Padmavati [1] accomplished Disease detection of plant leaves using three fundamental concepts of Image Processing- Grayscale Images, Color
Conversions Histogram Equalizations. Increasing the quality of grayscale images by applying cumulative histogram equalization generating a new image by applying new values and creating a new histogram for the resultant image increases their ease of processing and implementation. The concept of histogram equalization improves the contrast of the images to create a clearer visual output for human observation. Although MATLAB was utilized here to implement and test the functions, better computational tools could have been used to achieve better efficiency and accuracy.

Several machine learning classification techniques were compared in [2] to recognize the most accurate algorithm to detect plant diseases. As compared to other classifiers, SVM classifiers perform better when applied to disease classification as the CNN classifier detects more diseases with greater accuracy than other existing statistical models and neural networks. However, further advancements may include models like naive Bayes and decision trees that can achieve a higher degree of probability accuracy.

In their research paper, Sachin D. Khirade and A. B. Patil [3] focused on examining techniques to detect plant diseases using images of the leaves. The paper also addressed various methods of segmentation and feature extraction that can be utilized for diagnosing plant diseases. The use of image processing can assist in diagnosing and categorizing plant diseases in accordance with the recommended model, critical for improving the efficiency of crop management. An extensive dataset would have provided better results.

Melike Sardogan Adem Tuncer Yunus Ozen [4] demonstrated an approach that successfully detects four different forms of tomato leaf diseases. The output feature vector generated from the convolution component was fed into the LVQ for network training. Deep learning has the advantage of being able to automatically extract characteristics from images. The process of learning to extract features is accomplished through training the neural network. CNN is a well-known deep learning model that involves multiple layers in a feed-forward neural network. One flaw in the proposed approach is that only a small number of leaves in each class have been mistakenly classified, and it is clear from the table in which categories the incorrect classifications have been included.

Omkar Kulkarni [5] defined the purpose of this research as applying Deep CNN to both crop species classification and disease identification on images. The dataset contains 13 distinct crop species and 26 varieties of diseases. As part of the implementation, the following phases were carried out: Gathering datasets, Pre-processing them, Training the CNN model to recognize the type of crop, Training the CNN model to identify diseases, and Validating the accuracy of the model using the obtained data. There was a major drawback when images with multiple orientations were taken and classified in an uncontrolled environment.

Garima Shrestha, Deepshikha et al. [6] demonstrated a method for detecting plant diseases using a CNN-based approach. A simulated investigation and analysis utilizing image processing were done on example photographs to determine the complexity of time and the size of the infected area. The recommended strategy is successfully used to teach the system. The accuracy rate on the test set is 88.80 percent when there is no overfitting. Additionally, this technique may be leveraged in the agriculture industry to help farmers track their harvest and individuals track houseplants in their homes. In the future, an app might be developed for improved user involvement.

Thus, in order to yield better and more satisfactory results in agriculture, extensive research is always being conducted.
3 DATA DESCRIPTION

For this project we have used “potato leaves” datasets to build our “Plant leaf disease detection” system. This dataset contains various categories of healthy and diseased plants. The diseases include Tomatomosaicvirus, TargetSpot, BacterialSpot, TomatoYellowLeafCurlVirus, Lateblight, LeafMold, Earlyblight, Spidermites Two-spottedspidermite, Tomatohealthy, Septorialeafspot. For the disease detection and its treatment, we have used “Tomato Leaf Disease detection using CNN” dataset from Kaggle, which contains 3 folders (train, test, Val). The “train” folder consists of different kinds of diseases. Fig 1 shows Each disease folder contains various images of each disease. The train and Val folder are used for training model.

4 WORKFLOW

The primary objective of this project is to enable Indian farmers to enhance the quality of their crops. This project named “Cropable” detects the crop disease and recommends treatment accordingly. Using our application, the farmer can upload the image of the respective damaged crop our application will identify the disease and suggest treatment accordingly. The use of deep learning algorithms, such as Convolutional Neural Networks (CNN), makes it possible to identify crop diseases through image classification. The process of detecting diseases in crops involves two steps: the first step is to identify the specific disease affecting the crop, and the second step is to determine the appropriate remedy or treatment for that disease. This application will help farmers improve the quality of their crops. Figure 1 shows the workflow of our web app - Cropable.

5 METHODOLOGY

In order to analyse data in a valid manner, it is necessary to perform data pre-processing. The process of encoding or transforming data into a format that can be easily understood by machines is a significant aspect in the construction of neural networks or deep learning networks. This process begins with the data pre-processing stage. Mostly, it is necessary to make sure the data is readable by machine learning models and make sure it is accurate, complete, and has no duplicate, redundant, or missing data. It is essential that the updated data be free from inconsistencies and errors, and it must also be interpreted correctly. For a deep learning model to make accurate predictions/outcomes, it is crucial that the data is pre-processed.
We studied datasets of potato leaves to create a system for identifying plant leaf illnesses, and in order to receive reliable suggestions, we cleaned the data before feeding it to the model. The dataset includes a range of categories for both healthy and diseased plants. At times, there is a chance of overfitting the data for categories that have a significantly larger number of images compared to the rest. If the dataset is not balanced or the model does not perform well on new or unseen data, overfitting is a common problem in deep learning models, and it can create obstacles for achieving balanced results. Therefore, Fig 2 shows it is important to detect overfitting as the first step in addressing this issue. In such a case, it becomes important to conduct cross-validation.

To handle this situation, we have carried out pre-processing of data in the following manner:

Fig. 2. Architectural Diagram
5.1 Data Augmentation

Augmentation is a method that can artificially enlarge the size of a training dataset by producing modified data based on the existing one. In order to solve the over fitting problem, data augmentation may be used to produce more data from existing data. By inflating the data we already have, it helps to enhance the performance of the model.

If a neural network is not consistent, it may produce inaccurate predictions for the output. Thus, for a model to fit well, it is necessary to achieve a certain level of accuracy of the trained data. For our project in the case of disease detection app, we utilize image augmentation techniques. When working with TensorFlow or Keras as our deep learning framework, we have the option to create our own image augmentation pipelines or layers using the "tf.image" module or use built-in tools such as Keras preprocessing layers and ImageDataGenerator to perform image augmentation.

Using image augmentation, a small set of images can be augmented into a rich, diverse set of images for classifying images, detecting objects, or segmenting images. In image augmentation, the content of a base image is combined with the appearance of another image to produce new, high-quality images. Using the newly generated images to pre-train the existing neural network can lead to a more efficient training process. For this process, we have taken into account several factors, including flipping, rotating, scaling, and translating the crop images.

5.2 Train- Test and Validation of Data

It is important to test the accuracy of any neural network model before deploying it for any application. Once we have applied data augmentation to an existing dataset, we have partitioned the sample dataset into a training set and a separate independent test set, with the former being used for development and the latter for evaluation of the classification algorithm. By using the training set, the algorithm can learn the patterns and behaviors that are present in the data, and the testing set establishes whether the algorithm has the correct behavior. Here, we have allocated 2/3 of the original dataset as the training set and 1/3 as the testing set. The testing set is used for model validation purposes. As part of our training process, we collect validation data from a different set than the training set in order to validate our model's performance. The main reason for separating the data into a validation set is to prevent the model from overfitting. Overfitting can occur when the model becomes highly proficient at identifying samples in the training set, but is unable to accurately classify data it has never encountered before. By using a separate validation set, we can evaluate the model's performance on unseen data and prevent overfitting.

5.3 Model Implementation

We have utilized deep learning models to train the datasets after all three datasets have undergone data processing and augmentation for two tasks: crop recommendation and plant disease detection. Convolutional layers, pooling layers, and activation functions—often known as Rectified Linear Units—are the three primary building blocks of CNNs (ReLUs). The uniqueness of any architecture is defined by the number of layers used, their arrangement, and the inclusion of additional processing units. We have created a content-based model that makes use of content-based filtering technology to suggest treatments for the identified condition. AlexNet on ImageNet ResNet, VGG, and Inception architectures are commonly used forms of architectures while DenseNet or ResNetXt are newer
architecture forms. To detect plant diseases, we have utilized two deep neural network models, VGG16 and Sequential on the datasets of potato. After extracting the CNN model's features from the layers of the CNN architecture, these extracted features are then utilized for the classification of the thirty-eight different categories of healthy and diseased crop images. The CNN model is constructed as a multi-layered structure, where each layer generates a response and extracts significant features from the input image. These features are then forwarded to the subsequent layer in the model.

5.3.1 Recommendation system

We adopted a content-based approach for the recommendation system, examining solely the Content Description. This strategy advises taking into account prior decisions. When creating a model, the model first determines which dataset pairings are similar to one another, and it then utilizes the items that are the most similar to create a list of suggestions. In this methodology, the term frequency-inverse document frequency has been used to establish similarity amongst agricultural datasets (TF-IDF). We have used the inverse document frequency to denote that we are examining the rarity of each element in the collection. It has aided in providing the rare in the dataset a better ranking. We created TF-IDF vectors for each element in the dataset after computing TF and IDF. We utilized cosine similarity to compare comparable vectors. For cosine, the angle between the vectors has been determined to assess how similar the entries are.

5.3.2 Plant Disease Detection system

The CNN model in our implementation has two convolutional layers and a Max-Pooling layer. Each convolutional layer is subjected to the ReLU activation function. We favour the ReLU function in this situation because it helps our input pictures become more nonlinear, which will produce accurate prediction. A Max-Pooling layer is used between the convolutional layers with a pool size of (2, 2) to remove the maximum parameter from the input and discard the rest. To satisfy the probability density requirement, we utilize the softmax activation function in the output layer of the convolutional model to ensure that the logits sum up to 1. Layers of CNN architecture have been used to extract the CNN model’s characteristics, as shown in Fig 3. The output’s form is represented by the dense unit in the output layer. Our model has a dense unit of 38 due to the fact that there are 38 different output categories. The input shape and target size for each network in the proposed CNN model are (64, 64), (64, 64, 3), and (64, 64), respectively. The model’s performance is evaluated using the categorical cross-entropy loss function by determining the loss, and Adam optimizer. It refers to a set of particular algorithms used to modify the neural network’s characteristics, including its learning rate and weights, to minimize loss. The batch size is 32 and has been trained for 25 epochs. For the categorization of X-ray images, this CNN model employs eight pre-trained deep Convolutional Neural Networks (CNNs) that are fully integrated with keras-core in the following manner:

Sequential model The model is well-suited for a simple stack of layers where each layer has only one input tensor and one output tensor. Sequential models can include any number of layers; in this case, by utilizing the '.add()' method and the ReLU activation function, we incorporated two convolutional layers into the model. It is possible to further modify this model according to our specific requirements. This model was trained using the input shape of (64, 64, 3) which implies that data that consisted of images with a resolution of 64 pixels in width and 64 pixels in height, and with 3 color channels (red, green, and blue).

VGG16 Model The VGG16 is a pre-trained convolutional neural network model that comprises 16 layers and was developed by K. Simonyan and A. Zisserman from the
University of Oxford in 2014. 11 kernel-sized filters were used in this pretrained model, five of which were applied in the first layer and many

Fig. 3. Proposed Architecture for Crop Disease Detection

3x3 kernel-sized filters sequentially. The model has a fixed input size of 224 x 224 and it uses a stride of up to 1 while performing convolutions. Additionally, it applies a padding of 1 pixel to the input image

6 RESULTS and DISCUSSIONS

To make it easier for users to interact with the system and to integrate the entire system within the application, we developed a programme. The process begins with the user utilizing the program to capture an input image, using the programme, either manually choosing an image from the gallery or through the device camera. After choosing the image, the algorithm detects the plant illness. The technology determines if the plant is ill or not after processing the image. Two deep neural network models, Sequential and VGG16, were trained to identify plant diseases. The VGG16 model had the best accuracy, 97.53%. After a thorough examination, we put our VGG16 model to the test since it had the highest accuracy on test photos selected at random. While the majority of our system accurately identifies diseased plants within our dataset, there are times when predicting other classes, particularly healthy plants, can be difficult since we have less photos of them than other classes. We evaluated our strategy for our recommendation model using the combined dataset. The systems that are offered are founded on illness and recognition prediction. A suggestion function in the system exhibited may be used to find comparable plants that can be planted in that region.

7 FUTURE SCOPE & DISCUSSIONS

In order to make our recommendation system more effective and versatile, we plan to address the issue of suggesting crops taking into account farmers’ budgets and available farmland in future work. Our recommendation system currently uses images of the affected crops as an input and further predicts the disease followed by the treatment procedure. Currently, this approach relies on manually cropping photos taken using smartphone cameras to identify diseases. However, the technique may be improved by using cameras or
drones to observe the crops in real time. Future efforts might also focus on increasing the model’s precision and strengthening the website’s security. Additionally, we may combine the API and services, build a database, and display the user’s prior search results via an interactive dashboard.

8 CONCLUSION

The project Cropable, harnesses the power of Artificial intelligence and Machine Learning to help owners and farmers of gardens, plantations, or anyone who is interested in finding out the disease and its corresponding remedies. Through the project we aim to develop an interface which is simple and easily understandable for the user. The project will be developed in compliance with Norman’s heuristics and Schneiderman’s 8 Golden Rules. This product will be an app-based business solution used to monitor crop health. The software would improve the efficiency of crop harvesting and would help the farmers to keep track of crop health. The software would be only accessible by the farmers. The software provides a dynamic dashboard for farmers to keep reminders to water the crop and the prediction system helps the crop to be disease-free, this avoids the loss of farmers.

REFERENCES

1. Dr. K. Thangadurai , K. Padmavati, Computer Vision image Enhancement for Plant Leaves Disease Detection, 2014 World Congress on Computing and Communication Technologies, IEEE, INSPEC: 14220375
5. Garima Shrestha, Deepsikha, Majolica Das, NaiwritaDey,”Plant Disease Detection Using CNN”, 2020

19. Science, Engineering and Technology Vol. 3, Issue 2, February 2016, ISSN (Online) 2393-8021, ISSN (Print) 2394-1588


25. Dataset link:https://www.kaggle.com/kaustubhb999/tomatoleaf
