Disease Detection on Tomato Leaves Using CNN

Shruti Chaudhari, Mrunmayi Gujar, Snehal Pawar and Gauri Ghule

November 29, 2024

Diseases Detection of Tomato Leaves Using CNN

Shruti Chaudhari
Electronics And
Telecommunication Engineering
Vishwakarma Institute of
Information Technology
Pune , India
shruti.22110334@viit.ac.in

Mrunmayi Gujar
Electronics and
Telecommunication engineering
Vishwakarma Institute of
Information Technology
Pune , India
mrunmayi.22110411@viit.ac.in

Snehal Pawar
Electronics and
Telecommunication Engineering
Vishwakarma Institute of
Information Technology
Pune , India
snehal.22110005@viit.ac.in

Guided by

Prof. Gauri Ghule
dept. Electronics and
Telecommunication Engineering
Vishwakarma Institute of
Information Technolgy
Pune , India
gauri.ghule@viit.ac.in

Abstract*— *The ever-increasing population of the world has led to a shortage in raw materials and food resources. Hence, the agricultural sector has turned out to be this dominant and important source for counteracting such constraint. We will give a short overview of published solutions and concentrate on the less complex machine learning model using a traditional CNN designed as our contribution. This model of machine learning can be implemented in mobile phones, and drones and cameras that farmers can use to detect the affected crops on a large scale and take precautions measures not to allow the disease climb up high and impact supply production. Three, the paper uses the analysis of this mechanism and results obtained by model

Keywords—Machine learning; convolutional neural network ; Food security (key words**)**

I. Introduction

Diagnosis of plant diseases is both art and science A symptomatic presence characterizes many diseases and they are indicators to go for rapid diagnosis at field level. In diagnosing, identifying symptoms and signs rely on observational visual examination and intuitive judgment as much as they rely on a scientific approach. Photographic images depicting plant symptoms of disease are an important tool for improving understanding of such diseases and play a significant role in research, teaching, and diagnostics. Recently, the digital images and transfer tools have frequently been used by plant pathologists for diagnosis. At present, experts identify plant diseases manually;

however, due to limited accessibility of specific specialists in this field for farmers, consulting an expert can be very expensive. Hence, there is an increasing demand for automated systems that can detect the symptoms of plant diseases in the early stages. This early stage detection can save a farmer from losing out on significant amounts of revenue. By supporting the early detection of diseases, lowering pesticide expenses, and increasing output, technological advances can make agriculture attractive. Scientists are trying alternatives that minimize the use of hazardous chemicals, and a new approach is automating disease diagnosis as well as control. Fortunately, over the past couple of years, researchers have made significant advancements through image processing and artificial intelligence to leverage a few important technologies. Plant diseases lower agricultural productivity, and if they do not get diagnosed during initial stages of the disease cycle, food insecurity is a big threat. Due to the importance of these pathogens in agricultural practice and decision-making, early detection is crucial for developing an effective control strategy against them. Plant diseases have become one of the concerned topics in recent years. Plants that are diseased indicate their problem usually by spots or lesions on the leaves, stems, flowers and fruits. In general each disease has its own visual pattern that will aid in diagnosing the trouble. In fact, leaves are the major indicator of diseases because most symptoms initially happen in them. Generally, a farmer or an expert in this field recognizes the disease during field visits or from experience. This method, however, is subjective, takes too long and is often ineffective in identifying serial offenders.

II. Methodology

A. Data Gathering and Preparation

B. **Dataset:** The dataset containing different categories of disease for tomato leaves is downloaded and kept in a particular folder**.**

C. B. Data Preprocessing

Image Augmentation:

For the Train Set: The training data is augmented with an ImageDataGenerator. We apply various techniques such as zoom, shear and horizontal flipping to increase variation among training images to avoid overfitting. Validation Set — A fresh ImageDataGenerator instance is created for validation data that preprocesses images with the model-specific preprocess_input method.used (e.g., VGG19).

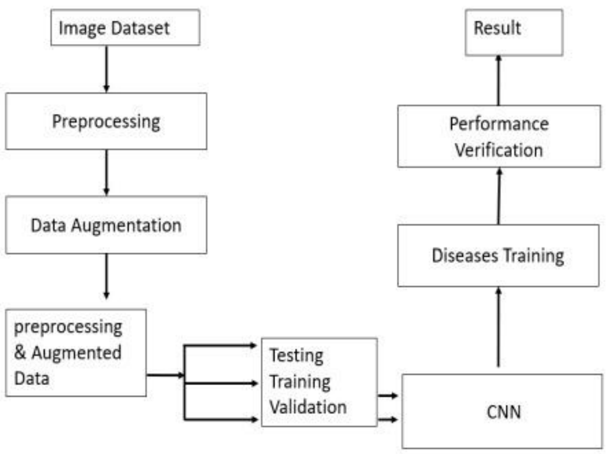


Image Rescaling: Both training and validation images are rescaled to a range of [0, 1] by dividing pixel values by 255 to normalize the input data.

D. Model Selection

Base Model*:* The base model from which transfer learning begins, with the VGG19 model pre trained and chosen because it has been efficient at image classification problems.

A**. Freezing Layers**: freeze the base model's layers so that they are not updated.

The model is trained on data ← updated during training, so it allows the model to preserve its gunth properties as well only training the classifier on the new layers.

Flatten Layer: A Flatten layer added

to squash the 3D outputs from the base

model and reshape it into a 1D array to pass into the dense layer. Output Layer: A Dense layer with n unitsn , the softmax activation which is designed to ' Trained on data until October 2023'.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 64)	0
flatten (Flatten)	(None, 186624)	0
dense (Dense)	(None, 256)	47,776,000
dense_1 (Dense)	(None, 11)	2,827

Total params: 47,798,219 (182.34 MB)

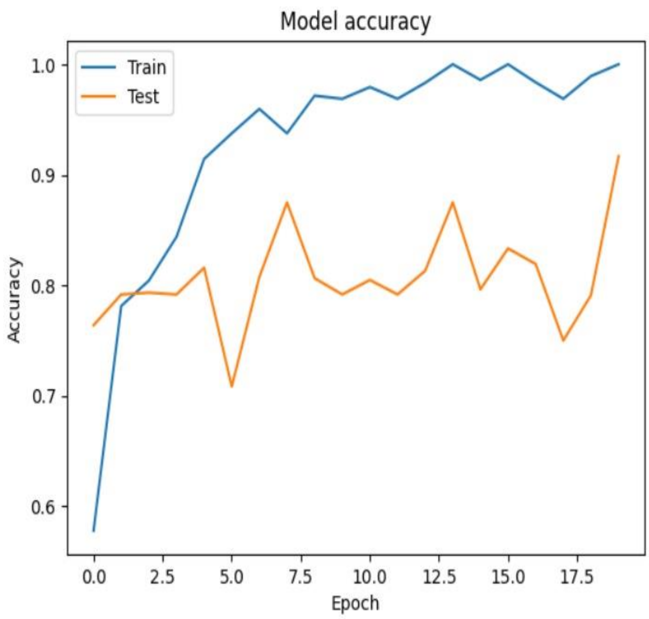
Trainable params: 47,798,219 (182.34 MB)

Non-trainable params: 0 (0.00 B)

A**. Call-backs:**

Early Stopping: To mitigate over fitting an Early Stopping call back is used to monitor validation accuracy and stop training when validation accuracy has not improved for a fixed number of epochs. This call back stores the best version of the model by saving it after every epoch if we observe an improvement in the validation accuracy. A. Plots on Performance: In the end, we will plot training and validation accuracy and loss versus epochs to visualize how our model performs throughout epochs. That aids in determining potential problems such as over fitting or under fitting.

```
Epoch 14/20
635/635 ━━━━━━━━━━━━━━━━━━━━ 1s 274us/step - accuracy: 1.0000 - loss: 0.0102 - val_accuracy: 0.8750 - val_loss: 0.8964
Epoch 15/20
635/635 ━━━━━━━━━━━━━━━━━━━━ 597s 941ms/step - accuracy: 0.9866 - loss: 0.0441 - val_accuracy: 0.7963 - val_loss: 1.0950
Epoch 16/20
635/635 ━━━━━━━━━━━━━━━━━━━━ 1s 272us/step - accuracy: 1.0000 - loss: 0.0153 - val_accuracy: 0.8333 - val_loss: 0.8798
Epoch 17/20
635/635 ━━━━━━━━━━━━━━━━━━━━ 527s 830ms/step - accuracy: 0.9855 - loss: 0.0470 - val_accuracy: 0.8192 - val_loss: 1.1405
Epoch 18/20
635/635 ━━━━━━━━━━━━━━━━━━━━ 1s 200us/step - accuracy: 0.9688 - loss: 0.0478 - val_accuracy: 0.7500 - val_loss: 2.9625
Epoch 19/20
635/635 ━━━━━━━━━━━━━━━━━━━━ 525s 827ms/step - accuracy: 0.9918 - loss: 0.0336 - val_accuracy: 0.7909 - val_loss: 1.3978
Epoch 20/20
635/635 ━━━━━━━━━━━━━━━━━━━━ 1s 206us/step - accuracy: 1.0000 - loss: 0.0102 - val_accuracy: 0.9167 - val_loss: 0.1581
```



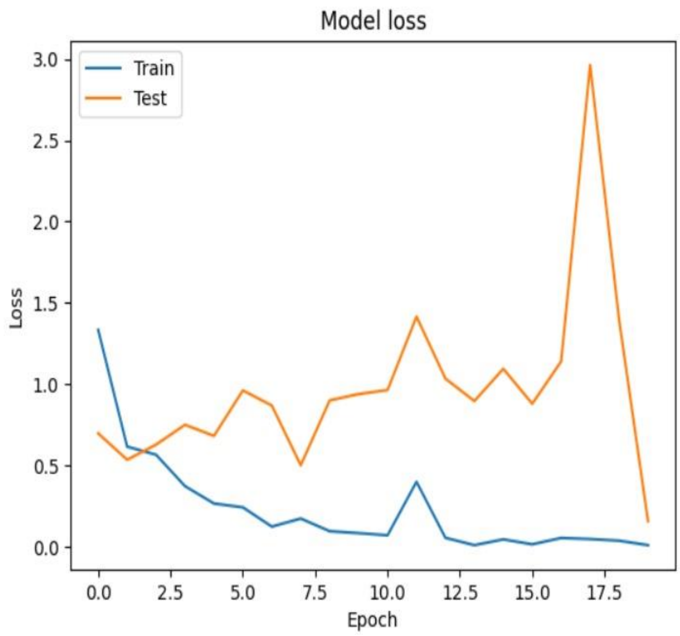


Image Prediction: A function to make predictions on new images is defined. We load our image and pre process it just like the images used during training. A. The model predicts the image's class, and the predicted label is printed.

III. LITERATURE SURVAY

The identification of plant diseases is critical to maintaining global food security. and traditional disease surveillance methods are too labor-intensive and impractical to operate a large-scale agricultural business. Abstract: Recent studies on artificial intelligence (AI), machine learning

(ML), and deep learning techniques. Learn (DL), for example, methods using visualization have reported promising results. Support vector machines (SVM) and logistic regression (LR) classifiers are known to be remarkably accurate in detecting plant diseases. There is also an artificial neural network. Convolutional Neural Networks (CNN) are gaining a lot of attention. This is especially true when the attention system and learning transfer are integrated. Despite these advances But accurately locating the disease remains a major challenge. Several studies have been conducted on other crops such as grapes, tomatoes, corn, oranges, apples, and potatoes. Many of these studies evaluate model performance using metrics such as precision, recall, precision, F1 score, and confusion matrices. In this case, the findings are exceptionally accurate. One study found an accuracy rate of 99.14% in identifying grape diseases using CNN, respectively. [1]

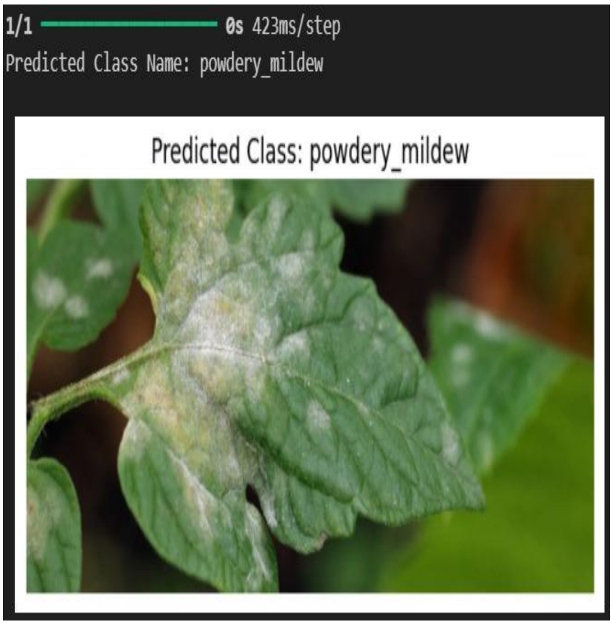
Acknowledge that Convolutional Neural Networks (CNNs) have been of primary help to plant disease identification with high precision and efficacy by using the new technology. CNNs were first used in the transfer learning of plant diseases and the results were majorly good. For example, DenseNet121 obtained the accuracy of 99.75% which was nearly perfect on the PlantVillage dataset. Other architectures that rely on ResNet and Inception have shown equally inspiring performances. To overcome computational challenges, various lightweight CNN models such as MobileNetV2 and SqueezeNext have been designed.Usage of the attention modules results in the release of additional features which, in turn, has positively influenced the classification accuracy of diseases. Thus, techniques such as multi-scale convolution kernels and coordinate attention have significantly improved the performance of detecting plant diseases. Convolutional Neural Networks (CNNs) are more capable than traditional machine learning techniques by reducing the drawbacks created by manual feature extraction. It is now possible to detect plant diseases in real-time using CNNs on mobile devices thanks to the rapid progress of technology. These CNN-based methods clearly excel in finding out various types of crop diseases from leaf spots to blights. Transfer learning involves applying pre-trained models to a particular disease of the crop and fine-tuning these models to the desired accuracy. Techniques for data augmentation aid in CNN tendency under different complex environments, in turn merging several CNNs yields even higher precision. Accordingly, CNN-based plant disease detection is a promising candidate smart agriculture technology that could enable automated diagnosis and corrective action to ultimately lead to less crop loss. [2] Several convolutional neural network (CNN) architectures have been researched in the field of identifying diseases in tomato crops, where an early diagnosis is important to increase both production and quality. The deep learning methods have been superior to the conventional machine learning methods in this area. Compact CNN architectures have been proposed to solve the computational problems. Data augmentation has been the strategy that helped models to be more resilient to different circumstances. Additionally, transfer learning has been used to change the pre-trained models for plant disease classification. Among the popular architectures used in this study are ResNet, DenseNet, and Inception. Although larger networks may overfit when trained on smaller datasets, the models used in this study were not overfitting. The new six-layer convolutional neural network (CNN) was specifically built for tomato disease classification and it got better marks than the pre-trained models. The model's performance was enhanced by the data augmentation process, which made it more robust in different environments. CNNs that are compact are among the advantages of devices with the limited processing power of the hardware that will allow sites to plant diseases in the field in real-time with mobile devices. Automated disease identification using deep learning techniques has been shown to be the main factor that allows for timely interventions to prevent crop losses, and deep learning methods have been very accurate in classifying plant diseases. [3]Stain disease on sugarcane leaves is one of the major hurdles in the sugar industry, identifying it on time, however, becomes very important for its growth and rural mass's economic development. Non-availability of a timely diagnosis can make the farm production go down due to a decline in crop yield and thus lowered production. To overcome this problem, the researchers employed Speeded-Up Robust Features (SURF) as a tool to extract image features and Fast Library Approximated Nearest Neighbor (FLANN) for the final decision. This study highlights an application that is capable of recognizing different forms of leaf stain disease and providing treatments to farmers resulting in considerable benefit for sugarcane growers. Various previous studies have proven SURF to be very efficient in tasks including plant image recognition, matching, and traffic sign detection, with an 89% accuracy rate of 92%. These results thus verify the procedure's ability to identify leaf stain disease. Future investigation is required to test the abilities of SURF and FLANN vis-a-vis other techniques in the disease identification. [4] This study introduces a Convolutional Neural Network (CNN) as a model that is able to detect the leaf diseases of corn automatically, which has been one of the problems of agriculture in India. The model trained with Plant Village Dataset has a success rate of sensational level and brings a reliable and affordable solution to the farmers. The CNN model with a 98.78% accuracy in classifying the three types of corn diseases has been the most successful so far. The further optimizations have not only increased the model's accuracy, but also its efficiency, and thus, this development of digital agriculture has helped farmers to produce more corn. [5] Automating the detection of sugarcane diseases is essential for ensuring food security and supporting the livelihoods of small-scale farmers. Traditional manual detection methods are often inefficient and time-consuming. This research employs

Convolutional Neural Networks (CNNs) for the early identification of diseases in sugarcane crops. [6] Blight diseases including late blight, bacterial spot, septoria leaf spot, and yellow curved leaf disease can produce monetary losses and constitute a menace to tomato crop quality. This study, therefore, purports the discovery of a convolutional neural network (CNN) model along with the learning vector quantization (LVQ) algorithm in doing both classification and detection of the leaf diseases of the tomatoes. The dataset of 500 images of tomato leaves implies that each leaf is affected by one of the four diseases. The Heatmap Network captures and classifies the CNN's automatic features by using a segmented logo function of filter objects to gain the RGB input. The LVQ learns from the output feature vector of the convolution layer of CNN. The experimental results show that this method correctly cross-references and confirms the fact that it is the only one showing the four types of leaf diseases in tomato. [7]Surampalli Ashok et al's method has an accuracy of 98% that uses a CNN algorithm for hierarchical feature extraction. This maps pixel intensity to an input image and then compares it with the previously trained dataset. Slightly a minimization of error in training and optimization of adjusting parameters of the leaf section. The images are then classified as diseased or healthy by application of image classification techniques in addition to artificial neural networks. Fuzzy logic And hybrid algorithms can be added to the disease detection process.[8]R. Venkatesan et al. hybrid deep learning architecture, targeting the early detection of tomato leaf diseases. Deep learning models especially the neural networks classifies and yields more accuracy because deep learning models will automatically do several feature extraction stages, providing advantages over other traditional classification methods. Models like AlexNet and GoogleNet come up for better identification of disease along with, it speedup in classification processes. Although the earlier models such as LeNet and the LVQ algorithm have delivered good results, they are limited in that they are unable to detect early disease indications. [9] Huiqun Hong et al. compared five different CNNs, all having different parameters and average accuracy levels in the study on deep learning-based tomato disease detection and classification. The models compared are ResNet50, Xception, MobileNet, ShuffleNet, and DenseNet_Xception. The network architectures were compared with different learning rates. In this evaluation, the researcher found that DenseNet Xception holds 97.10% accuracy although it has used the most number of parameters. In contrast, ShuffleNet demonstrated an accuracy of 83.68% while requiring much fewer parameters.[10]

IV. RESULT

The suggested technique showed a great accuracy of 80.72%. The intricacies found in the pictures are explored using CNN (Convolutional Neural

Networks) algorithm which determines the correspondences of pixel intensities of the input images to the trained dataset and thus, mapping and comparing are done. The shown architecture of CNN includes three convolutional layers and max-pooling layers, every one of which gets a different number of filters. We conclude that the current models are the best ones..



V. Conclusion

This study presents a novel system aimed at identifying diseases in tomato leaves, with the goal of enhancing tomato yield in the global agricultural industry. Through this proposed approach, we assert that an effective model has been created for the detection of diseases affecting tomato leaves. The integration of image processing techniques with Convolutional Neural Network (CNN) algorithms has gained widespread acceptance and inspired us to investigate this emerging technology. This research meticulously analyzes diseased tomato leaves, ensuring a detection system that offers greater accuracy compared to current methods.

VI. References

[1] D. H, G. E. O and K. M, "Disease detection on the leaves of the tomato plants by using deep learning," in *International conference on agro-geoinformatics*, 2017.

[2] P. H, E. J. S and . K. S. H, "Image-based disease diagnosing and predicting of the crops through the deep learning mechanism," in *international conference on information and communication technology convergence (ICTC)*, 2017.

[3] R. M, . W. M. R, J. N and I. A. B, "Detection and classification of leaf disease using artificial neural network," *International journal of technical research and applications*, vol. 3, no. 3, pp. 331-333, 2015.

[4] W. S, P. M and B. C, "Soybean plant disease identification using convolutional neural network," in *The thirty-first international flairs conference*, 2018.

[5] I. S and R. M, "Leaf disease classification using artificial neural network," *Jurnal Teknologi*, vol. 17, no. 1, p. 77, 2015.

[6] M. V. S. S. P. K. M. K. P. N. C. &. K. N. Shobana, "Plant disease detection using convolution neural network," in *International Conference on Computer Communication and Informatics (ICCCI)*, 2022.

[7] S. K. G. R. V. S. S. S. S. G. &. P. B. Ashok, "Tomato leaf disease detection using deep learning techniques," in *5th International Conference on Communication and Electronics Systems (ICCES)*, 2020.

[8] H. J. L. a. F. H. Hong, "Tomato disease detection and

classification by deep learning," in *International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2020.

[9] H. E. David, "Literature review of disease detection in tomato leaf using deep learning techniques," in *7th International conference on advanced computing and communication systems (ICACCS)*, 2021.

[10] S. Ashok, "Tomato leaf disease detection using deep learning techniques," in *5th International Conference on Communication and Electronics Systems (ICCES)*, 2020.