Apple Disease Expert

Norin Hossam
Computer Engineering
AAST
Alexandria, Egypt

Zyad Shrin
Computer Engineering
AAST
Alexandria, Egypt

Habiba Ahmed
Computer Engineering
AAST
Alexandria, Egypt

Farah Atef
Computer Engineering
AAST
Alexandria, Egypt

ABSTRACT

Yearly, there is a significant loss in financials of the apple industry due to diseases and pests. Conventionally, disease classification is carried out by farmers exploring the crops to see if there are any visible diseases making it expensive and time-consuming. Images chosen for this study were apple scab, rot, and rust, as well as images of healthy apples. All models we propose performed better than 90% on testing images.

INTRODUCTION

Since apple plants are one of the most grown trees. And, since apples are extremely susceptible to disease, it was essential to explore it. Identifying diseases is difficult eve in the eyes of an expert. And so the need for an efficient system for early detection of diseases was obvious.

Early disease identification can be critical in preventing crop and plantations loss. Recent enhancement in the field of automated plant disease detection have proven to enhance production especially in massive apple farms.

To achieve automated the apple disease diagnosis and detection we first need to discover the medical condition and figure out its type. So, a classification model is used to determine the specific disease that has affected the crop from its leaves.

RELATED WORK

Shivaram Dubey and Anand Singh Jalal (2012) discussed three apple illnesses in their study: apple scab, apple rot, and apple blotch. The primary step in detecting these diseases using image processing and analysis is image segmentation through K-mean clustering. The next step involves feature extraction from the segmented images. These features include the global color histogram (GCH), color coherence vector (CCV), local binary pattern (LBP), and complete local binary pattern (CLBP). Among these features, the complete local binary pattern achieves a 93% correct classification accuracy, outperforming the others, as it considers the sign, magnitude, and center value of every pixel.

Mr. Abhijeet and Prof. A.P. Patil's research paper by (2017), a solution was proposed for detecting diseases in apples using the k-means clustering and Learning Vector Quantization neural network techniques. The approach involved numerous steps, starting with data preparation and preprocessing, to image segmentation, followed by feature extraction. Finally, a neural network was used to train and test the model. The results of the study indicated that the algorithm could achieve a recognition rate of above 95% accuracy to identify apple plant diseases.

Yan et al. presented a model used for identifying diseases in apple plant leaves using images. The proposed system successfully classified three common illnesses, scab, frog-eye spot, and cedar rust. The method used the enhanced Convolutional Neural Network (CNN) that utilized both the VGG16 and ResNet50 networks.

PROPOSED MODEL

Two different classification models have been proposed for this dataset. These are as follows:
ResNet50: ResNet, short for Residual Network, is a type of convolutional neural network (CNN) consisting of 50 layers, including 48 convolutional layers, one max pool layer, and one average pool layer.
Vgg16: VGG16, also known as VGGNet is a CNN model that supports 16 layers.
These models were chosen based on a thorough review and exploration of numerous research papers, which consistently indicate that both models are the most effective ones for classifying apple diseases.

Experimental Work

Plant Disease Expert Dataset

The Plant Disease Expert dataset, which can be accessed and downloaded from Kaggle (https://www.kaggle.com/datasets/sadmansakibmahi/plant-disease-expert), is an image dataset used for plant disease detection. It includes more than 200,000 images of healthy and diseased plants divided over 58 separate classes. Only apple crops were considered in this study, with a total of 18600 images representing 3 different apple plant diseases, as well as healthy apples. Figure 1 shows a picture from every class label of apple leaf illness in the database.

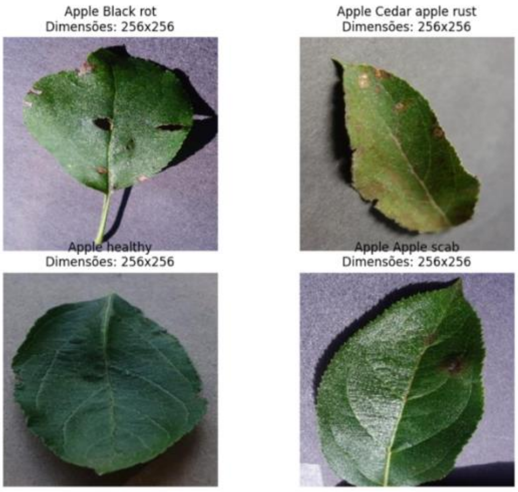

Figure 1

Table 1. The distribution of pictures among each class label of the Plant Disease Expert dataset.

Table 1

	Class Name	Number of images
1	Apple Cedar apple rust	2640
2	Apple healthy	3948
0	Apple Black rot	5964
3	Apple Apple scab	6048

Figure 2 is a visualization of the proportion of each class in the dataset.

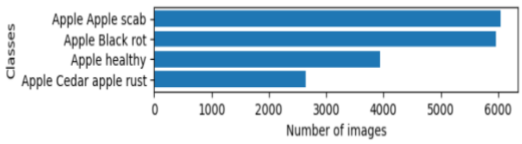

Figure 2

Evaluation Metrics

To achieve a comprehensive understanding of our model we used a multitude of evaluation techniques. First metric we used though the training and testing process was the accuracy. It calculates the ration of correctly predicted instances to the total number of instances to see how the model preforms.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

We also used a cost function throughout the process. Which measures the difference between the predicted output and the actual target values. This function was used to achieve regularization preventing overfitting during model training.

For visualization we used a graph representing the variation versus validation to understand the way our model performed during training across epochs. Typically shows the Improvement Over Epochs and if the model overfits or underfits. Figure 4 shows the training and validation graphs for our ResNet50 with transfer learning model. Figure 5 is for the VGG16 model with transfer learning.

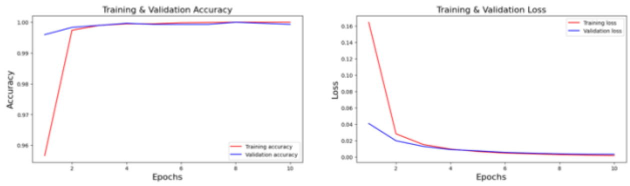

Figure 3: training and validation curves ResNet50

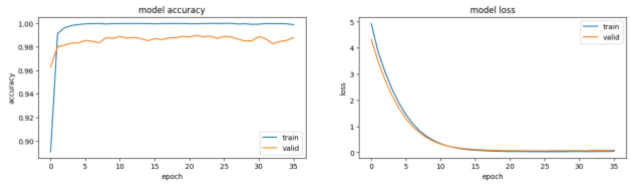

Figure 4: training and validation curves VGG16

Another visualization method we implemented was the confusion matrix. A confusion matrix is a tabular representation of the performance of a classification model, showing the counts of true positive, true negative, false positive, and false negative predictions on a set of data. It can be used to calculate other metrics i.e. Precision, Recall, and F1 Score.

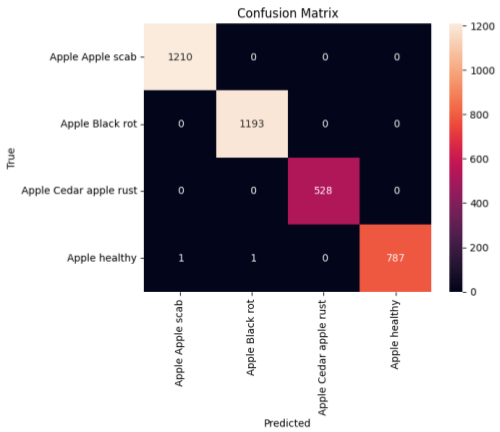

Figure 5: confusion matrix RestNet50

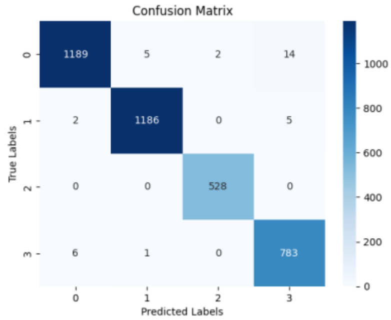

Figure 6: confusion matrix VGG16

Results

Table 2. shows the performance classification accuracy (%) of transfer learning (TL) with ResNet50, and VGG16, using k-fold cross validation, and hold-out cross validation.

Table 2

	Best train accuracy	best train loss	best test accuracy	best test loss
TL-ResNet50-Kfold	99.95%	6.97%	99.856%	0.30%
TL-ResNet50-holdout	99.97%	3.325%	99.95%	1.342%
TL-VGG16-Kfold	98.76%	4.553%	99.76%	7.315%
TL-VGG16-holdout	99.96%	10.985%	99.3%	6.529%

Conclusion

The goal of the "Apple disease expert" project is to classify apple diseases. This research utilizes two deep learning algorithms, ResNet50 and Vgg16. To assess the performance of these models, we computed the accuracy score and created a confusion matrix. The main limitation of our work was the already augmented dataset. We improved our results by employing various types of cross-validation and adjusting hyperparameters. The findings indicate that ResNet50 performs better than Vgg16, achieving an accuracy of 99.95%.

References

[1] Miss.Kambale, Anuradha Manik, Dr.Mrs.Chougule "GRADING OF APPLE FRUIT DISEASE", ISSN: 2277-9655 (I2OR), Publication Impact Factor: 3.785

[2] Uravashi Solanki, Udesang K. Jaliya and Darshak G. Thakore," A Survey on Detection of Disease and Fruit Grading", International Journal of Innovative and Emerging Research in Engineering Volume 2, Issue 2, 2015.

[3] Q. Yan, B. Yang, W. Wang, B. Wang, P. Chen, and J. Zhang, "Apple leaf diseases recognition based on an improved convolutional neural network," Sensors, vol. 20, no. 12, p. 3535, 2020.

[4] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255).