TOMATO LEAF DISEASE DETECTION

TEAM

TEAM NO . 34

TIME . 10:00 AM

DAY . SAT 6/5/2023

ENG. EMAN ISMAIL

ID	الاسم
202000114	إسراء شعبان السيد
202000155	أمنية عرفات فهمي
202000183	آية بهجت عبد السلام
202000307	رانا كمال أحمد البدري
202000487	عائشة محمد أحمد عطيه
202001003	نور خالد علي سويفي

PAPER

ABSTRACT

Tomato disease control remains a major challenge in the agriculture sector. Early-stage recognition of these diseases is critical to reduce pesticide usage and mitigate economic losses.

While many research works have been inspired by the success of deep learning in computer vision to improve the performance of recognition systems for crop diseases, few of these studies optimised the deep learning models to generalise their findings to practical use in the field. In this work, we proposed a model for identifying tomato leaf diseases based on in-house data and public tomato leaf image databases. Three deep-learning network architectures (VGG16, Inception_v3, and Resnet50) were trained and tested. We packaged the trained model into an Android application named TomatoGuard to identify nine kinds of tomato leaf diseases and healthy tomato leaves. The results showed that TomatoGuard could be adopted as a model for identifying tomato diseases with a 99% test accuracy, showing significantly better performance than APP Plantix, a widely used APP for general-purpose plant disease detection.

PAPER OVERVIEW

INTRODUCTION

MATERIALS AND METHODS

RESULTS

DISCUSSION

CONCLUSIONS

REFERENCES

INTRODUCTION

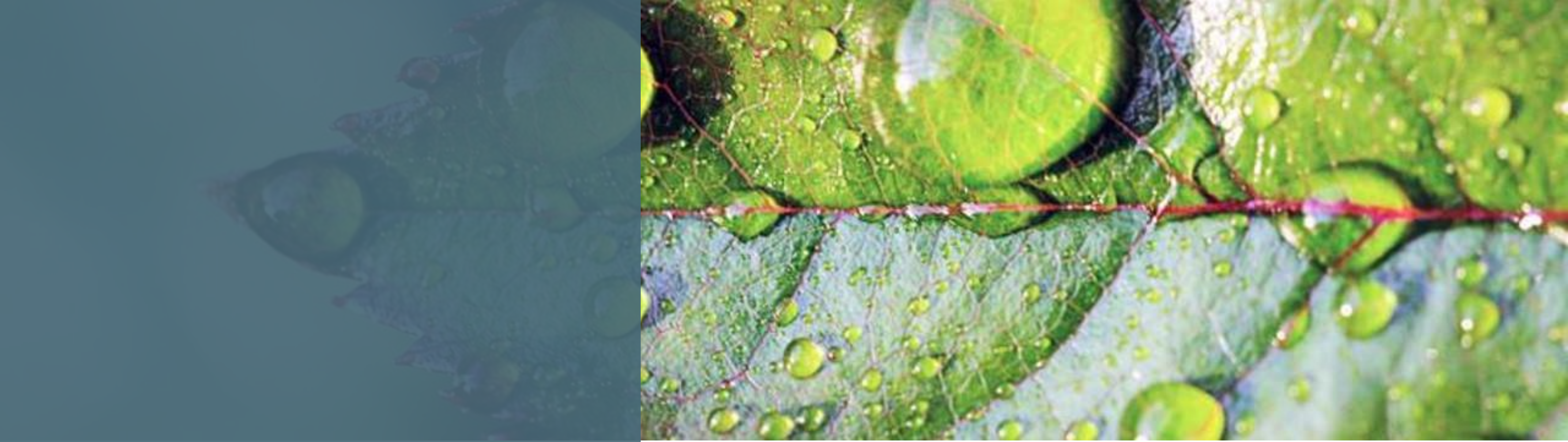
Food and Agriculture Organization of the United Nations

According to the FAO (The Food and Agriculture Organization of the United Nations, FAOSTAT), China was the largest producer of tomatoes from 2013 to 2017. Diseases in tomato leaves cause major production and economic losses by reducing both the quality and quantity of productivity in the tomato industry. Moreover, crop disease control issues are closely related to the development of sustainable agriculture. In China, crop disease diagnosis still heavily relies on the experiences of farmers or crop producers. Excessive use of pesticides leads to long-term drug resistance of the bacteria, which severely impairs the disease resistance in the crop To reduce the use of pesticides, the timely and accurate detection of tomato leaf disease in the early stages

Traditionally, identification of the type and the severity of the plant disease mainly relied on experienced farmers or trained experts .However, the process is not ideal, and difficult to meet the requirements of modern agriculture because of low efficiency, small scope, and poor real-time performance. With the rapid development of machine learning algorithms in computer vision, especially deep artificial neural network technology, it is now possible to perform an automatic, timely, and accurate diagnosis of crop diseases with the support of massive agricultural information data .Researchers have used Bayesian classifiers, support vector machines, and artificial neural networks to develop several expert diagnosis systems for disease detection in various crops. Intelligently identifying crop diseases has achieved reasonable success in many specific cases. The traditional machine learning methods consist of image preprocessing, feature extraction, and classifier training has become a critical and urgent challenge to tackle.

Among these three steps, the feature extraction step is the key challenge. Traditionally, features were extracted based on prior knowledge to obtain the image's color, shape, texture, and other characteristics, such as scale-invariant feature transform and histogram of oriented gradient. Generally, the extracted features do not generalize into new images for universality which is one of the main reasons that the developments currently stay in the laboratory phase and cannot be applied in practice. More recently, the emerging deep Convolution Neural Network (CNN) generated via reference to the structure of the visual system has performed very well in image feature extraction. Researchers have achieved excellent performances in various types of image recognition tasks. Additionally, Fuentes et al. (2017) presented a deep-learning based approach to detect diseases and pests in tomato plants, and the mean average precision for their whole system showed a performance of more than 80% for the best cases, but they only used their private dataset from areas of the Korean Peninsula and did not provide an open source. Generally, most of these studies are mainly limited to their curated data and have not yet demonstrated the application in practice.

This work aimed to provide a solution for early control of tomato diseases from the perspective of image recognition. For this purpose, we have comprehensively considered data acquisition, model optimization, result analysis, and application deployment. The specific contributions of this paper



MATERIALS AND METHODS

1) Identification and control methodology of various diseases of tomato leaves
2) In-house and public tomato image datasets
3) Image augmentation
4) Deep convolution neural network and transfer learning

IDENTIFICATION AND CONTROL METHODOLOGY OF VARIOUS DISEASES OF TOMATO LEAVES

Tomato diseases are caused by various factors including fungi, bacteria, virus, mite (Chowdhury et al., 2021).

1. Fungi are the predominant plant pathogens, and can cause multiple diseases, including early blight, septoria leaf spot, target spot, and leaf mould.
2. Bacteria are also major plant pathogens. Bacterial spot is a plant disease caused by bacteria. First, dark brown water-soaked spots appear on the leaves; later, these spots become blackish, and eventually, the affected tissue drops out, leaving a hole in the leaf. Copper sprays provide some control. Good sanitation practices, including prompt plough-down of stubble and weed control, help prevent the disease.
3. Moulds are also a major cause of plant diseases. The characteristics of late blight of tomato plants caused by mould are that lesions on leaves appear as large water-soaked areas that eventually turn brown and papery. Fruit lesions are large irregular greenish brown patches having a greasy rough appearance. Green to black irregular lesions are also present on the stems.
4. Mites are a common pest that attacks vegetable and fruit crops . Tomato plants attacked by mites often have a mottled or speckled dull appearance on the top leaf surfaces due to feeding injury . Leaves then turn yellow and drop. Large populations produce visible webbing that can completely cover the leaves. The use of overhead-sprinkler irrigation may provide short-term relief for mite infestations .

septoria

Bacteria

IN-HOUSE AND PUBLIC TOMATO IMAGE DATASETS

THE FIRST DATASET

The first dataset was acquired from the College of Water Conservancy and Civil Engineering greenhouse at South China Agricultural University. The tomato seedlings were inoculated with tomato yellow leaf curl virus (TYLCV) and tomato mosaic virus (TMV) at the Plant Protection Research Institute of Guangdong Academy of Agricultural Sciences, then transplanted to the conservatory .Tomato was raised after seven-day acclimatization. During the period of growth, the Hoagland nutrient solution was two days by using a sprinkling can to provide sufficient nutrients for the plants. Tomato images were collected as the dataset during the growth cycle by a digital camera (SONY DSC-HX400), which was set to adjust the focal length and aperture, auto white balance, and without flash. The image resolution was 5184×3888 pixels. To maintain the diversity of the data for the test of the generalization ability of the algorithm, tomato images were captured in all the weather conditions.

THE SECOND DATASET

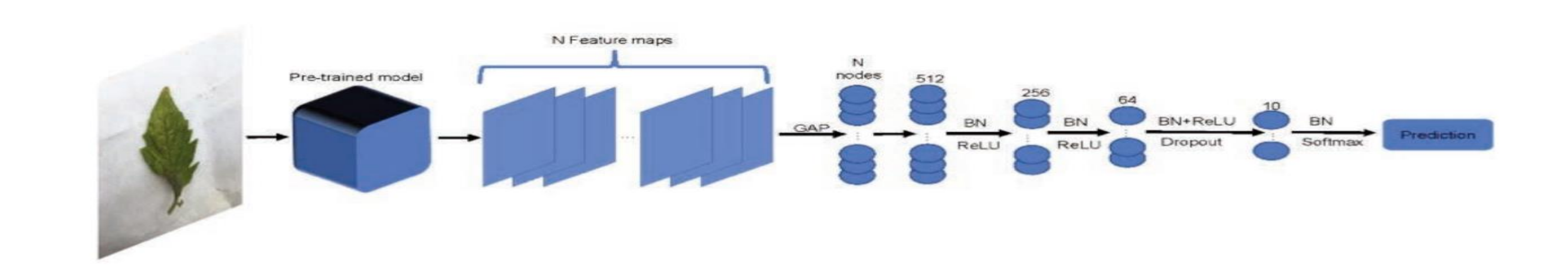
The second dataset, Plant Village data (Hughes et al., 2015), was an open-access dataset with more than 50,000 images of leaves, from which we extracted 18,160 images of 10 different tomato leaf classes. These images' size was 256×256. To complement the in-house dataset and enhance data diversity, we also created the third dataset as a supplement and named it the Internet dataset. We downloaded 295 images from multi-sources, such as the Arkansas Plant Diseases Database, the American Phytopathological Society database, the Bug wood image database, and several other academic sources. So, these images' sizes varied from 35×47 to 2156×2232. The sampled images of these three datasets are shown in Figure 1. The differences among the images from these three sources included the shooting environment, the growing season, and the growing location. The specific type of diseases and quantities of the images in these

IMAGE AUGMENTATION

Accurate and comprehensive annotation of the images for training and validation was crucial for developing an appropriate and reliable detecting model because deep learning models tend to fit the labelled samples for learning the features during the training. Thus, the ideal training images should be photographed from various angles, scales, and scopes, which was difficult to achieve in actual shooting situations because of the photograph restrictions .
Furthermore, the lighting condition was another disturbance. Many factors interfered with the lighting condition, such as the weather, occlusion between leaves, shadows, and the disturbance of sand and dust. Only using these disturbing images as training data could lead to the over-fitting problem of deep learning models . Besides, both the in-house and Internet datasets were much smaller than the PlantVillage dataset. The augmented images were used to enrich the dataset at the experimental phase to expand the data amount and better generalization .
Image preprocessing is a common technique for enhancing data .
This technique involves removing low-frequency background noise, normalizing the intensity of images of individual particles, removing reflections, and masking portions of images. In this work, we eliminated the interference of the shooting position by randomly rotating within 90-degree angles, flipping, and mirroring. We also shifted the brightness value to eliminate the problem of uneven illumination. In addition, cropping the original image randomly also augmented the original data.

DEEP CONVOLUTION NEURAL NETWORK AND TRANSFER LEARNING

CNN has been developed explosively since its successful application in classifying the MNSIT dataset. The close connection and spatial information between layers in CNN make it particularly suitable for image processing and understanding. Deep learning is an end-to-end approach. Researchers have proven that CNN can automatically extract rich, relevant features from images. Given a large amount of input image data and output labels, a CNN model automatically learns the features in the data. The learned features are effective because the data annotation is accurate. For a long time, increasing the hidden layer was a common strategy to improve the performance of the networks, and there are many excellent models which have been proposed based on that, such as AlexNet , VGG16, Inception_v3, ResNet50 and DenseNet . According to literature , six state-of-the art architectures AlexNet, DenseNet-169, Inception_v3, ResNet50, SqueezeNet-1.1, and VGG13) were trained on the PlantVillage dataset. Inception_v3 gave the best accuracy for the deep training strategy. The inception in Inception_v3 is the most prominent characteristic. Its core idea is factorization, replacing the big convolutional kernel with multiple smaller kernels. The inception mechanism not only reduces calculations but also improves feature extraction capabilities. However, when the network depth increases, the network accuracy becomes saturated or even decreases; this problem is called the degradation problem. ResNet was developed to solve this problem; it can achieve a deeper network without causing the gradient to disappear or the gradient explosion problem. VGG16 is one of the most utilized classical sequential networks.



RESULTS

DEEP CONVOLUTION NEURAL NETWORK CAN ACCURATELY CLASSIFY TOMATO LEAF DISEASE IMAGES

FEATURE VISUALIZATION

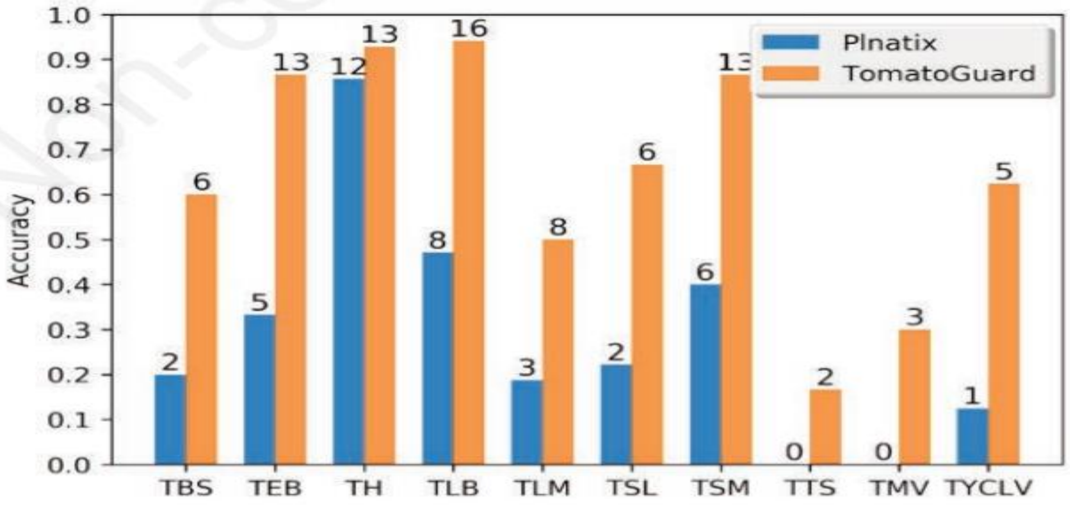
TOMATOGUARD APPLICATION

The training proceeded on the training set; after that, the evaluation in each epoch was performed on the validation set, and the final evaluation was done on the testing set. The validation set is a technique used for minimizing over-fitting and is a typical way to control training processing.
The network was trained using mini-batch stochastic gradient descent with a momentum factor. The number of samples per small batch was 32, and the momentum factor was set to a fixed value of 0.9. The initialization of the weight affected the convergence speed of the network. In this work, the Glo rot uniform initializer was used to initialize the weights of all network layers. The biases of all convolutional layers and fully connected layers were initialized to 0. The same learning rate was adopted for all layers in the network, and the initial learning rate was set to 0.001 and decayed 1e-6 for each training epoch. All the processes of training and testing the tomato disease identification model described in this work were implemented on one machine, whose configuration parameters were Intel Core i9- 7900X 3.3 GHz Processor, an Nvidia GeForce 2080Ti GPU, and 11GB memory. Observing the training process curve, we found that the validation loss converged to a small value and did not change in many epochs before the minimum validation loss was obtained. So, we set up the early stop strategy to avoid over-fitting. Once the validation loss does not change in 30 epochs, stop the training. As we can see from Table 3, an early stop strategy can significantly reduce model training time. Compared with the regular strategy, the early stop strategy ended the training about 100 epochs in advance and kept the test accuracy high. Finally, we choose the Inception_v3 model to package an application.

DEEP CONVOLUTION NEURAL NETWORK CAN ACCURATELY CLASSIFY TOMATO LEAF DISEASE IMAGES

For most of the similar research, they ended their work after completing the model training. To move forward a single step, applying the findings in practice, we deployed the proposed deep learning model into Android devices. Nowadays, the smartphone has become indispensable tool in people’s life.
Android system has been widely used in smartphones and tablet computers. To implement our research, deploying the trained deep learning model in a smartphone is meaningful. We use Android Studio to deploy our model into the Android device. To optimize the model for smartphone applications, we compressed our trained model with TensorFlow lite and compared it. Compared with regular model packaging applications, compressed model applications not only had smaller installation packages but also had faster calculation speeds. However, the test accuracy was slightly reduced. 12.6% reduction in installation package size in exchange for 8.3% accuracy reduction. Since both operation speeds were at the millisecond level, accuracy should be prioritized when hardware conditions permit. Our goal is to provide an immediate and effective tool for people who want to recognize tomato leaf diseases. So, the process of our application is inputting a tomato leaf image, computing it by the advanced trained model, and displaying the result of recognition. We developed an Android APP named TomatoGuard. It can identify 9 kinds of tomato leaf diseases from healthy tomato leaves. It claimed that it could detect more than 200 diseases covered from more than 40 crops, including 10 kinds of tomato leaf images discussed in this work. However, since some of the test set in this work was augmented from some of the training set, the test set was unsuitable for testing Plantix. So, we collected 127 more new images, including 10 classes mentioned above from multi-sources for comparing with Plantix strictly.

TOMATOGUARD APPLICATION



The symptoms of tomato diseases are various. Some disease characteristics are spots on the leaf or/and the edge of the leaf. Some diseases infect the whole leaf, it which will change the leaf color, shape, and texture, such as TYLCV. These diseases' characterization is a global feature. These tomato disease diagnoses depend on the entire leaf. Therefore, it is unsuitable to localize these disease parts on the leaf. We visualised the model output layer to analyse how the neural network works with an image. First, we generated 10 arbitrary images and updated them until we maximised each node activation of the network's last layer. These images represented the feature of each class exacted from this model and allowed us to understand what sort of input patterns activate a particular filter, as shown in Figure 6. Distinctly, most of these features do not conform to any observation in the human sense. This is because the CNN classifies images by decomposing the visual input space into a hierarchical-modular network of convolution filters mapping the probabilities between certain combinations of these filters and a set of arbitrary labels. But some class feature maps made sense, such as tomato bacterial spot (TBS) feature map containing multiple prominent bright spots, similar to TBS pathological symptoms.

FEATURE VISUALIZATION

DISCUSSION

Because of the complexity of the patterns shown in each class, especially in terms of infection status and background, the model tends to be confused with several classes. Figure 7 presents a confusion matrix of the Inception_v3 test results. Based on the confusion matrix, we can evaluate the classifier's performance and determine what classes were more highlighted by the neurons in the network. Furthermore, it helped us analyze further

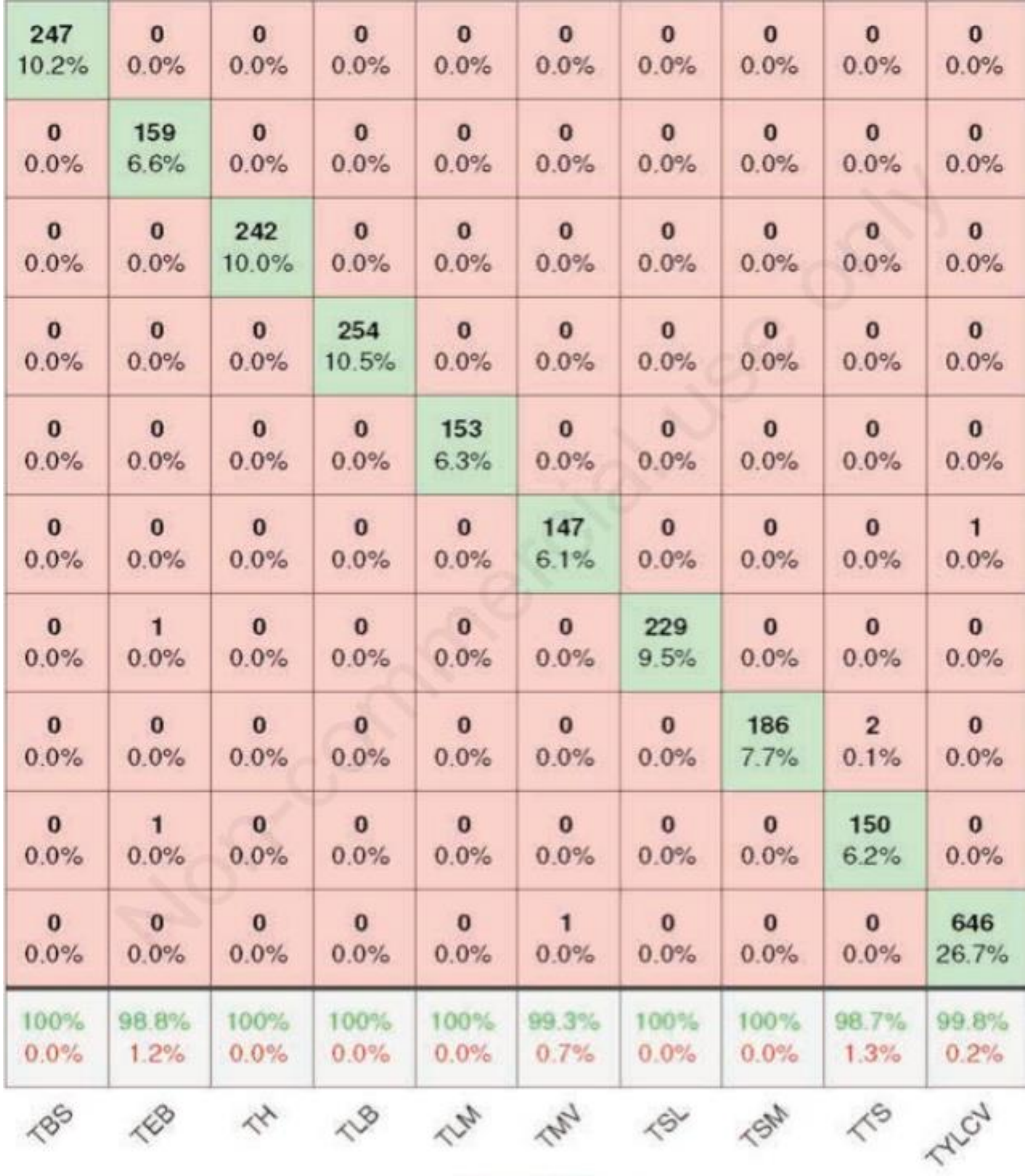
ther procedures to avoid inter-class confusion. For instance, 2 tomato target spot (TTS) images were being incorrectly classified into tomato spider mites (TSM). The TSM class precision (98.9%) was shown to be the lowest. These two classes have shown relatively strongly missed up. This was related to their disease symptoms. TSM is not a virus disease, spider mites cause it. Mostly mites live on the underside of leaves and feed by piercing leaf tissue and sucking up plant fluids. Feeding marks show up as light dots on the leaves. As feeding continues, the leaves turn yellow .

DISCUSSION

TTS disease first appears as small, necrotic lesions with light-brown centers and dark margins on tomato leaves. In brief, these diseases have common symptoms: they all have spots (marks) on the leaf and change the color to yellow. Moreover, two tomato early blight (TEB) images were misidentified as TTS (1) and tomato septoria leaf spot (TSL) (1). Comparing TEB, TTS, and TSL, they all have similar disease spots. TEB causes small dark spots to form on older foliage near the ground initially. Leaf spots are round and brown and can grow up to a half-inch in diameter. Larger spots have a target like concentric rings, and tissue around spots often turns yellow .

Similarly, spots caused by TSL are circular with dark brown margins and tan to gray centers with small black fruiting structures. Characteristically, there are many spots per leaf. If leaf lesions are numerous, the leaves turn slightly yellow, brown, and wither. When these diseases’ unique characteristics are not obvious, there is the possibility of being confused. There are also 1 tomato mosaic virus (TMV) image classified as TYLCV and 1 TYLCV image classified as TMV. Considering the dense tomato growth environment of the in-house dataset, tomato plants may be cross-infected, which means there are probably two disease features shown on one leaf. Overfitting is the most prone problem in deep neural networks. Adding the dropout and batch normalization layers (BN) to the model is a very effective way to solve the problem of overfitting .

The idea of the dropout layer can be described very simply: randomly drop units (along with their connections) from the neural network during training .



TBS	TEB	TH	TLB	TLM	TMV	TSL	TSM	TTS	TYLCV
247 10.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
0 0.0%	159 6.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
0 0.0%	0 0.0%	242 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
0 0.0%	0 0.0%	0 0.0%	254 10.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	153 6.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	147 6.1%	0 0.0%	0 0.0%	0 0.0%	1 0.0%
0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	229 9.5%	0 0.0%	0 0.0%	0 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	186 7.7%	2 0.1%	0 0.0%
0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	150 6.2%	0 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	646 26.7%
100% 0.0%	98.8% 1.2%	100% 0.0%	100% 0.0%	100% 0.0%	99.3% 0.7%	100% 0.0%	100% 0.0%	98.7% 1.3%	99.8% 0.2%

Target Class

CONCLUSIONS

This work presents a deep-learning-based classifier for nine tomato diseases and healthy tomato leaf recognition, which achieved 99% test accuracy. We also developed a tool named TomatoGuard on Android devices to help people recognize tomato diseases. The experiment results showed that TomatoGuard dramatically overcame the state-of-the-art of this field as it recorded a higher test accuracy resulting in higher performance than the APP Plantix. We expect this tool to significantly contribute to the crop protection research area. Although TomatoGuard stays at the TRL5 level, to make this tool more applicable, future studies will need to detect the diseased crop planted on site, and the application needs to optimize interaction logic and supplement more functions.

REFERENCES

Journal of Agricultural Engineering 2023; volume LIV:1432

Tomato leaf diseases recognition based on deep convolutional neural networks Kai Tian,1 Jiefeng Zeng,2 Tianci Song,3 Zhuliu Li,3 Asenso Evans,4 Jiuhao Li4 1National Supercomputing Center in Shenzhen, Shenzhen, Guangdong, China; 2College of Water Conservancy and Civil Engineering, China Agricultural University, Beijing, China; 3Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, Minnesota, United States of America; 4College of Water Conservancy and Civil Engineering, South China Agricultural University, Guangzhou, China

THANK YOU