Plant Disease Detection using Machine Learning

Authors's Name/s Shrey Sharma
Department of Electronics and Communication
Raj Kumar Goel Institute Of Technology
Ghaziabad,India

Authors Name/s Tushar
Department of Electronics and Communication
Raj Kumar Goel Institute Of Technology
Ghaziabad,India

Abstract— Plant disease is an ongoing challenge for smallholder farmers, which threatens income and food security. The recent revolution in smartphone penetration and computer vision models has created an opportunity for image classification in agriculture. Convolutional Neural Networks (CNNs) are considered state-of-the-art in image recognition and offer the ability to provide a prompt and definite diagnosis.. A dataset containing 163k leaf images; captured in a controlled environment, is established for training and validating the model. Validation results show that the proposed method can achieve an accuracy of 86.87%. This demonstrates the technical feasibility of CNNs in classifying plant diseases and presents a path towards AI solutions for small holder farmers

INTRODUCTION

BY 2050, GLOBAL CROP PRODUCTION MUST INCREASE BY AT LEAST 50% TO SUPPORT THE PREDICTED DEMAND [1]. THE MAJORITY OF PRODUCTION CURRENTLY OCCURS IN AFRICA AND ASIA, WHERE 83% OF FARMERS ARE FAMILY RUN WITH LITTLE TO NO HORTICULTURAL EXPERTISE [2, 3]. DUE TO THIS, YIELD LOSSES OF GREATER THAN 50%; AS A RESULT OF PESTS AND DISEASES ARE COMMON [4]. IN CLASSIFYING CROP DISEASES, THE TRADITIONAL METHOD OF HUMAN ANALYSIS BY VISUAL INSPECTION IS NO LONGER FEASIBLE. THE DEVELOPMENT OF COMPUTER VISION MODELS OFFERS A QUICK, STANDARDISED AND ACCURATE SOLUTION TO THIS ISSUE. ONCE TRAINED, A CLASSIFIER CAN ALSO BE DEPLOYED AS AN APPLICATION [5]. EASY TO USE, ALL THAT IS REQUIRED IS AN INTERNET CONNECTION AND CAMERAEQUIPPED SMARTPHONE. POPULAR COMMERCIAL APPS 'INATURALIST' [6] AND 'PLANTSNAP' [7] DEMONSTRATE HOW THIS CAN BE EXECUTED. BOTH APPS HAVE ATTAINED SUCCESS IN NOT ONLY DELIVERING EXPERTISE TO USERS BUT ALSO IN BUILDING AN INTERACTIVE ONLINE SOCIAL COMMUNITY. EACH YEAR, SMARTPHONES CONTINUE TO BECOME MORE ACCESSIBLE AND AFFORDABLE. IN 2020 THERE ARE APPROXIMATELY 5 BILLON SMARTPHONE USERS IN THE WORLD [8]. OF THIS, ONE BILLION USERS ARE LOCATED IN INDIA AND A FURTHER ONE BILLION ARE LOCATED IN AFRICA. ACCORDING TO STATISTA, THESE FIGURES HAVE CONSISTENTLY RISEN EVERY YEAR FOR THE LAST DECADE [9]. WITH THESE FACTS IN MIND, IT IS BELIEVED THAT AI APPS WILL PLAY AN IMPORTANT ROLE IN SHAPING THE FUTURE OF FARMING.

OBJECTIVE

The specific goals of this research are to:

I)Determine the model's overall effectiveness in classifying diseases using both a validation and test dataset.

ii) Compare the model's accuracy when tested with various image sizes and augmentation settings.

iii) Deploy the trained model to create an easy to use web application

This research will be carried out with the needs of smallholder farmers in mind. The classifier and web application will require both a smartphone and internet connection, which as previously mentioned, continue to reach remote regions. In acknowledging the limitations of basic camera phones, the model will be tested with a variety of image sizes and augmentation settings.

MATERIALS AND METHODS

THIS SECTION DESCRIBES THE STEPS INVOLVED IN CREATING AND DEPLOYING THE CLASSIFIER. CLASSIFICATION BY CNN IS DIVIDED INTO THREE PHASES WHICH TACKLE SEPARATE TASKS

A) Data Acquisition

All imagery derive from 'The PlantVilllage Dataset, and open-access repository which contains in total 163k images.

B) Data Pre-Processing

First, augmentation settings are applied to the training data. These are generated 'on the fly', with each operation carrying a weighted probability of appearing in each epoch.. Finally, all images are re-sized and normalized. Resizing is carried out using a compress function, to 150 x 150. As a pre-trained model is used, the RBG ImageNet statistics are used to normalize

C) Classification by CNN

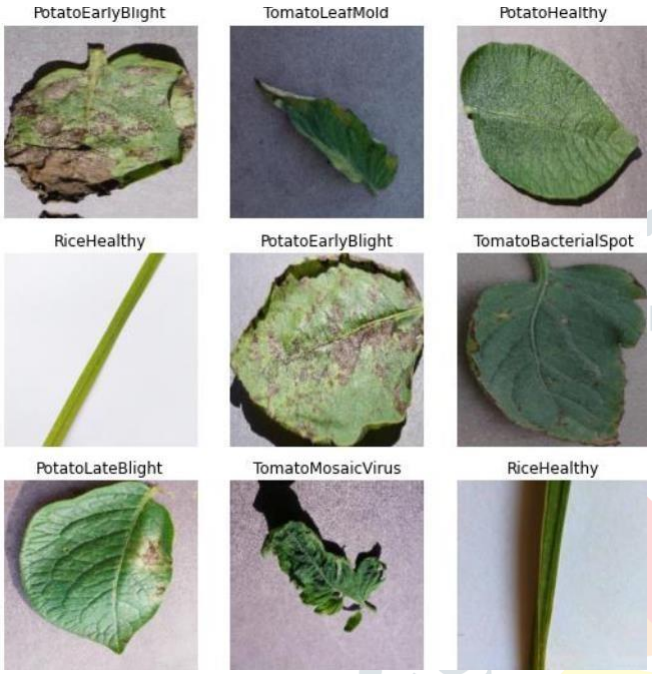
1) Phase One – Trialling of Image size

Phase one aims to investigate the effect that image size has on model performance. In total, five images sized are tested to 255 x 255.
To begin, the kaggle pre-trained weights are downloaded. As a default of transfer learning, all layers with the except of the final two layers are frozen. These contain new weights and are specific to the plant disease classification task. Freezing allows these layers to be disease separately trained, without backpropagating the gradients.

With this complete, the remaining layers are released. To aid the fine-tuning process, a plot displaying learning rate vs loss is generated and analysed. From this, a suitable learning is selected, and the model is run.

2) Phase Two – Model Optimisation

Next, the final two layers are isolated and trained at the default learning rate. With this complete, fine tuning is performed, running multiple trials to test a series of learning rates and number of epochs.



3) Phase Three – Visualisations

For the purpose of interpretation, a series of visualisations are generated based on the validation and test datasets. Additionally, the model is deployed to create a web application. To achieve this, the completed essential files are stored in a Pycharm Community Edition and the model is exported as a main.py file.

RESULTS

1) Phase One – Trialling of Image Size.

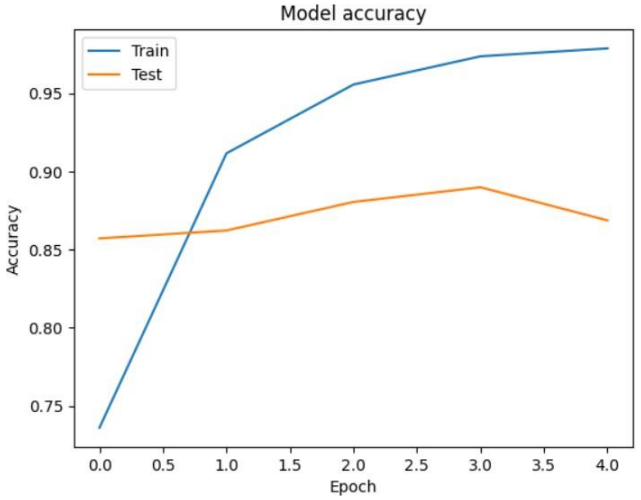
The results of Phase One prove that it is possible to achieve an accuracy score of greater than 85% for image sizes 256 x 256. As expected, an increase in image size not only improves feature extraction but also increases running time

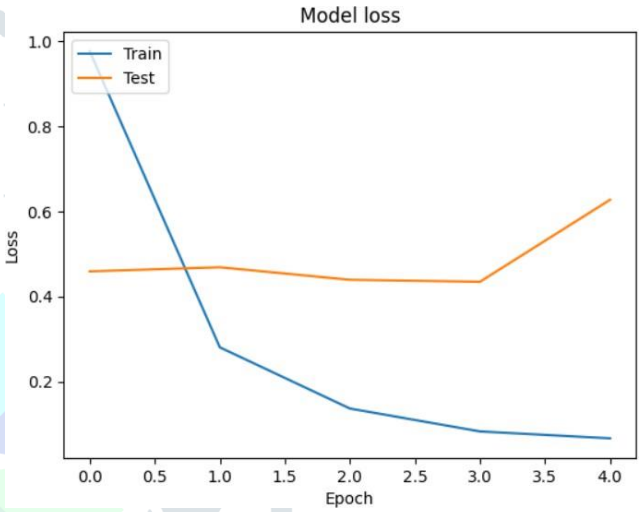
This initial analysis produced excellent results. As previously stated, the model would be accepted if it reached an accuracy of at least 80%. Even at this early stage, results far exceed the acceptance criteria.

Overall, image size 256 produced the best results including the highest accuracy

2) Phase Two – Model Optimisation

Prior to fine-tuning, the model attained an accuracy of 0.8675.To aid fine-tuning, a plot depicting learning rate (logarithmic scale) v loss was analyzed. This demonstrates a relatively low loss between learning rates





3) Phase Three– Visualisasations

An analysis of heat maps reveals the inner workings of the CNN. Colour, shape and texture appear to be important factors in working to extract plant disease features

Colour appears to be especially crucial, helping to clearly differentiate similar diseases, by adding an extra dimension of characterisation. This explains the importance of RGB data to disease classification tasks

For all three species, the CNN shows effectiveness in recognising features.

CONCLUSION

To prevent losses, small holder farmers are dependent on a timely and accurate crop disease diagnosis. In this study, a pre-trained Convolutional Neural Network was fine-tuned, and the model was deployed online. The final result was a plant disease detection app. This service is free, easy to use and requires just a smart phone and internet connection. Thus, the user's needs as defined in this paper have been fulfilled.

A thorough investigation exposes the capabilities and limitations of the model. Overall, when validated in a controlled environment, an accuracy of 86.72% is presented. This achieved accuracy depends on a number of factors including the stage of disease, disease type, background data and object composition. Due to this, a set of user guidelines would be required for commercial use, to ensure the stated accuracy is delivered. As the model was trained using a plain background and singular leaf, imitation of these features is best.

Augmentation and transfer learning in this case, proved beneficial to the model, helping the CNN to generalize more reliability. While this improved the model's ability to extract features r, it was not enough when the model was presented with 'in field' imagery. In this case, the classifier ranked an accuracy of just 44% Above all, this highlights the importance of diversifying the training dataset to include alternative background data, additional plant anatomy and varying stages of disease.

Overall, this study is conclusive in demonstrating how CNNs may be applied to empower small-holder farmers in their fight against plant disease. In the future, work should be focused on diversifying training datasets and also in testing similar web applications in real life situations. Without such developments, the struggle against plant disease will continue.

REFERENCES

[1] M. Hunter, R. Smith, M. Schipanski, L. Atwood and D. Mortensen, "Agriculture in 2050: Recalibrating Targets for Sustainable Intensification", BioScience, vol. 67, no. 4, pp. 386-391, 2017. Available: 10.1093/biosci/bix010.

[2] "PlantVillage", Plantvillage.psu.edu, 2020. [Online]. Available: https://plantvillage.psu.edu/. [Accessed: 31- Jan- 2020].

[3] D. Klauser, "Challenges in monitoring and managing plant diseases in developing countries", Journal of Plant Diseases and Protection, vol. 125, no. 3, pp. 235-237, 2018. Available: 10.1007/s41348-018-0145- 9.

[4] A.Muimba-Kankolongo, Food crop production by smallholder farmers in Southern Africa. Elsevier, 2018, pp. 23-27.

[5] J. Boulent, S. Foucher, J. Théau and P. St-Charles, "Convolutional Neural Networks for the Automatic Identification of Plant Diseases", Frontiers in Plant Science, vol. 10, 2019. Available: 10.3389/fpls.2019.00941.

[6] "iNaturalist", iNaturalist, 2020. [Online]. Available: https://www.inaturalist.org/.

[7] I. Ltd., "Success-stories - PlantSnap: Training the world's largest plant recognition classifier. | Imagga Technologies Ltd.", Imagga.com, 2020. [Online]. Available: https://imagga.com/success-stories/plantsnapcase-study.

[8] "Smartphone users worldwide 2020 | Statista", Statista, 2020. [Online]. Available: https://www.statista.com/statistics/330695/number-ofsmartphone-users-worldwide/.

[9] K. Liakos, P. Busato, D. Moshou, S. Pearson and D. Bochtis, "Machine Learning in Agriculture: A Review", Sensors, vol. 18, no. 8, p. 2674, 2018. Available: 10.3390/s18082674.

[10] Y.Toda and F. Okura, "How Convolutional Neural Networks Diagnose Plant Disease", Plant Phenomics, vol. 2019, pp. 1-14, 2019. Available: 10.34133/2019/9237136.