A Comparative Study on Plant Disease Detection Using Machine Learning Algorithm

Muhammad Solihin Kadir*[1],Syahid Anuar[2]

[1,2]*Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur*

[1]*muhammadsolihin@graduate.utm.my*,

[2]*syahid.anuar@utm.my*

Article history

Received:
15 Sept 2022

Received in revised form:
25 Oct 2022

Accepted:
19 Nov 2022

Published online:
15 Dec 2022

*Corresponding author
*muhammadsolinin@graduate.utm.my

Abstract

The crop diseases are major problem in agriculture industry that requires an accurate and fast crop disease detection method to prevent and limiting major loss. Many researchers utilize machine learning algorithm to achieve this solution. Majority of the solution either using traditional machine learning algorithm or deep learning-based algorithm. For traditional machine learning algorithm, the algorithm usually used feature extraction algorithm paired with machine learning algorithm such as Support Vector Machine, Logistic Regression and K-Nearest Neighbor. Deep learning-based algorithm utilize either fully connected neural network or use convolution neural network as feature extractor and paired it with machine learning classifier. However, evaluating those algorithms are quite difficult due to different settings in each experiment done in evaluating deep learning-based algorithm and traditional machine learning based algorithm. The purpose of this paper is to evaluate those algorithms with same dataset which is Plant Village dataset to give them fair comparison in performance. The results show that both machine learning and deep learning algorithm achieve great result with the highest accuracy achieve around 97% accuracy.

Keywords: *Crop diseases, deep learning, convolutional neural network, machine learning, Support Vector Machine, K-Nearest Neighbor*

1. Introduction

Agriculture industry is facing one major problem, which is crop diseases. Due to climate change, crop diseases become deadlier than ever which according to Food and Agriculture Organization of the United Nations (FAO), on 2018 crop diseases account for an estimated 10-16 percent loss of global agriculture harvest [1]. Crop diseases, without some sort of early detection method, will greatly affect the agriculture industry. Current detection method is simply by using naked eye, which expensive and also unscalable method to deploy in large farm. Thus, many researchers are using artificial intelligent and computer vision to detect agriculture disease in agriculture goods. By using artificial intelligence, result can be produced faster, precise, and efficient compare to the human beings that prone to the human error.

. On the current information era, devices around the world has more computing power than ever. Machine learning based solution, which is computably expensive before become more viable nowadays. Thus, many researchers use this opportunity to use machine learning algorithm to classify crop diseases. K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) is ones of the popular

machine learning algorithm used in crop diseases classification [2], [3]. The machine learning algorithm usually paired with some feature extraction algorithm. However, this algorithm come with huge drawback which is reliance on pre-processing technique that computationally demanding processing time and required high domain of knowledge [4], [5].

Other popular Artificial Intelligence (AI) algorithm is Convolution Neural Network (CNN). CNN, as name suggest is inspired from biological neural network. CNN consist of various layers, which primally convolution, pooling and fully-connected layer. The first layer in CNN will be convolution layer, which can be followed by another convolution layer or pooling layers. The last layer will be fully connected layer for classification purposes. The main purpose of CNN is to create automated and trained models which greatly reduced human intervention in finetuning an Artificial Intelligence algorithm. These algorithms can take up clusters of data and also employ more data points to increase the accuracy. However, CNN requires a lot of data and a lot more of training time to have an acceptable accuracy [6]. One of the CNN architecture, Resnet50 takes around 14 days to train on ImageNet dataset [6]. And ImageNet itself have thousands of images in their dataset [7].

On the most literature published to overcome agriculture crop diseases, most of the algorithm used are usually machine learning algorithm such as KNN and SVM coupled with feature extraction algorithm such as Gray Level Co-occurrence Matrix (GLCM) and Scale-invariant Feature Transform (SIFT) or by using deep learning algorithm such as CNN to identify crop diseases. Both of this method has its own advantages over the others. However, it is difficult to compare both of this method since different paper tend to have different settings and dataset for their experiment. Thus, they are need for study to evaluate these two different algorithms with same settings and dataset

2.Literature Survey

AI as a tool have become a significant part for solving real problem regarding problems related to prediction and regression. One of the branches of AI is machine learning. One of problem that machine learning algorithm have been applied by researchers is crop diseases classification. In image recognition field, a great data collection is a significant part of the research. One of data collection used by using handheld devices to take images and integrating image processing technique to their methodology [8]. From the Table 1, this technique still leave room for improvement. The summary of literature review for traditional machine learning algorithm in plant disease detection are summarise in Table 1.

A part of deep learning algorithm, CNN is also used for crop disease image classification by the researchers. CNN have no need for image segmentation and feature extraction because of their layers have already fulfil similar role. However, CNN usually need a huge amount of dataset for training purposes. This problem can somehow be mitigated by using transfer learning. One of the researchers that use this method is [9] in order to create a CNN model that can classify crop diseases. Some other researchers are using CNN model that's are trained on different dataset as feature extractor and use machine learning classifier such as KNN and SVM to classify [10]. The summary of CNN based algorithm are summarise in Table 2.

Table 1 Related work in Crop Disease Detection using Machine Learning Algorithm

Author	Title	Process	Dataset	Result
[3]	Detection of plant leaf diseases using image segmentation and soft computing techniques	1. Images in dataset were applied with image processing techniques. 2.The datasets were applied with Green pixel masking. 3.Segmentation were used using GA. 4.Color co-occurrence feature were computed. 5.Classify using Minimum Distance Criterion and SVM.	Various leaves images of banana, beans, lemons and rose were taken using camera.	Accuracy with SVM 95.71%
[11]	Rice Blast Disease Detection and Classification Using Machine Learning Algorithm	1. Images in dataset were converted using color conversion. 2. K-means clustering were used for segmentation. 3. Extract feature such as Mean Value, GLCM and Standard Deviation 4. Classifier used is (Artificial Neural Network) ANN	Images of paddy leaves	90% testing accuracy
[12]	Detection of potato diseases using image segmentation and multiclass support vector machine	1. Pixel masking were used for automatic region segmentation. 2. Feature extracted is colour and texture statistical analysis. 3. SVM were used as classifier.	Plant Village Dataset	95% total accuracy
[13]	Improved Segmentation Approach for Plant Disease Detection	1. Color negation and conversion used for image enhancement. 2. Colour thresholding and morphological operation for image segmentation. 3. Histogram and GLCM used for feature extraction. 4. Used fully-connected layer classifier.	Plant Village dataset	99.25% accuracy

Table 2. Related work in Crop Disease Detection by using Convolution Neural Network

Author	Title	Process	Result	Dataset
[14]	Deep Learning for Image-Based Cassava Disease Detection	1. Inception V3 CNN model is tested with varieties of classifiers.	SVM achieve highest accuracy.	Cassava Image Dataset
[15]	Basic investigation on a robust and practical plant diagnostic system	1. Image augmentation were performed on the dataset. 2. Two datasets were used to trained their custom CNN models, which one have bad images and other have good images.	The CNN achieve average accuracy of 82.30%.	Cucumber Leaf images
[16]	Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases	1. In order to detect plant diseases, convolutional autoencoders is used for unsupervised feature learning 2. The output are then feed to SVM classifier.	The SVM classifiers outperform the fully-connected networks.	Plant Village dataset
[17]	Plant disease and pest detection using deep learning-based features	1. Pretrained CNN models are tested with different classifier. Classifier used are SVM, Extreme Learning Machine (ELM) and KNN.	Resnet50 achieve highest accuracy of 97.86% but with SVM classifier.	Images of leaves and pest from Turkey.
[18]	Experimental study on crop disease detection based on deep learning	1. 60 experimental configurations were used for CNN testing. 2. AlexNet and GoogLeNet CNN model were used for testing.	GoogLeNet get highest accuracy of 99.35%	Plant Village Dataset

3. Experimental works and results

In this study, multiple general machine learning and convolutional neural network algorithm were evaluated to classify plant disease. All algorithm will be

trained and tested with same dataset which is part of public plant disease dataset named Plant Village Dataset [19]. For traditional based machine learning algorithm, the dataset will first be segmented with green pixel masking. Then, the feature will be extracted from the images. The feature extracted is Hu moments, Haralick and histogram. To classify the image several classifiers will be used which are Support Vector Machine, Logistic Regression, Linear Discriminant Analysis, K-Neighbours, Decision Tree, Random Forests and Gaussian Naïve Bayes.

Six CNN models are also evaluated in this experiment. The CNN models that are tested were VGG19, VGG16, Resnet50, Mobile net, InceptionResnetV2 and InceptionV3. All these models are fully connected layers with SoftMax at Dense layers. All models were train for 40 epochs with 16 batch size. The training will run on GPU to short the training time of the CNN models.

3.1. Dataset

Plant Village dataset is used in this experiment, which is an online image database which consist of variety images of crop diseases. In this experiment, the dataset used is tomato leaves images which the leaves infected by diseases. The 5 categories of images are Bacterial Spot, Early Blight, Late Blight, Leaf Mold and also healthy tomato leaves. 160 images were uses for training and 40 for testing for each class. The training images are split to 80:20 of train and test ratio.

Table 3. Categories in plant diseases dataset

Category	**Original Dataset**	
	Training Set	**Testing Set**
Healthy	160	40
Early Blight	160	40
Late Blight	160	40
Bacterial Spot	160	40
Leaf Mold	160	40

3.2. Results

This section provides the result of experiment done to compare the performance of CNN based algorithm and traditional machine learning algorithm. The precision, recall, accuracy and f1-score of models and classifier are tabulated in Table 4 and Table 5. Feature used for the classifier in Table 5 are Histogram, Hu moments and Haralick feature. Based on the result below, for CNN based algorithm InceptionResnetV2 gain the highest accuracy of 0.95 while for the traditional machine learning algorithm, Decision Tree classifier also got accuracy of 0.95. Mobile net accuracy is the lowest amount CNN based algorithm with 0.92 accuracy. The Gaussian Naïve-Bayes classifier achieve lowest accuracy of 0.73 accuracy.

Table 4 Results for CNN algorithm experiment

Model	Precision	Recall	Accuracy	F1- Score
VGG19	0.89	0.89	0.89	0.89
VGG16	0.85	0.85	0.85	0.85
Resnet50	0.92	0.92	0.92	0.91
Mobile net	0.84	0.83	0.83	0.83
InceptionResnetV2	0.96	0.96	0.95	0.95
InceptionV3	0.94	0.93	0.94	0.94

Table 5 Result for traditional machine learning method

Classifier	Precision	Recall	Accuracy	F1- Score
Logistic Regression	0.86	0.86	0.85	0.85
Linear Discriminant Analysis	0.85	0.85	0.85	0.85
K-Neighbours Classifier	0.82	0.83	0.83	0.83
Decision Tree Classifier	0.81	0.81	0.81	0.81
Random Forest Classifier	0.96	0.96	0.95	0.95
Gaussian Naïve-Bayes	0.74	0.73	0.73	0.70
Support Vector Machine	0.85	0.86	0.85	0.85

5. Conclusion

From the experiment result in Table 4 and Table 5, both traditional machine learning algorithms and CNN algorithms get very satisfying result. Random Forest Classifier and InceptionResnetv2 both gain 95% accuracy. Both algorithms have its own advantages. From the literature review, some research even gains higher accuracy of 99.35% by using CNN algorithm which also tested on Plant Village Dataset [18]. In the end, it all depend on research ability to fine tune the model or pick relevant features to extract. From this experiment, there were some advantages found on CNN based algorithms and traditional machine learning algorithm. CNN algorithm takes very long time to train compared to traditional machine learning algorithm even though the CNN algorithm were run on GPU to further speed up the training time. Traditional machine learning algorithm however depend heavily on green pixel masking, which is not very robust algorithm since green pixel value were put manually from the researcher. This make green pixel masking not effective on dataset which taken outside of the lab which have more vibrant colour of green.

For the future work, some further experiment can be done on CNN based algorithm. Right now, the models are fully connected model. The models can use which machine learning classifier for more variant result. For the traditional machine learning algorithm, the feature can be tested singularly instead of concatenated

together to feed the classifier. There also other feature extraction method that are not tested yet such as GLCM and SIFTS.

Acknowledgments

While doing this paper, I meet many people from Universiti Teknologi Malaysia (UTM) in search of guidance to finishing this paper. I wish to express my great gratitude to Dr.Syahid Anuar for his guidance and kind words to push me to finish this paper. I also in debt to my family members in giving their support and motivation to me. Lastly, I would like to show my appreciation to UTM staff who provide me with various help to complete this paper.

References

[1] "FAO - News Article: Global body adopts new measures to stop the spread of plant pests." https://www.fao.org/news/story/en/item/1118322/icode/ (accessed Nov. 15, 2022).

[2] S. B. Jadhav, V. R. Udupi, and S. B. Patil, "Soybean leaf disease detection and severity measurement using multiclass SVM and KNN classifier," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, pp. 4077–4091, Oct. 2019.

[3] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, Mar. 2017.

[4] J. Francis, Anto Sahaya Dhas D, and Anoop B K, "Identification of leaf diseases in pepper plants using soft computing techniques," Oct. 2016, pp. 168–173.

[5] S. Zhang and Z. Wang, "Cucumber disease recognition based on Global-Local Singular value decomposition," *Neurocomputing*, vol. 205, pp. 341–348, Sep. 2016.

[6] Y. You, Z. Zhang, C.-J. Hsieh, J. Demmel, and K. Keutzer, "ImageNet Training in Minutes," Sep. 2017.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," Mar. 2010, pp. 248–255.

[8] C. G. Dhaware and K. H. Wanjale, "A modern approach for plant leaf disease classification which depends on leaf image processing," in *2017 International Conference on Computer Communication and Informatics, ICCCI 2017*, Nov. 2017.

[9] V. K. Shrivastava, M. K. Pradhan, and M. P. Thakur, "Neural Networks for Rice Plant Disease Classification," pp. 1023–1030, 2021.

[10] J. Lu, R. Ehsani, Y. Shi, J. Abdulridha, A. I. de Castro, and Y. Xu, "Field detection of anthracnose crown rot in strawberry using spectroscopy technology," *Comput Electron Agric*, vol. 135, pp. 289–299, Apr. 2017.

[11] S. Ramesh and D. Vydeki, "Rice blast disease detection and classification using machine learning algorithm," in *Proceedings - 2nd International Conference on Micro-Electronics and Telecommunication Engineering, ICMETE 2018*, Sep. 2018, pp. 255–259.

[12] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," in *Canadian Conference on Electrical and Computer Engineering*, Jun. 2017.

[13] M. A. Rahman, M. M. Islam, G. M. S. Mahdee, and M. W. Ul Kabir, "Improved Segmentation Approach for Plant Disease Detection," in *1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019*, May 2019.

[14] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Front Plant Sci*, vol. 8, Oct. 2017.

[15] E. Fujita, Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic investigation on a robust and practical plant diagnostic system," in *Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016*, Jan. 2017, pp. 989–992.

[16] H. F. Pardede, E. Suryawati, R. Sustika, and V. Zilvan, "Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases," *2018 International Conference on Computer, Control, Informatics and its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018 - Proceeding*, pp. 158–162, Jan. 2019

[17] M. Türkoğlu and D. Hanbay, "Plant disease and pest detection using deep learning-based features," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 3, pp. 1636–1651, 2019,

[18] X. Xia, Y. Wu, Q. Lu, and C. Fan, "Experimental study on crop disease detection based on deep learning," in *IOP Conference Series: Materials Science and Engineering*, Aug. 2019, vol. 569, no. 5.

[19] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," 2015.