Diagnosis of Multiple Cucumber Infections with Convolutional Neural Networks

Hiroki TANI Hosei University Tokyo, Japan hiroki.tani.2j@stu.hosei.ac.jp

Hiroyuki UGA Saitama Agricultural Technology Research Center Saitama, Japan uga.hiroyuki@pref.saitama.lg.jp Ryunosuke KOTANI Hosei University Tokyo, Japan ryunosuke.kotani.58@hosei.ac.jp Satoshi KAGIWADA Hosei University Tokyo, Japan kagiwada@hosei.ac.jp

Hitoshi IYATOMI Hosei University Tokyo, Japan iyatomi@hosei.ac.jp

Abstract—Recent machine learning approaches have shown promising results in the field of automated plant diagnosis. However, all of the systems were designed to diagnose single infections, thus they do not assume multiple infections. In this paper, we created our original on-site cucumber leaf dataset including multiple infections to build a practical plant diagnosis system. Our dataset has a total of 48,311 cucumber leaf images (38,821 leaves infected with any of 11 kinds of diseases, 1,814 leaves infected with multiple diseases, and 7,676 healthy leaves). We developed a convolutional neural networks (CNN) classifier having the sigmoid function with a tunable threshold on each node in the last output layer. Our model attained on average a 95.5% classification accuracy on the entire dataset. On only multiple infected cases, the result was 85.9% and it accurately identified at least one disease in 1,808 out of the total of 1,814 (99.7%).

Index Terms—Convolutional Neural Networks, Plant Diagnosis, Image Processing, Multiple Infections

I. INTRODUCTION

Plant diseases affect agricultural production on a global scale [1], [2]. Identification and appropriate treatment of the diseases in their early stage are essential to minimize the damage. However, plant diagnosis is generally conducted through visual examination by experts, with subsequent genetic testing applied as necessary; thus, it is expensive and time-consuming. An accurate, fast and low-cost automated plant disease diagnosis system is required for solving such problems.

Recent machine learning approaches to automated plant diagnosis have shown promising results [3]–[5]. In particular, the effects of deep learning using models like convolutional neural networks (CNN) have become understood. They have attracted many researchers and, as a result, many achievements have been seen [6]–[12].

Kawasaki et al. [6] proposed an automated cucumber diagnosis system with CNN. They demonstrated that the new methodology did not need to be tackled with essential but difficult traditional issues such as the extraction of the areas

978-1-5386-9306-3/18/\$31.00 ©2018 IEEE

involved (e.g. leaf) in the pre-processing and embedding the knowledge of experts in the classification. Mohanty et al. [7] used 54,306 plant leaf images consisting of a total of 38 classes of crop-disease pairs from Plant Village [13] and built CNN models to identify crop and plant diseases. They achieved an excellent classification accuracy of 99.35%. On the other hand, all the leaves used in their study were physically cropped and each leaf was separately placed in front of a uniform coloured background in advance and photographed. These conditions are very different to what we observed in the field. In fact, they noted in their manuscript that the accuracy dropped to around 31% under the different setting from the training images. In addition, we found a significant number of inappropriate labels assigned on the Plant Village dataset. This is one of the hazards of the data collection type open dataset. Liu et al. [12] proposed a novel apple diagnosis system for classifying four kinds of diseases based on CNNs. They claimed that their own CNN model showed the best classification performance of 97.62% compared with that using well-known CNN models such as VGG [14] or ResNet [15]. However, their study also used cropped leaf images as well as the PlantVillage dataset and thus these systems cannot be directly applied to practical situations with complex backgrounds, for example.

Along with the development of deep learning techniques for general object recognition, more practical systems have been proposed also in this field. Lu et al. [10] proposed a wheat diagnosing system which performs localization of the areas involved and diagnosis simultaneously based on a full convolutional VGG-based [14] network with their own wheat disease database (WDD2017 Dataset). Their system achieved a mean recognition accuracy of 97.95%. Fuentes et al. [11] proposed a detection and identification system for diseases and pests in tomato with three types of object detection method (Faster Region-based CNN [17], Single Shot Multibox Detector [18] based on VGG [14] and ResNet [15]). Their dataset has a total of 5,000 field tomato leaf images. We think these systems are practical because their images based on the system were photographed in the field. However, to the best of our knowledge, all of the systems were designed to diagnose single infections and do not assume multiple infections. In practice, multiple infections are often seen, but systematic studies on this have not been proposed so far. The symptoms of the multiple infections are not an independent sum of each and have an interaction, especially in the case of viral diseases. For this reason, the conventional systems cannot handle these complex cases and would produce undesired output. One difficulty behind this problem is that creating a reliable dataset with multiple infections is demanding and costly. There is no such available dataset in public. Against this background, we built our own on-site cucumber leaf dataset including multiple infections and the classifier to address this problem. To the best of our knowledge, our research is the first systematic study on automated diagnosis on multiple infections.

II. METHOD

A. Dataset and preprocess

We created our own on-site cucumber leaf dataset including healthy and diseased leaves with single and multiple infections to develop a robust diagnosis system. We handled seven viral (CCYV, CMV, KGMMV, MYSV, PRSV, ZYMV and WMV) and four fungal (Brown spot, Downy mildew, Grey mould, and Powdery mildew) diseases. Multiple infected leaves have two or three diseases from the abovementioned single diseases (11 disease types). In this study, we made and used 13 combinations of multiple infections based on the consideration of the actual situation. Table I shows the name and the number of images in each class in the first and second column. Our dataset is composed of a total of 48,331 cucumber leaf images comprising 38,821 single, 1,814 multiple infections, and 7,676 healthy leaves. These cucumbers were planted under a strictly controlled environment at Saitama Agricultural Research Center. The target disease was transmitted in an isolated environment to avoid contamination.

Each leaf image was photographed under only one loose requirement; a leaf must be near the middle of the picture. As the table shows, the variation in the number of datasets in each class is broad; single and multiple infections are from 250 to 10,210 and 1 to 923, respectively.

Our dataset consists of different size and aspect ratio images. Therefore, we crop each image to a square at its center and resize it to 316×316 pixels beforehand in order to follow the data augmentation process proposed in the literature [8]. We rotate the images clockwise with the incremental step size of 20 degrees and then generate their mirror images with respect to the horizontal axis for data augmentation. In summary, these processes artificially increase the training data by 36 times. Finally, we resize all images to 224×224 pixels.

B. Architecture of the classifier

Our CNN-based plant diagnosis system is composed of eight convolutional and two fully connected layers. This architecture is designed with reference to the VGG-net [14] model, and it has been miniaturized in consideration of the data size. Fig. 1 shows the network architecture. Each fully



Fig. 1. Our model architecture.

connected layer performs a batch normalization [16] and all the convolutional layers have smaller feature maps than the original VGG-net [14] to reduce the degree of freedom. Two fully connected layers have 2,048 nodes each, with a dropout ratio of 0.5. We optimized the architecture for our task according to several preliminary experiments. In our preliminary experiments, our model was 9x faster than the original VGG architecture with equivalent accuracy.

To classify multiple infections, we omit the soft-max function at the last output layer of the network and instead utilize a sigmoid function with tunable thresholds on each node. Since the number of data in each class is unbalanced, we set the thresholds for each output value beforehand by applying gridsearch on the F1-score result domain.

III. EXPERIMENT

In our experiments, all performances were evaluated with the 4-fold cross-validation. In this study, each of the multiple infections (e.g. WMV + Downy Mildew, WMV + Brown spot) is treated as an independent class. Namely, we handle our classification problem as a 25-class classification task (11 single, 13 multiple diseases and healthy), because the appearance of a leaf infected with multiple diseases is not the sum of each disease, as mentioned earlier.

In order to evaluate the classification performance for multiple infections, we introduce two performance evaluation criteria for them; the strict evaluation criteria and the weighted evaluation criteria [19]. The former evaluates the result as correct only when all the diseases of the target are correctly detected and allows no false detection. The latter does it in a weighted manner. For example, if the system detects one out of three infections, this criterion is counted as 1/3 of a true positive. This affects both the precision and recall rates. False detection also reduces the precision. In each criterion, we calculate the precision, recall, and F1-score. Note here that the recall for the sum of several classes (e.g. the total of single infections, total of plural infections, overall) corresponds to the average accuracy of the total data.

IV. RESULTS

Fig. 2 shows examples of single infection images ((a): MYSV and (b): ZYMV) and a multiple infection image ((c): MYSV & ZYMV). Our system correctly identified the left image as MYSV with a score of 0.859 (decision threshold



Fig. 2. Example of single infection images and multiple infection image.

 θ_{MYSV} = 0.811), the middle image as ZYMV with 0.729 (θ_{ZYMV} = 0.726) and the right image as multiple infections of these two viruses with 1.000 (>0.811) and 0.792 (>0.726), respectively.

Table I summarizes the performance of our cucumber diagnosis model. With the strict evaluation criteria, our model achieved 95.5% accuracy on average. For the disease cases, the model showed an average of 94.9% precision and 95.0% for recall (=sensitivity). Among them, for single infection cases only, the values were 95.5% and 96.3%, and for only multiple infection cases, they were 85.9% and 85.1%, respectively. For healthy cases, our model attained the precision of 98.5% and the recall (=specificity) of 97.9%. With the weighted evaluation criteria, these scores were improved, especially in the multiple infection cases. They reached 93.3% and 93.2% in precision and recall, respectively.

In the ten categories from the fewer data in the multiple infection classes (i.e. where the number of training data is equal to or less than 100), the performance was limited under the strict evaluation criteria. Nonetheless, our model showed a reliable performance for all the categories with the weighted evaluation criteria, and for multiple infection leaves our model detected at least one disease correctly in 1,808 cases out of the total 1,814 cases.

V. DISCUSSION

From Table I, we confirmed that our model showed an excellent diagnosis performance not only for single disease cases, but also for multiple infection cases (91.1% and 91.7% in the F1 score with the strict and weighted evaluation criteria, respectively).

Our model has an output threshold for each class, enabling the output of multiple infections. In our experiments, the output of our model was lower than any of the thresholds of 12 classes on 264 cases (0.68%). Although these data were incorrectly classified, we consider that the ratio is in the acceptable range, and the ability of the model to yield "others" for unknown objects is an important characteristic considering the diversity of diseases.

We would like to note here that, when we excluded the multiple infection images from our dataset, the diagnosis performance for those cases dramatically declined and eventually failed.

Finally, in order to diagnose multiple infections appropriately, the preparation of a reliable dataset is crucial, regardless of its creation cost. However, the creation of a dataset for each combination is not practical at all. We will seek effective learning methods in the future.

VI. CONCLUSION

In this paper, we created our original on-site cucumber leaf dataset including multiple infections to build a practical plant diagnosis system. We created our own reliable dataset and developed a CNN-based classifier having a sigmoid function with tunable thresholds at the last output layer of the network instead of the soft-max function. Our model achieved an excellent diagnosis performance on both single and multiple infections, with 95.5% accuracy on average and 85.5% for the strict F1-score on the multiple infections only dataset.

ACKNOWLEDGMENT

This research was supported in part by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Fundamental Research (C), 17K08033, 2017–2020.

REFERENCES

- D. Majumdat, D. K. Kole, A. Chakraborty and D. D. Majumder, "Review: Detection and diagnosis of plant leaf disease using integrated image processing approach," International Journal of Computer Engineering and Applications, vol. 6, issue 3, pp. 1-13, 2014.
- [2] "2014 Georgia plant disease loss estimates," http://extension.uga.edu/publications/detail.cfm?number=AP102-7, accessed on May 5, 2018.
- [3] Q. Yao, Z. Guan, Y. Zhou, J. Tang, Y. Hu and B. Yang, "Application of support vector machine for detecting rice diseases using shape and color texture features," Proc. Engineering and Computation, Hong Kong, 2009, pp. 79-83.
- [4] F. Qin, D. Liu, B. Sun, L. Ruan, Z. Ma and H. Wang, "Identification of alfalfa leaf diseases using image recognition technology," PLoS ONE, vol. 11, issue 12, e0168274. doi:10.1371/journal, 2016.
- [5] L. Hallau, M. Neumann, B. Klatt, B. Kleinhenz, T. Klein, C. Kuhn, M. Roehrige, C. Bauckhage, K. Kersting, A.-K. Mahlein, U. Steiner and E.-C. Oerke, "Automated identification of sugar beet diseases using smartphones," Plant Pathology, vol. 67, issue 2, pp. 399-410, 2017.
- [6] Y. Kawasaki, H. Uga, S. Kagiwada and H. Iyatomi, "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks," Lecture Notes in Computer Science, vol. 9475, pp. 638-645, 2015.
- [7] S. P. Mohanty, D. P. Hughes and M. Salathe, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, pp. 1419, 2016.
- [8] E. Fujita, H. Uga, S. Kagiwada and H. Iyatomi, "Basic investigation on a robust and practical plant diagnostic system," Proc. Int. Conf. on Machine Learning and Application, 2016, pp. 989-992.
- [9] Q. H. Cap, E. Fujita, S. Kagiwada, H. Uga, and H. Iyatomi, "A deep learning approach for on-site plant leaf detection," Proc. Signal Processing and its Applications, 2018, pp. 120-124.
- [10] J. Lu, J. Hu, G. Zhao, F. Mei and C. Zhang, "A field automatic wheat disease diagnosis system," Computers and Electronics in Agriculture, pp. 369-379, 2017.
- [11] A. Fuentes, S. Yoon, S. C. Kim and D. S. Park, "A robust deep-learningbased detector for real-time tomato plant diseases and pests recognition," Sensors, vol. 17, no. 9, 2017.
- [12] B. Liu, Y. Zhang, D. J. He and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," Symmetry, vol. 10, issue 1, pp. 11, 2018.
- [13] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," CoRR, abs/1511.08060, 2015.
- [14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, abs/1409.1556, 2014.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

	# Images	F1 score [%]		Precision [%]		Recall [%]	
	0	Strict	Weighted	Strict	Weighted	Strict	Weighted
Healthy	7,676	98.2	96.5	98.5	98.5	97.9	97.9
ССҮҮ	3,034	98.1	96.5	98.4	98.4	97.8	98.1
CMV	4,565	97.1	94.4	97.3	97.3	96.8	96.9
KGMMV	3,181	91.6	84.8	92.8	92.8	90.5	90.9
MYSV	8,287	97.4	95.3	97.6	98.1	97.1	97.4
PRSV	2,318	91.7	85.2	92.1	92.1	91.3	92.1
WMV	2,790	94.4	90.0	95.7	95.9	93.2	93.8
ZYMV	10,210	97.0	94.4	97.4	97.5	96.5	96.8
Brown Spot	2,238	95.5	94.5	94.1	97.8	96.9	98.3
Downy Mildew	1,457	95.1	93.3	96.3	98.0	94.0	96.4
Grey Mould	491	90.6	87.5	91.1	94.6	90.0	94.3
Powdery Mildew	250	82.3	77.5	87.1	94.6	78.0	84.0
MYSV & ZYMV	923	95.0	96.7	95.9	97.9	94.1	97.0
WMV & Downy Mildew	365	95.9	97.2	93.7	97.9	98.1	98.8
WMV & Brown Spot	228	92.6	95.7	91.5	98.9	93.9	97.4
Brown Spot & Downy Mildew	100	53.2	79.7	56.8	83.5	50.0	76.0
Brown Spot & Powdery Mildew	51	46.2	76.2	52.5	76.2	41.2	70.6
WMV & Brown Spot & Downy Mildew	37	48.1	85.6	76.5	92.2	35.1	77.5
Brown Spot & Downy Mildew & Powdery Mildew	30	35.0	74.2	70.0	90.0	23.3	62.2
Brown Spot & Grey Mould	27	44.4	72.2	44.4	70.4	44.4	70.4
Downy Mildew & Grey Mould	25	54.1	76.2	56.5	76.1	52.0	76.0
Powdery Mildew & Grey Mould	13	-	66.6	0.0	50.0	0.0	50.0
CMV & Brown Spot	9	-	66.7	0.0	50.0	0.0	50.0
Downy Mildew & Powdery Mildew	5	36.4	78.1	33.3	91.7	40.0	80.0
Brown Spot & Downy Mildew & Grey Mould	1	-	50.0	0.0	33.3	0.0	33.3
Total Single Infection	38,821	95.9	96.0	96.3	96.8	95.5	96.0
Total Multiple Infection	1,814	85.5	91.7	85.1	93.3	85.9	93.2
Disease Total	40,635	95.0	95.4	94.9	96.7	95.0	95.9
Total	48,311	95.5 [†]	96.1	95.5 [†]	97.0	95.5 [†]	96.2

 TABLE I

 Diagnosis performance for healthy, single infected, and multiple Infected cases

[16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. 32nd Int. Conf. on Machine Learning (ICML), 2015, pp. 448-456. 2016, pp. 21-37.

- [19] G. Tsoumakas, I. Katakis and I. Vlahavas, "Mining multi-label data," in Data Mining and Knowledge Discovery Handbook, 2nd ed, O. Maimon and L. Rokach, Eds. Springer, 2010, ch. 34, pp. 667-685.
- [17] S. Ren, K. He, G. Kaiming, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems, pp. 91-99, 2015.
- [18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A. C. Berg, "SSD: Single shot multibox detector," in Proc. European Conf. on Computer Vision—ECCV, Amsterdam, The Netherlands, Oct. 8-16,

[†]This value corresponds to accuracy.