Variety Classification by Image Recognition of Grape Leaf and Berry

Ryo Sugiura, Mitsunori Ishihara, Jihyun Lim,
Chisa Iwasaki, Yu Oishi
Research Center for Agricultural Information Technology
National Agriculture and Food Research Organization (NARO)
Tsukuba, Ibaraki, Japan
e-mail: rsugiura@naro.affrc.go.jp, ishihara.mitsunori@naro.affrc.go.jp, j_lim@naro.affrc.go.jp, i.chisa@naro.affrc.go.jp, oishi.yu@naro.affrc.go.jp

Hiroshi Yakushiji, Yoshihiko Kuwana,
Toshiya Yamamoto
Department of Intellectual Property
National Agriculture and Food Research Organization (NARO)
Tsukuba, Ibaraki, Japan
e-mail: aoe@naro.affrc.go.jp, kuwana@naro.affrc.go.jp, toshiya@naro.affrc.go.jp

***Abstract*— Protection of the rights of plant breeders is essential to preserve the value of new varieties. However, table grapes, for instance, are easy to propagate vegetatively by grafting or cutting, and once they become popular and the cultivation area increases, they face a high risk of unauthorized cultivation and propagation, leading to overseas outflow. To prevent such infringements, effective methods for promptly identifying protected varieties have been desired, and deep learning-based image recognition can be one of the key techniques. The objective of this study is to verify the possibility of identifying grape varieties using images of leaves and berries. The images of leaves, young berries, and mature berries of Shine Muscat and two similar varieties were captured using smartphone cameras, and an image dataset was created to train and test classification models. Image classification models named VGG16, ResNet50, and Vision Transformer (ViT) were applied and redesigned to classify three categories. After training, these models were tested on 51 images of leaves, 174 images of young berries, and 171 images of mature berries. The models achieved classification accuracies of more than 96.1% for leaves, over 99.4% for young berries, and 100% for mature berries. Although additional testing at different sites or in different years will be needed, these results indicate that image recognition techniques can help identify plant varieties toward infringement detection.**

Keywords-image recognition; variety classification; grape leaf; grape berry.

I. Introduction

Measures to prevent superior crop varieties from being grown and propagated without permission are necessary, and breeders must receive profits from licensing the varieties. Furthermore, the protection of the rights of plant breeders is essential for preserving the value of new varieties. However, table grapes, for instance, can easily propagate vegetatively by grafting or cutting; once they become popular and their cultivation area increases, they face a high risk of unauthorized cultivation and propagation, leading to overseas outflow.

Several cases of infringements have been observed worldwide. A private company with a breeder's right of a blueberry variety filed a lawsuit in the Federal Court of Australia, alleging that some farmers were growing and selling the variety without any license [1]. Another US company that has several patented grape and cherry varieties launched administrative complaints against the illegal sale of propagation and harvested material of their varieties [2]. An Italian Court admitted the infringement of a breeder's right and stopped an alleged grape producer from harvesting and selling the infringed variety [3]. In Japan, young, high-quality grape vines have been cultivated outside the country [4]. These infringements were often confirmed using DeoxyriboNucleic Acid (DNA) profiling or comparison tests. Although DNA analysis and comparison tests are currently among the most reliable ways to identify varieties [5][6][7], they usually take time to obtain results and are sometimes tedious. It might be difficult to develop a method that completely replaces conventional methods at present. However, a simple and prompt complementary method is desired, and Deep Learning (DL)-based image recognition [8][9][10][11][12] can be one of the key techniques. This study proposes an image-based method to identify crop varieties using Deep Neural Networks (DNNs). DNNs are often used for numerous tasks, including image classification, and DL is the easiest and best way to generate classification models.

The objective of this study is to verify the possibility of identifying crop varieties using images. Grapes of the three varieties were selected, and images of their leaves and berries were captured to create the training dataset. Three classification models were trained using this dataset, and their classification accuracies were compared.

The plan of the paper is as follows. In Section II, we mainly describe the details of the image dataset and the DNN-based classification models. Section III shows the classification accuracies obtained by the trained models and discusses the results of visualizing the regions focused by the model for classification. Section IV addresses the concluding remarks.

II. Materials and Methods

A. Image Dataset

The grape varieties used in this study were Shine Muscat [13] and its similar varieties, Muscat of Alexandria and Rosario Bianco. Images of the leaves and berries of these

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

 ISBN: 978-1-68558-324-8

varieties were captured using three smartphone cameras (SONY Xperia 10IV, Google Pixel 7, Apple iPhone 13 Mini). A single leaf was laid face up on a black plate immediately after sampling off the trees, and an image was captured using one of the cameras. The cameras were held by hands facing down at a distance of approximately 35 cm from the leaves. All leaves were sampled from trees in the experimental field of National Agriculture and Food Research Organization (NARO) in Akitsu (34.33°N, 132.82°E), Hiroshima, Japan, from July 3rd to July 20th in 2023. Similarly, images of young berries sampled from June 20th to June 28th and mature berries sampled from September 21st to September 28th were captured at a distance of approximately 20 cm from the berries.

The image sizes in pixels were 4000 × 3000 (Xperia), 4080 × 3072 (Pixel 7), and 4032 × 3023 (iPhone). The spatial resolutions of leaf and berry images were approximately 0.11 and 0.07 mm, respectively. Image data were created by cropping a square region of 2500 × 2500 pixels with a leaf at the center and 800 × 800 pixels for berries. Consequently, 186 image data of leaves, 665 image data of young berries, and 673 image data of mature berries were obtained. Fig. 1 shows examples of image data for the three varieties.

B. Training and Evaluation of Classification Models

The image classification models VGG16 [14], ResNet50 [15], and Vision Transformer (ViT) [16] were applied and redesigned to classify the three categories. VGG16 is one of the simplest and most classical Convolutional Neural Networks (CNNs). It stacks 13 convolutional layers and three fully connected layers. ResNet50 is the most commonly used classifier. It has residual connections across layers in the network, which improve network performance in training process. This model is characterized by fast convergence of the trainable weight despite the large number of layers. ViT is an image classifier based on Transformer architecture, which has been the de facto standard for natural language processing. It divides an input image into a sequence of patches and examines their contextual relationships with each other using Self-Attention, which contributes to model robustness [17].

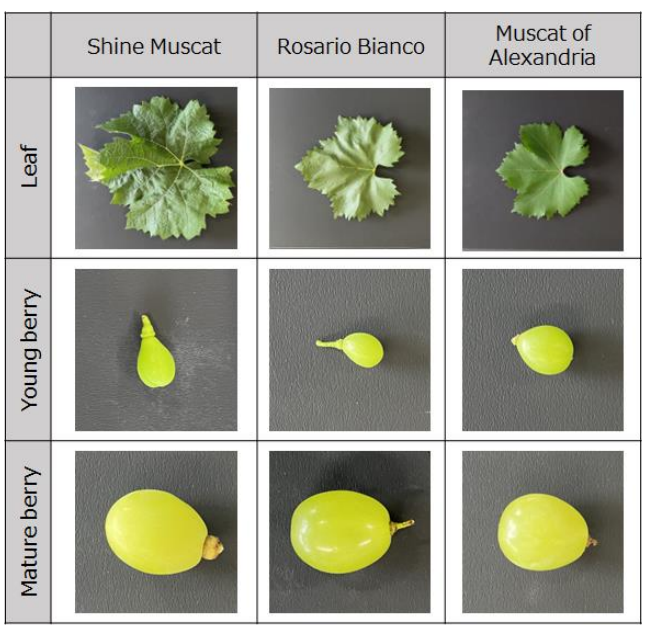


Figure 1. Image data examples of three varieties.

The Python programs of these three models and the corresponding pre-trained weights were obtained from TorchVision [18], which is part of the DL framework of the PyTorch project [19]. The output layer of each model was modified to match the classification of the three varieties.

To train and test the models, the dataset mentioned in the previous section was randomly divided into those for training, validation and test as summarized in Table I. The models were trained by fine-tuning the pre-trained weights as the initial values. The weights of each model were updated repeatedly using the training data such that the loss (cross entropy) decreased. The loss on the validation data was also calculated for each update. After 100 iterations (epochs), the weights with the lowest validation loss were selected as the best. During training, accuracy was calculated as the percentage ratio of the number of correct predictions to the total number of samples in each epoch.

After the training process, the best models were tested on 51 images of leaves, 174 images of young berries, and 171 images of mature berries to obtain the final accuracies.

C. Visualization of Interest Area

For the VGG models, Gradient-Weighted Class Activation Mapping (Grad-CAM) was applied to show the important portions of an input image for making predictions. Grad-CAM uses the gradient information flowing into the last convolutional layer of a CNN to assign importance values to each neuron for a particular decision of interest [20]. The magnitude of influence on the prediction results can be obtained as a heat map overlaying the input image. This helps us understand where the models focus on the leaf or berry to predict variety. The Grad-CAM program was obtained from the website of [21].

III. Results and Discussion

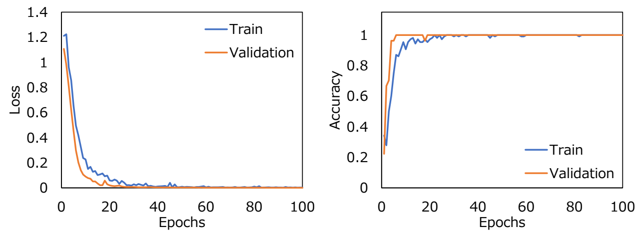
A. Loss and Accuracy Variations in Training Proccess

As an example, Fig. 2 shows the variations in loss and accuracy with an increase in the number of epochs for the training process of VGG16 for leaves. The curves of the training and validation losses gradually converged, and those of the training and validation accuracies quickly reached 1.0

TABLE I. Number of Image Data

	Train			Validation			Test		
Leaf	108			27			51		
	37	36	35	9	9	9	17	17	17
Young berry	401			90			174		
	135	141	125	30	30	30	58	58	58
Mature berry	409			93			171		
	147	129	133	31	31	31	57	57	57

The three numbers in the bottom row of each cell are of image data of Shine Muscat, Rosario Biance, and Muscat of Alexandria.



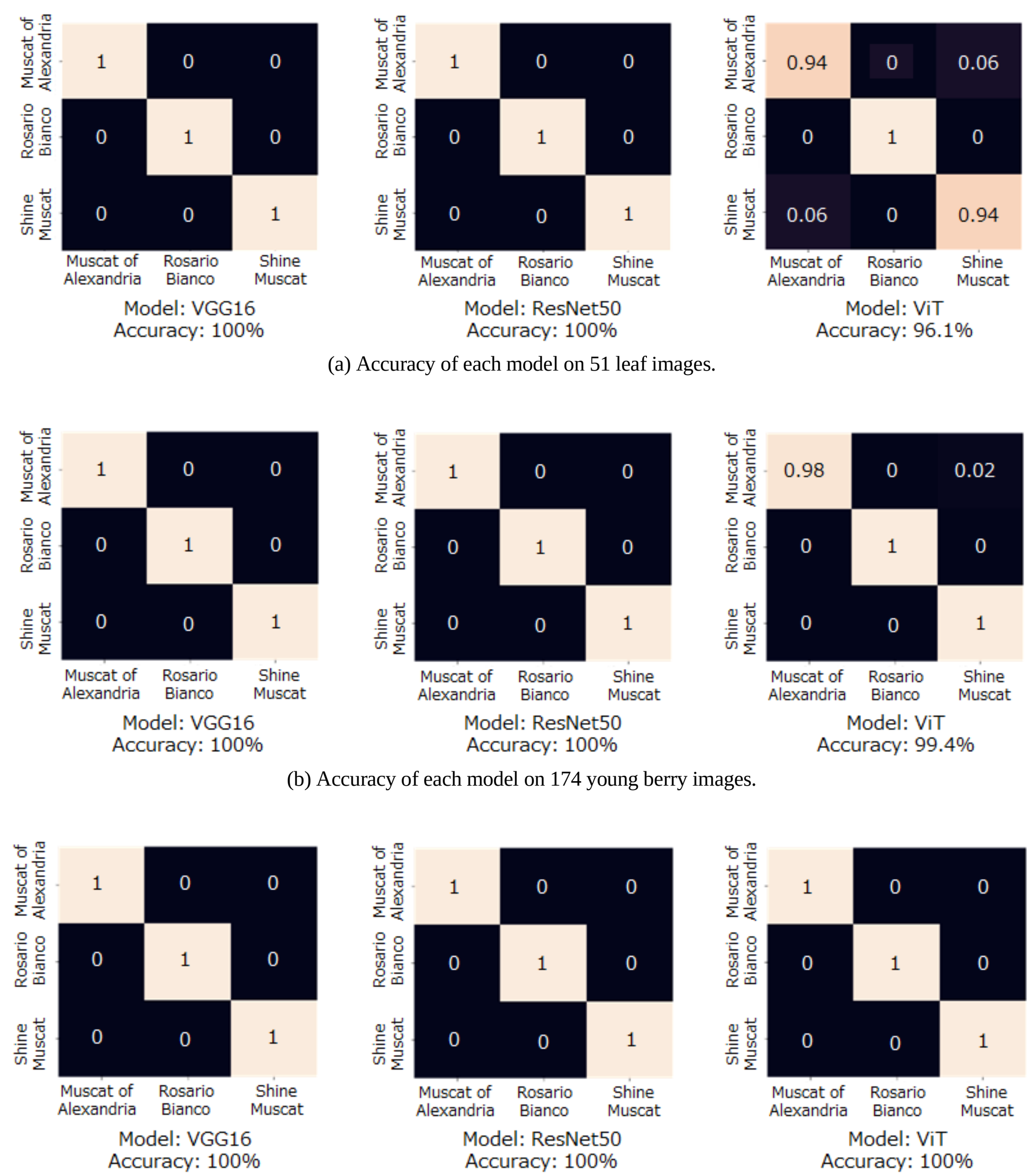
(a) Training loss (b) Training accuracy

Figure 2. Variation in loss and accuracy in the training process of VGG16 with leaf images.

within the early epochs. This implies that the training process was successfully performed, and a model was created to classify an image into three varieties. As the variations in loss and accuracy for both young and mature berries on the other models were similar, their figures were omitted from this report.

B. *Accuracies on Test Data*

The test accuracies of the trained models for leaf classification were obtained and are summarized as confusion matrices, as shown in Fig. 3(a). VGG16 and ResNet50 correctly classified all varieties on the 51 test images with 100% accuracy. However, ViT had some misclassifications and its accuracy across the three varieties was 96.1%. One of the 17 images of Muscat of Alexandria was misclassified as Shine Muscat, resulting in a recall of



(a) Accuracy of each model on 51 leaf images.

(b) Accuracy of each model on 174 young berry images.

(c) Accuracy of each model on 171 mature berry images.

Figure 3. Confusion matrices of test accuracies of variety classification. The vertical and horizontal axes are true and predicted labels, respectively.

 ISBN: 978-1-68558-324-8

94.1%, and one of the 17 image of Shine Muscat was misclassified as Muscat of Alexandria, whereas the images of Rosario Bianco were classified correctly.

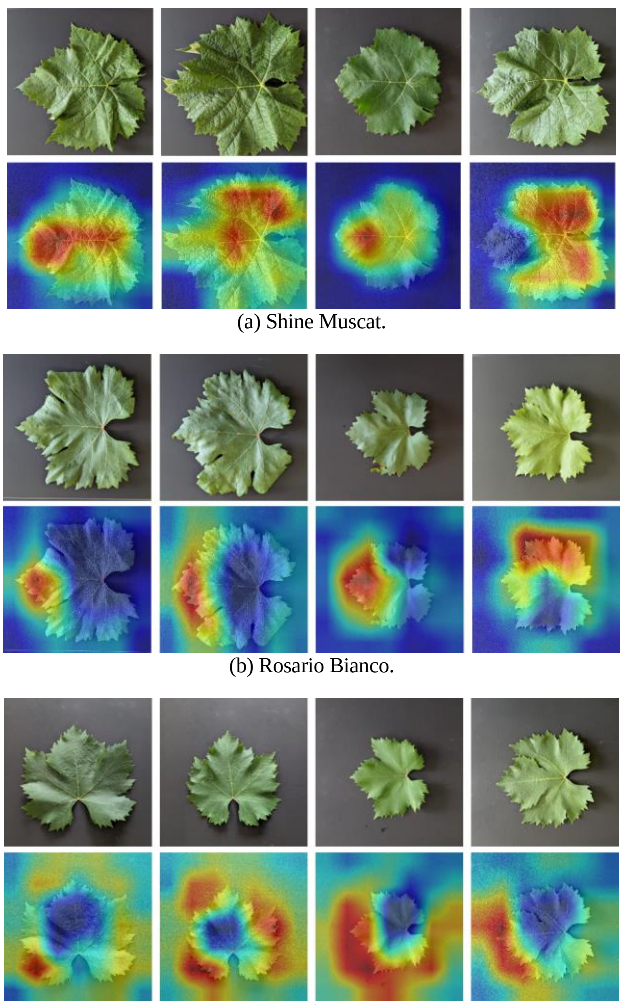
The confusion matrices for the young berries are shown in Fig. 3(b). The accuracies of VGG16 and ResNet50 were 100% similar to the leaf classifications, whereas that of ViT was 99.4% on the 174 images. One of the 58 images of Muscat of Alexandria was misclassified as Shine Muscat and its recall was 98.2%. In the case of mature berries, the accuracies were 100% across all varieties for the 171 test images, as shown in Fig. 3(c). The accuracy of ViT for young berries was 99.4%, whereas that for mature berries was 100%. This implies that the growth of berries might make the features of the variety on its shape and texture more distinct. The accuracies of ViT for leaves and young berries were not 100%. A large amount of data is required for the ViT to achieve optimal performance. However, ViT still achieved good results across leaves and berries.

VGG16 and ResNet50 provided perfect classification of the test data. The method of capturing images contributes to obtaining such good results, even with these basic models. Using a black background and capturing images immediately above the subjects made it easy to capture morphological characteristics and might lead to perfect classifications. Assuming that the images of leaves and berries on trees had been captured at several angles or distances in the outside environment and the background was not uniform, as in [22][23], it would have been difficult to obtain such good accuracies.

Although additional testing at different sites or in different years may be required, these results indicate that image recognition techniques can help identify plant varieties for infringement detection.

C. Visualization by Grad-CAM

Four heat maps generated from leaf images using the Grad-CAM program were used as examples, as shown in Fig. 4. VGG16 is assumed to have classified the images into Shine Muscat by focusing on the leaf surface (Fig. 4(a)), whereas it seems to observe some leaf tips to classify the

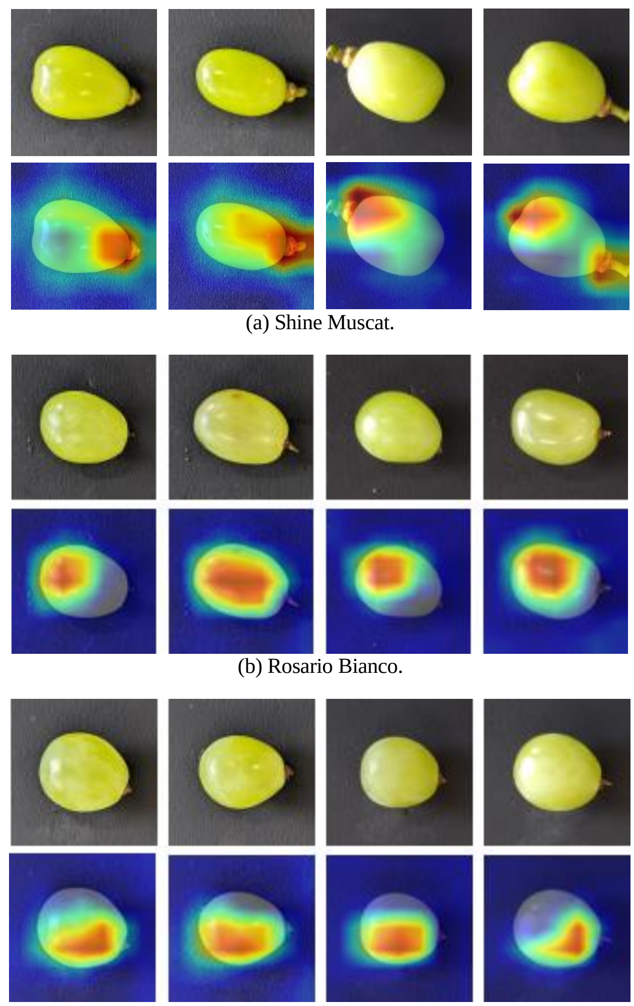


(a) Shine Muscat.

(b) Rosario Bianco.

(c) Muscat of Alexandria.

Figure 4. Heat maps of leaf images of each variety using VGG16 and Grad-CAM.



(a) Shine Muscat.

(b) Rosario Bianco.

(c) Muscat of Alexandria.

Figure 5. Heat maps of mature berry images of each variety using VGG16 and Grad-CAM.

 ISBN: 978-1-68558-324-8

images of Rosario Bianco (Fig. 4(b)). For the images of Muscat of Alexandria, the outline of a leaf appeared to be the focus of attention (Fig. 4(c)). The model for mature berry images may observe the surface of the berries in Rosario Bianco and Muscat of Alexandria (Fig. 5(b) and (c)). However, it possibly focused on the pedicel to recognize the Shine Muscat (Fig. 5(a)).

IV. Conclusions

To examine whether image recognition can be used to distinguish crop varieties, images of the three grape varieties were collected, and classification models were generated. The test results demonstrated that all models were highly accurate. VGG16 and ResNet50 attained test accuracies of 100% for the leaf and berry images. The accuracies of ViT were 96.1% and 99.8% for leaves and young berries, respectively. Grad-CAM clarified that VGG16 focused on different parts of an image depending on the variety. Based on these results, any other DNN-based classification model is most likely to provide high accuracy. In other words, these three types of classification issues may be comparatively easy to handle in DNN-based image recognition. This study showed that the three grape varieties could be classified based on images of both leaves and berries. However, a model that can classify only three varieties is not sufficiently practical for making infringement decisions, and the number of varieties to be recognized should increase. Additional training data must be collected to update and verify the models.

In general, image recognition models that maintain 100% classification accuracy for any unseen image are rare, even though they are trained on a large amount of training data. The models trained in this study are not expected to maintain high accuracy and may produce some misclassifications in unseen images, such as those captured at different sites and in different years. Therefore, distinguishing variety infringements based on images alone is unreliable, and DNA analysis or other conventional methods are still required to obtain a correct judgement. Image recognition techniques should be utilized to screen for suspicious varieties prior to DNA analysis, which may provide an effective method for infringement detection.

In conclusion, this study represents a stepping stone for future development of infringement prevention technology. By increasing the number of cultivars that can be identified with image recognition while maintaining the accuracy, this image-based approach will become more practical, contributing to prompt detection of illegal cultivation and overseas outflow of registered varieties.

References

[1] Queensland District Registry, Federal Court of Australia, No: QUD184/2019, Oct. 2019, [Online] Available from: https://www.comcourts.gov.au/file/Federal/P/QUD184/2019/3851895/event/30050693/document/1482902 2024.10.22

[2] AgTech MediaGroup, “IFG Triumphs in Plant Variety Rights Case in China,” iGrow News, July 2023, [Online] Available from: https://igrownews.com/ifg-triumphs-in-plant-variety-rights-case-in-china/ 2024.10.22

[3] C. Collen, “Italian court issues landmark grape ruling,” EUROFRUIT, Nov. 2022, [Online] Available from: https://www.fruitnet.com/eurofruit/italian-court-issues-landmark-grape-ruling/247547.article 2024.10.22

[4] The Yomiuri Shimbun, “Protect Japan’s Valuable Agricultural Products through Revised Seed Law,” The Japan News, Dec. 2020, [Online] Available from: https://japannews.yomiuri.co.jp/editorial/yomiuri-editorial/20201225-157388/ 2024.10.22

[5] M. Okamoto et al., “A target cultivar-specific identification system based on the chromatographic printed array strip method for eight prominent Japanese citrus cultivars,” Breeding Science vol. 73(2), pp. 146–157, April 2023, doi: 10.1270/jsbbs.22065.

[6] G. Würtenberger, ”The Object Of Plant Variety Protection and the Scope Deriving Therefrom,” CPVO, Mar. 2017, [Online] Available from: https://cpvo.europa.eu/en/news-and-events/articles/object-plant-variety-protection-and-scope-deriving-therefrom 2024.10.22

[7] Ministry of Agriculture, Forestry and Fisheries Japan, “Plant Variety Protection System of Japan,”, April 2022, [Online] Available from: https://www.hinshu2.maff.go.jp/en/about/leaflet.pdf 2024.10.22

[8] Y. LeCun and Y. Bengio, “Convolutional Networks for Images, Speech, and Time Series,” The handbook of brain theory and neural networks, 1995, [Online] Available from: https://www.researchgate.net/publication/216792820_Convolutional_Networks_for_Images_Speech_and_Time-Series 2024.10.22

[9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature vol. 521(7553), pp. 436-444, May 2015, https://doi.org/10.1038/nature14539

[10] M. Albahar, “A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities,” vol. 13(3):540, 2023, https://doi.org/10.3390/agriculture13030540

[11] I. V. Srihith, “A Short Review on Deep Learning in Agriculture,” Journal of Advanced Research in Artificial Intelligence & It's Applications vol. 1(3), pp. 23-39, June 2024, https://doi.org/10.5281/zenodo.11612440

[12] I. Kunduracioglu and I. Pacal, “Advancements in Deep Learning for Accurate Classification of Grape Leaves and Diagnosis of Grape Diseases,” Journal of Plant Diseases and Protection vol. 131, pp. 1061–1080 Mar. 2024, https://doi.org/10.1007/s41348-024-00896-z

[13] K. Shirasawa et al., “*De novo* Whole-Genome Assembly in an Interspecific Hybrid Table Grape ‘Shine Muscat’,” DNA Research vol. 29, pp. 1–9, Nov. 2022, https://doi.org/10.1093/dnares/dsac040

[14] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv:1409.1556, April 2014, https://doi.org/10.48550/arXiv.1409.1556

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385, Dec. 2015, https://doi.org/10.48550/arXiv.1512.03385

[16] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” In International Conference on Learning Representations, Jan. 2021, [Online] Available from: https://openreview.net/pdf?id=YicbFdNTTy 2024.10.22

[17] D. Zhou, Z. Yu, E. Xie, C. Xiao, A. Anandkumar, J. Feng, and J. M. Alvarez, “Understanding the Robustness in Vision Transformers,” arXiv:2204.12451, Nov. 2022, https://doi.org/10.48550/arXiv.2204.12451

[18] The PyTorch Foundation, TorchVision [Online] Available from: https://pytorch.org/vision/stable/models.html 2024.10.22

[19] The PyTorch Foundation, PyTorch [Online] Available from: https://pytorch.org/ 2024.10.22

[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vadantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," arXiv:1610.02391, Dec. 2019, https://doi.org/10.48550/arXiv.1610.02391

[21] J. Gildenblat, *Introduction: Advanced Explainable AI for computer vision*, [Online] Available from: https://jacobgil.github.io/pytorch-gradcam-book/introduction.html 2024.10.22

[22] B. Franczyk et al., "Deep learning for grape variety recognition," Procedia Computer Science vol. 176, pp. 1211–1220, Oct. 2020, https://doi.org/10.1016/j.procs.2020.09.117

[23] C. S. Pereira, R. Morais, and M. J. C. S. Reis, "Deep Learning Techniques for Grape Plant Species Identification in Natural Images," Sensor vol. 19(22):4850, Nov. 2019, https://doi.org/10.3390/s19224850