Enhancing agricultural classification models through data augmentation and advanced deep learning techniques

Tien Dang*, **& Long D. Phan**

Faculty of Information Technology, Nong Lam University, Ho Chi Minh City, Vietnam

ARTICLE INFO

Research Paper

Received: August 31, 2024
Revised: October 10, 2024
Accepted: November 18, 2024

Keywords

Agricultural datasets
Agricultural informatics
CNNs
Data augmentation
ViTs

***Corresponding author**

Dang Minh Tien
Email:
tien.dangminh@hcmuaf.edu.vn

ABSTRACT

In the field of agricultural data analysis, achieving high quality classification modeling remains a significant challenge due to the inherent variability and complexity of agricultural datasets. This study investigated cutting-edge approaches to enhance model performance through data augmentation techniques and the application of advanced deep learning models to artificially enlarge the training dataset, thereby improving model generalizability and robustness. Additionally, the study evaluated the efficacy of state-of-the-art models (i.e., ViT-Ti/16, CaiT-XXS-24, XCiT-T12, Resnet26, ConvNeXt-T) for agricultural data analysis. The experimental results revealed a marked improvement in terms of accuracy and F1-Score when applied data augmentation into the training session. This underscored the potential of these techniques to significantly advance the field of agricultural informatics. Briefly, the findings contributed to the development of more reliable and high performance models for agricultural practices.

Cited as: Dang, T. M., & Phan, L. D. (2024). Enhancing agricultural classification models through data augmentation and advanced deep learning techniques. *The Journal of Agriculture and Development* 23(Special issue 2), 25-32.

1. Introduction

In recent years, the application of information technology in agriculture has attracted significant interest, particularly in the monsoon-influenced climate region such as Vietnam. This interest is driven by the potential of advanced technologies to transform traditional agricultural practices and enhance productivity. Among these technologies, machine learning has emerged as a powerful tool, offering innovative solutions to various agricultural challenges. In information technology, especially in computer vision, models which are designed for image classification, have shown remarkable success in tasks such as crop disease detection, yield prediction, and soil health monitoring. These models can analyze vast amounts of visual data, providing insights that were previously unattainable.

However, one of the critical challenges in applying deep learning to agricultural datasets is the limited availability of labeled data. Therefore, data augmentation has been proposed as a technique that generates new training samples from existing data, has proven to be an effective solution to this problem. By increasing the diversity and size of training datasets, data augmentation enhances the robustness and generalization capabilities of deep learning models. For instance, Vision Transformers (ViT) (Dosovitskiy et al., 2020) have garnered attention for their ability to process and analyze visual data using self-attention mechanisms, capturing long-range dependencies in images. Class-Attention in Image Transformers (CaiT) (Touvron et al., 2021) and Cross-Covariance Image Transformers (XCiT) (El-Nouby et al., 2021) extend the capabilities of ViT by introducing novel architectural modifications that enhance performance and efficiency. Residual Networks (Resnet) (He et al., 2015), renowned for its deep residual learning framework, addresses the vanishing gradient problem, allowing the training of very deep networks. A family of pure ConvNets dubbed ConvNeXt (Liu et al., 2022) was introduced with their hierarchical structure of convolutional layers, remain a cornerstone in the field of image recognition, offering robust feature extraction and classification capabilities.

In this paper, we presents a comprehensive investigation of different deep learning models applied to classification tasks for a specific agricultural dataset, specifically the Vietnamese leaf dataset (Phan & Tran, 2022), with a particular focus on the impact of data augmentation. The models under consideration include ViT, CaiT, ConvNeXt. In the context of agricultural image classification, these models offer unique strengths and limitations. By evaluating their performance, this paper aims to identify the most effective approaches for leveraging data augmentation to improve classification accuracy and robustness.

2. Related Works

The application of artifitial inteligent in agricultural image classification has seen significant advancements, with each model offering unique advantages. In this section, we inform several key models and emphasizes the role of fundamental data augmentation techniques in enhancing classification performance.

2.1. Efficient classification models

Recently, Convolutional Neural Networks have been becoming famous due to it performance relevant to image tasks. For instance, ConvNeXt have established themselves as foundational models in image classification due to their hierarchical structure of convolutional layers. In agriculture, ConvNeXt is widely used for tasks such as soil texture classification, crop yield estimation, and remote sensing analysis, continuing to provide robust feature extraction and classification capabilities (Bhuyan & Singh, 2024). Their widespread use and proven effectiveness make ConvNeXt a staple in agricultural image processing. Moreover, ResNet build on the strengths of CNNs with their deep residual learning framework, which mitigates the vanishing gradient problem and allows for the training of very deep networks. In agricultural contexts, ResNet has demonstrated robust performance in tasks like crop disease detection and fruit sorting by effectively capturing hierarchical features from images. Studies indicate ResNet's reliability and efficacy in various agricultural applications.

Beside, in the past few years, Transformer-based models have been proposed as powerful alternatives to convolutional neural networks. ViT leverages self-attention mechanisms to capture long-range dependencies within images, making them particularly effective in handling complex visual data. Research has demonstrated ViT's ability to differentiate between healthy and diseased crops by analyzing detailed visual features, leading to significant improvements in agricultural image classification tasks. However, ViT has several drawbacks, one of those is the lack of local communication. Therefore, CaiT enhances ViT by introducing class-attention layers, which focus on the Class token information, and XCiT addresses some limitations of ViT by incorporating cross-covariance information between image patches, and it Local Patch Interaction (LPI) was introduced as a module for capturing the local information. This approach enables XCiT to capture fine-grained details in high-resolution images, making it particularly effective for classification tasks.

2.2. Data augmentation

Data augmentation is essential for improving the performance of machine learning models, particularly when working with limited datasets. In computer vision, where obtaining diverse and sufficient training data can be challenging, augmentation techniques address this issue by artificially expanding the training set with various transformations. By increasing both the size and diversity of training data, data augmentation boosts the robustness and generalization capabilities of models, resulting in enhanced performance across a range of tasks. Traditional data augmentation methods include geometric transformations such as rotation, scaling, cropping, and flipping, as well as color jittering and noise addition. These techniques can further enhance model performance by simulating different lighting conditions and image quality degradations, and have been shown to effectively reduce overfitting and improve model robustness across various tasks (Shorten & Khoshgoftaar, 2019). For instance, rotations and translations can help models become invariant to slight changes in object position, while color adjustments can aid in generalizing across different lighting conditions.

In addition to these basic techniques, more sophisticated methods such as MixUp and CutMix have been introduced to further enhance data augmentation. MixUp, proposed by (Zhang et al., 2017), generates new training examples by blending two images and their corresponding labels, creating a smoother decision boundary and improving model robustness. CutMix, introduced by (Yun et al., 2019), goes a step further by cutting and pasting patches from one image onto another, which helps the model focus on different regions and learn more invariant features.

Overall, data augmentation remains a critical component in the development of robust machine learning models, particularly in scenarios with limited data. Its ability to enhance model generalization and reduce overfitting makes it an indispensable tool in modern machine learning pipelines.

3. Experiments

In this section, we demonstrate the effectiveness of several models from the timm librabry such as: ViT, CaiT and XCiT - those models are represent for the transformer-based model, and for the CNNs models, we use Resnet26 and ConvneXt. By applying serveral data augmentation on the Vietnamese leaf dataset which proposed by Phan & Tran (2022), we will evaluate the performance of these data augmentation on different models.

3.1. Dataset

According to (Phan & Tran, 2022) they introduced a new Vietnamese leaf dataset. This dataset comprises of 6800 images for training set and 1707 images for testing. Those leaves are from 15 different food and industrial crops, as well as certain fruit trees. These include corn, sorghum, sweet potato, tapioca, potato, rice, cassava, rubber, coffee, cashew, pepper, rambutan, durian, green beans, and peanuts. However, the authors have eliminated peanuts, mung beans, sago, and rambutan from the dataset after testing due to insufficient sample quantity and poor image quality. Sorghum was also discarded because its leaf shape was too similar to that of corn. To address this issue, we apply several data augmentation and utilze some advanced machine learning models and we decided to use all 15 different type of leaves for our experiments.

3.2. Metrics for classification

Evaluating the performance of classification models involves several key metrics that provide insights into different aspects of the model's capabilities. Assuming that we have a confussion matrix with 4 similarity parameters TP, TN, FP, FN represent for True Positive, True Negative, False Positive, and False Negative, respectively. Here are several metrics that we used:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F}_1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In there, the F1-score is the special one due to it balances precision and recall, making it especially useful for classificational tasks. Evaluating models using the F1-score ensures that both FP and FN are considered, leading to a more robust performance.

3.3. Experimental setting

3.3.1. Data augmentation

Our objective is to evaluate the effectiveness of various data augmentation techniques in enhancing the performance of vision models for agricultural image classification. To begin with, we resize the full image to 224 x 224 pixels using bicubic interpolation to normalize the input image. After that, we decided to apply several augmentation methods to enhance the robustness of our models. We utilize Random Horizontal Flip with a probability of 0.5, allowing the model to learn from images that are flipped horizontally, which simulates different viewing angles. Random Rotation is applied to rotate images up to 50 degrees, helping the model become invariant to rotational changes of the input data. Random Erasing (Zhong et al., 2017) with a probability of 0.3 to randomly erase some pixels in the images, obliging the model to focus on the entire context rather than specific features. Additionally, Random Augment (Cubuk et al., 2019) is employed with a probability of 0.5, randomly applying a combination of augmentations to further diversify the training data.

Furthermore, during the training phase, we incorporate Mixup and Cutmix techniques, each with a probability of 0.15. Mixup creates new training samples by blending pairs of images and their corresponding labels, which makes model to be more generalized. On the other hand, Cutmix combines patches from different images, encouraging the model to recognize the label based on fragmented information.

3.3.2. Model and optimizer

We conduct experiments on several tiny variant model based on ViTs and CNNs, all implemented via the timm library. More precisely like, we explore transformer-based models including ViT-Ti/16, XCiT-T12, and CaiT-XXS-24. For CNN-based models, we use ResNet26 and ConvNeXt-T. For each model, we employ a tailored optimization strategy to ensure optimal performance. For the transformer models, we employ the AdamW optimizer due to its ability to handle sparse gradients and large learning rates. The learning rate is set to with a weight decay of 0.05 to mitigate overfitting. Additionally, we utilize a cosine annealing learning rate scheduler, which gradually reduces the learning rate following a cosine curve, and enhancing the model's convergence during training. Besides, for the CNN-based models, such as ResNet26 and ConvNeXt-T, we also use the AdamW optimizer with similar hyperparameters. However, to further enhance performance, we incorporate several regularization such as Stochastic Depth (Huang et al., 2016) and Label Smoothing (Szegedy et al., 2015), both with a probability of 0.1. Stochastic Depth introduces regularization by randomly skipping layers during training, while Label Smoothing helps to prevent overfitting by softening the target labels.

Table 1. Hyperparameter during training phrase

Training setting	Configuration	
	Transformer-based model	CNN-based model
Input size	224 x 224	
Interpolation	Bicubic	
Optimizer	AdamW	
Learning rate	Base LR = 5×10 − 4 with cosine decay	
Optimizer momentum	β_1, β_2 = 0.9,0.999	
Weight decay	0.05	
Batch size	128	
Total epochs	50	
Warmup epochs	50	
Random Horizontal Flip	0.3	
Random Rotation	50 degrees	
Random Erasing	0.3	
Random Augment	0.5	
Mixup	0.15	
Cutmix	0.15	
Stochastic Depth	-	0.1
Label Smoothing	-	0.1

3.4. Results

In this section, we illustrate our experimental result after training each model for 50 epochs and the setting of hyperparameter is shown in Table 1.

Table 2. Comparison of various model results on Vietnamese leaf dataset

Model	Params (M)	Layers Blocks	Embed-size	Accuracy	F1-score
ViT-Ti/16	5.8	12	192	92.72	92.06
XCiT-T12	6.7	12	192	97.89	97.41
CaiT-XXS-24	12	24 + 2	192	94.42	93.89
Resnet26-t	16	26 blocks {2,2,2,2}	{32, 24, 48, 64}	98.76	98.47
ConvNeXt-T	28.6	18 blocks {3,3,9,3}	{96, 192, 384, 768}	98.47	98.23

As shown in Table 2, all models on our experiments have easily achieved high performance. More precisely like, CNN-based models have surpassed Transformer-based models due to its advancements in local information. To be more specific, Resnet26-t achieved state-of-the-art accuracy and F1-score (98.76 and 98.47, respectively). The second winner is CovNeXt-T with 98.47 in accuracy and 98.23 in F1-score. For transformer models, because of the LPI layer helped collect the local information, XCiT gained a closed-call to Resnet26-t and ConvNeX-T (97.89 for accuracy and 97.47 for F1-score). Finally, although ViT's drawback is that it needs to be trained on a large dataset, it still achieved a high performance with our setting in such a small dataset.

3.5. Ablation study

This section presents the effects of numerous data augmentation and regularization methods. According to Section 3.3.1, the experiment was conducted to write down the performance of each data augmentation by removing one factor at the time, then we make a comparison with the default setting. Also, we employed a few more techniques including Random Crop, Gaussian Blur, Color Jitter, and Random Gray Scale.

Table 3. Comparison of the effects of various data augmentation on various models

Action	Augmentation	Model				
		ViT - T/14	XCiT-T12	CaiT-XXS-24	ResNet26	ConvNeXt -T
	Default	92.72	97.89	94.42	98.76	98.47
Adding	Random Crop	90.39	96.86	93.4	97.63	98.29
	Color Jitter	92.25	97.6	93.97	98.64	98.3
	Random Gray Scale	92.1	97.73	91.67	98.24	97.95
Removing	Random Horizontal Flip	91.97	96.3	92.38	97.68	96.72
	Random Rotation	91.82	97.12	92.95	98.42	97.68
	Random Erasing	90.62	95.32	92.51	98.6	97.96
	Random Augment	92.91	97.82	95.18	99.01	98.62
	Mixup	91.41	95.79	93.65	96.79	98.25
	Cutmix	91.68	95.89	91.67	97.3	98.13
	Stochastic Depth	-	-	-	98.21	96.4
	Label Smoothing	-	-	-	97.6	97.81

As shown in Table 3, although Random Augment was slightly improved the performance of XCiT, we still decided it was the only data augmentation method that restrain the performance of almost all models. Without it, Resnet26-t performance achieved 99.01% in accuracy which surpassed default setting. For that reason, we suggest that Random Augment should not be installed. Random Crop, Color Jitter, and Random Gray Scale was slightly decreased the accuracy of model, therefore, we also removed them out of our setting.

4. Conclusions

In this paper, we aimed to comprehensively evaluate the effectiveness of various data augmentation on different models on the Vietnamese leaf dataset collected by previous researchers We have presented a comprehensive result of both Vision Transformer-based models (ViT-Ti/16, XCiT-T12, CaiT-XXS-24) and Convolutional Neural Networks (ResNet26, ConvNeXt-T) on our setting. Our findings indicate that data augmentation techniques significantly enhance the robustness and generalization capabilities of AI models in agricultural image classification tasks. This improvement contributes to the overall advancement of precision agriculture. Our work underscores the importance of employing diverse data augmentation methods to address the challenges posed by limited agricultural datasets, leading to more resilient and effective agricultural technologies.

Conflict of interest

We hereby declare that this is a scientific research work conducted by our team. The data used in the analysis process has clear origins and has been published in accordance with the regulations. All research results are the product of an honest and objective process of inquiry and analysis. These results have not been published in any other research.

References

Bhuyan, P., & Singh, P. K. (2024). Evaluating deep CNNs and vision transformers for plant leaf disease classification. In Devismes, S., Mandal, P. S., Saradhi, V. V., Prasad, B., Molla, A. R., & Sharma, G. (Eds.), *Proceedings of The 20th International Conference on Distributed Computing and Intelligent Technology ICDCIT 2024, Bhubaneswar, India, January 17-20, 2024* (293-306). Zug, Switzerland: Springer Cham. https://doi.org/10.1007/978-3-031-50583-6_20.

Cubuk, E. D., Zoph, B., Shlens, J., & Le, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. In Boult, T., Medioni, G., & Zabih, R. (Eds.), *Proceedings of The 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, June 14-19* (3008-3017). New Jersey, USA: Institute of Electrical and Electronics Engineers - IEEE. https://doi.org/10.1109/cvprw50498.2020.00359.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* 2010, 11929. https://doi.org/10.48550/arXiv.2010.11929.

El-Nouby, A., Touvron, H., Caron, M., Bojanowski, P., Douze, M., Joulin, A., Laptev, I., Neverova, N., Synnaeve, G., Verbeek, J., & Jegou, H. (2021). Xcit: Cross-covariance image transformers. *arXiv* 2106, 09681v2. https://doi.org/10.48550/arXiv.2106.09681.

He, M. K., Zhang, G. X., Ren, Q. S., & Sun, J. (2015).

Deep residual learning for image recognition. In Tuytelaars, T., Li, F. F., & Bajcsy, R. (Eds.), *Proceedings of The 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 27-30* (770-778). New Jersey, USA: Institute of Electrical and Electronics Engineers - IEEE. https://doi.org/10.1109/cvpr.2016.90.

Huang, G., Sun, Y., Liu, Z., Sedra, D., & Weinberger, K. (2016). Deep networks with stochastic depth. In Leibe, B., Sebe, N., Matas, J., & Welling, M. (Eds.), *Proceedings of Computer Vision - ECCV 2016: 14th European Conference Part IV, Amsterdam, The Netherlands, October 11-14* (646-661). Zug, Switzerland: Springer Cham. https://doi.org/10.1007/978-3-319-46493-0_39.

Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Chellappa, R., Matas, J., Quan, L., & Shah, M. (Eds.), *Proceedings of The 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops New Orleans, Louisiana, June 19-24, 2022* (11976-11986). New Jersey, USA: Institute of Electrical and Electronics Engineers - IEEE. https://doi.org/10.1109/CVPR52688.2022.01167.

Phan, L. D., & Tran, T. S. (2022). Applying convolution neural networks for leaf image recognition with the vietnamese leaf image database. In *Proceedings of The 4th International Conference on Sustainable Agriculture and Environment* (81-94). Ho Chi Minh City, Vietnam: Nong Lam University.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data* 6(1), 1-48. https://doi.org/10.1186/s40537-019-0197-0.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the inception architecture for computer vision. In Tuytelaars, T., Li, F. F., & Bajcsy, R. (Eds.), *Proceedings of The 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 27-30* (2818-2826). New Jersey, USA: Institute of Electrical and Electronics Engineers - IEEE. https://doi.org/10.1109/cvpr.2016.308.

Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G., & Jégou, H. (2021). Going deeper with image transformers. In Berg, T., Clark, J., Matsushita, Y., & Taylor, C. J. (Eds.), *Proceedings of The 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, October 11-17* (32-42). New Jersey, USA: Institute of Electrical and Electronics Engineers - IEEE. https://doi.ieeecomputersociety.org/10.1109/ICCV48922.2021.00010.

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). CutMix: Regularization strategy to train strong classifiers with localizable features. In Lee, K. M., Forsyth, D., Pollefeys, M., & Tang, X. (Eds.), *Proceedings of The 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, October 27-November 2,* (6022-6031). New Jersey, USA: Institute of Electrical and Electronics Engineers - IEEE. https://doi.ieeecomputersociety.org/10.1109/ICCV.2019.00612

Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). Mixup: Beyond empirical risk minimization. In Bengio, Y., & LeCun, Y. (Eds.), *Proceedings of The 6th International Conference on Learning Representations - ICLR 2018, Vancouver, Canada, April 30-May 3* (1-13). https://doi.org/10.48550/arXiv.1710.09412.

Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2017). Random erasing data augmentation. In *Proceedings of The AAAI Conference on Artificial Intelligence* (13001-13008). https://doi.org/10.1609/aaai.v34i07.7000.