Assessing the Effectiveness of Augmentation Techniques in Enhancing Plant Leaf Disease Classification

MD Ilias Bappi*, David J. Richter*, Kyungbaek Kim
Dept. of Artificial Intelligence Convergence, Chonnam National University, Gwangju, South Korea

Abstract

Plant leaf disease identification is vital for food security, as plant diseases cause significant agricultural losses. Early detection of symptoms on plant leaves is crucial for minimizing yield loss. Traditional monitoring is labor-intensive, prompting the use of deep learning for automated detection. However, the lack of large-scale, high-quality, open datasets remains a challenge, with many being closed-source or suboptimal. Image augmentation techniques can expand dataset size, improving model performance without the need for additional data collection. In this work, we explore the impact of various augmentation techniques on the performance of deep learning models, particularly in lab and field datasets. Our results show that techniques like color, transformation, and noise augmentation significantly enhance model accuracy, with combined augmentation yielding the highest accuracy, especially for field datasets. These findings underscore the effectiveness of augmentation in improving deep learning models for plant leaf disease identification.

Keywords : Deep Learning | Plant Leaf Disease Classification | Transformer | CNN | Transfer Learning

I. INTRODUCTION

Agriculture is crucial for global food security, but crop diseases can reduce yields by 20–40% [1]. Early disease identification is vital, yet manual inspection is labor-intensive and slow [2]. Deep learning, particularly Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), offer potential for automating disease classification [3]. However, both methods require large datasets, and high-quality, domain-specific datasets are scarce [4]. Lab-based datasets are large but may lack real-world applicability, field-based datasets are more representative but often too small for robust training. Data augmentation techniques, such as color, transformation, and noise-based modifications, can help expand datasets and improve model generalization [5]. This study explores effective augmentation strategies for plant leaf disease classification using CNNs and ViTs, comparing the performance of lab and field datasets to guide future research and model development. This study provides contributions in the following ways:

- Comprehensive exploration of combined augmentation methods (color, transformation,

* The first 2 authors contributed to this work equally.
* This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry(IPET) through the Agriculture and Food Convergence Technologies Program for Research Manpower development, funded by Ministry of Agriculture, Food and Rural Affairs(MAFRA)(project no. RS-2024-00397026, 34%). This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(IITP-2024-RS-2022-00156287, 33%). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629, 33%) grant funded by the Korea government(MSIT).

Manuscript 2025.01.15
Revised : 2025.02.05
Confirmation of Publication: 2025.02.12
Corresponding Author email: kyungbaekkim@jnu.ac.kr

and noise based) to enhance potato leaf disease detection accuracy across diverse datasets.

- Evaluated lab-based (PLD) [14] and field-based (PLDDUE) [15] datasets, finding field datasets that benefited most from combined augmentations, improving generalization.
- Analyzed augmentation performance across CNNs (MobileNetV3, DenseNet121, EfficientNetV2B0, ResNet50V2, and VGG19) and transformer-based models (ViT and Transformer), for both augmentation methods and both datasets.
- Validated that augmentation strategies improve disease detection accuracy and model reliability, making datasets more suitable for real-world agricultural applications.

The rest of the paper is organized as follows. Section II describes the related work. Section III describes the background of this work, including datasets and deep learning techniques. Section IV mentions the methodology used in this study. Section V elaborates the results of each experiment and the conclusive summary can be found in Section VI.

II. RELATED WORK

Recent studies highlight the effectiveness of data augmentation in improving deep learning models for plant leaf disease detection. Eunice et al. [6] applied rotation, flipping, and zooming to the PlantVillage dataset, achieving up to 99.81% accuracy with DenseNet121. Fulle et al. [7] augmented a field dataset from Ethiopia using shearing and brightness shifting, reaching 99.93% accuracy with MobileNetV3 Small. Paiva et al. [8] used rotation and brightness adjustments on the cassava dataset, achieving 74.77% accuracy with DenseNet169. Lanjewar et al. [9] applied flip and rotation augmentations on a potato leaf dataset, reaching 99.67% accuracy with DenseNet169. Ramya et al. [10] achieved 99.7% accuracy on a subset of PlantVillage's tomato images using flipping and rotation. Salam et al. [11] augmented a mulberry dataset and reached 94.4% accuracy with ResNet50, while Shah et al. [12] applied multiple augmentations to a grape dataset, achieving 98% accuracy with ResNet50V2. These studies demonstrate that data augmentation significantly boosts model performance. Building on this, this study explores the impact of various augmentation methods for potato leaf disease detection across lab and field datasets, providing insights for future research.

III. BACKGROUND

Deep learning (DL) techniques, particularly convolutional neural networks (CNNs) and vision transformers (ViTs), offer promising solutions for automating disease detection through image classification. However, these models require large, diverse and high-quality datasets, and data augmentation techniques like color adjustments and geometric transformations help address this challenge by expanding datasets and improving model performance. While augmentation has been effective in controlled environments, less is known about its performance in real-world field conditions.

A. DATASETS

1) Lab Dataset

Lab datasets are datasets that were captured in perfect conditions in a lab like setting. This means that they are taken in front of a simple background, a single leaf at a time. This makes it much easier for DL models to identify and understand the leaves and tends to generate high accuracy easily, making it hard to judge the models. Also, models trained with lab data often lack the robustness to be used in real field conditions

2) Field Dataset

Field Datasets are datasets that were taken in real field conditions, with noisy backgrounds that include soil, other leaves, branches and other details not belonging to the actual leaf. This makes these images much harder to train and understand and creates more robust models.

B. AUGMENTATION

Data augmentation is the process of increasing the size of a dataset by altering the data in a way that does not falsify it, but in a way that makes it different enough from the original data to offer new and diverse data to train the models on. In image tasks, augmentation often includes noise augmentation (adding new noise to the image), transforming augmentation (e.g. rotating) and color altering augmentation (e.g. contrast shifting).

C. ARCHITECTURE

1) CNN based Architecture

CNNs are highly effective for image classification due to their hierarchical feature extraction. Here, we used MobileNetV3, DenseNet121, EfficientNetV2B0, ResNet50V2, and VGG19, all pretrained on ImageNet [6] to improve performance on plant leaf disease classification. These models require a 224x224x3 input size. MobileNetV3 uses depth wise separable convolutions, making it lightweight, with 16 layers and a 0.3 dropout rate to prevent overfitting. DenseNet121 uses dense blocks for feature reuse, with 4 blocks, 121 layers, and a 0.2 dropout rate. EfficientNetV2B0 applies compound scaling with 12 convolution stages, optimizing depth, width, and resolution [23]. ResNet50V2 incorporates residual blocks to prevent vanishing gradients, with 50 layers and a 0.5 dropout rate [20]. VGG19 has 19 convolution layers with 3x3 kernels, max pooling, and a 0.5 dropout rate [21]. All models use SoftMax activation for 3 class classifications: healthy, late blight, and early blight.

2) Transformer based Architecture

Recently, transformer-based models have gained attention for their ability to capture long range dependencies in images through attention mechanisms. In this study, we used the Transformer, and Vision Transformer (ViT), all pretrained on ImageNet [6].

These models process images by dividing them into patches (16x16) and using multi head self-attention to capture contextual relationships. The Transformer model divides the image into patches, flattens them, and passes them through 12 encoder layers with 8 attention heads and a hidden dimension of 512. Each attention block is followed by a position-wise feed-forward network with ReLU activation, and dropout is applied with a rate of 0.1. The output is then passed through a fully connected layer with SoftMax activation for classification. ViT also divides the image into non-overlapping patches (16x16), applies multi-head self-attention with 12 layers, each having 12 attention heads and a hidden size of 768. Position embeddings are included, and the final classification is done using an MLP head with dropout applied at 0.1. SoftMax activation is used to predict the three disease classes. These transformer models excel at capturing long range dependencies and are effective for plant leaf disease classification [21].

IV. METHODOLOGY

To test whether or not augmentation, or even specific augmentation, positively impact the results, we tested the impact of color, transformation, and noise-based data augmentation on a lab and a field dataset with MobileNetV3Large [19], DenseNet121 [22], EfficientNetV2B0 [23], ResNet50V2 [24], VGG19 [21], Transformer [25], and ViT [26].

A. DATASETS

To evaluate the effectiveness of different augmentation techniques for plant leaf disease classification, we trained models on two datasets: the lab-based "Potato Disease Leaf Dataset (PLD)" [13,14] with 4,072 images across 3 classes (healthy, late blight, early blight) and the field-based "Potato Leaf Disease Dataset in Uncontrolled Environment (PLDDUE)" [15,16] with 3,076 images across 7 classes (Bacteria, Fungi, Healthy, Nematode, Pest, Phytophthora, Virus). To ensure comparability, we reduced PLDDUE to 3 classes (Bacteria, Fungi, Healthy), resulting in 1,518 images before augmentation. All images were resized to 224x224x3. The data was split into 64% training, 20% testing, and 16% validation, with augmentation applied only to the training and validation sets. Test data remained original across all augmentation configurations. Sample images from both datasets are shown in Figure 1.

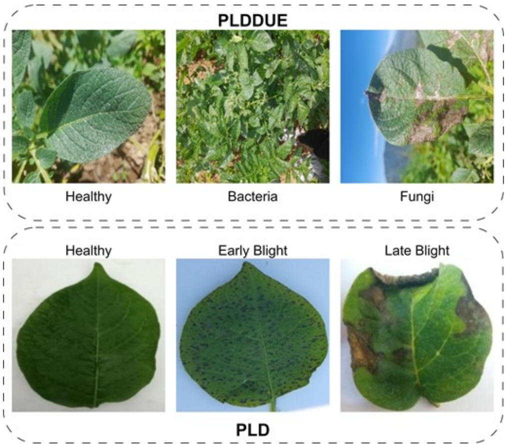


Fig. 1. Samples of both datasets, image classes.

B. DATA AUGMENTATION

To increase the size of the datasets a set of different data augmentation techniques were applied them and then tested them all combined. One thing to note is that the original image will always be kept with the augmented images, so if for example 5 augmentations are applied to an image, then this will result in 6 images being output (5 augmented and the original). Augmentation was carried out with numpy 1.24.3 and OpenCV 4.10.0.84 on Python 3.11.5. The exact code used can be found in [18] for replication.

1) Color

The color-based augmentation techniques used here are brightness adjustments, contrast adjustments, color space channel shifting, and hue channel adjustments, and all parameters used to augment are in Table 1.

Table 1. Color augmentation methods.

Method	Value
Brightness	-0.75, 0.75
Contrast	-0.25, 0.25
Channel Shift	75
Hue Shift	-20, 20

2) Noise

To add noise to the image, we use Gaussian noise that we apply to the image. The values we used mean: 0, std: 1.

3) Transformation

To augment the images with transformations, we apply flipping images (both horizontally and vertically) and rotation (90 degrees, 180 degrees, and 270 degrees. The parameters used are listed in Table 2.

Table 2. Transformation augmentation methods.

Method	Value
Flip	horizontal, vertical
Rotation	90°, 180°, 270°

C. MODELS

The models used to train on the different augmented PLD and PLDDUE datasets are MobileNetV3Large, DenseNet121, EfficientNetV2B0, ResNet50V2, VGG19, Transformer, and, ViT. CNN models were picked partially based on their performance in recent benchmarks [17] one per architecture family to obtain and present diverse results, as well as on their use in recent work. All CNN models are largely based on the idea of convolutional layers. The ViT based models on the other hand are transformer image classifier models. To train all of the models, transfer learning was employed. All models were pretrained on the ImageNet dataset before being retrained on the datasets mentioned above. All layers, including the feature-extracting layers in the CNNs and the ViTs were unfrozen for this and subsequently trained. The hyperparameters used can be seen in Table 3.

Table 3. Hyperparameters used to train the models.

Method	Value
Learning Rate	0.0001
Loss	Categorical Cross Entropy
Optimizer	Adam
Epochs	30
Batch Size	32

V. RESULTS

The analysis of Table 4 shows that augmentation strategies significantly improve model performance on the PLDDUE dataset. EfficientNetV2B0 achieves the highest accuracy (99.34%) and F1-Score (99.49%) with the "augmented all" strategy, while its performance drops to 97.70% without augmentation. MobileNetV3Large also benefits from this strategy, reaching 98.36% accuracy, up from 97.38% without augmentation. DenseNet121 shows its best performance (97.05%) with "augmented all" but slightly declines to 95.41% without. VGG19 achieves its highest accuracy (96.07%) with full augmentation but drops to 94.10% without. ResNet50V2 shows limited improvement, with 95.74% accuracy even with full augmentation, dropping to 93.77% without. Overall, the "augmented all" strategy is the most effective, especially for EfficientNetV2B0, demonstrating that combining spatial and visual augmentations improves model accuracy and robustness. Table 5 shows that augmentation strategies significantly improve performance on the PLD dataset. EfficientNetV2B0 achieves the highest accuracy (99.75%) with the "augmented all" strategy, highlighting the effectiveness of

Table 4. Results for the PLDDUE dataset.

Model	Dataset	Augmentation	Accuracy	F1-Score	Recall	Precision
DenseNet121	PLDDUE	augmented all	97.05%	96.64%	97.05%	97.05%
DenseNet121	PLDDUE	none	95.41%	94.34%	95.41%	95.41%
EfficientNetV2B0	PLDDUE	augmented all	99.34%	99.49%	99.34%	99.34%
EfficientNetV2B0	PLDDUE	none	97.70%	97.08%	97.70%	98.03%
MobileNetV3Large	PLDDUE	augmented all	98.36%	98.71%	98.36%	98.36%
MobileNetV3Large	PLDDUE	none	97.38%	97.42%	97.38%	97.38%
ResNet50V2	PLDDUE	augmented all	95.74%	94.86%	95.74%	96.05%
ResNet50V2	PLDDUE	none	93.77%	92.12%	93.11%	93.73%
VGG19	PLDDUE	augmented all	96.07%	95.60%	96.07%	96.07%
VGG19	PLDDUE	none	94.10%	92.95%	93.77%	94.39%

Table 5. Results for the PLD dataset.

Model	Dataset	Augmentation	Accuracy	F1-Score	Recall	Precision
DenseNet121	PLD	augmented all	99.75%	99.78%	99.75%	99.75%
DenseNet121	PLD	none	99.01%	98.93%	99.01%	99.01%
EfficientNetV2B0	PLD	augmented all	99.26%	99.23%	99.26%	99.26%
EfficientNetV2B0	PLD	none	99.26%	99.17%	99.26%	99.26%
MobileNetV3Large	PLD	augmented all	99.75%	99.73%	99.75%	99.75%
MobileNetV3Large	PLD	none	99.75%	99.73%	99.75%	99.75%
ResNet50V2	PLD	augmented all	99.26%	99.23%	99.26%	99.26%
ResNet50V2	PLD	none	98.52%	98.46%	98.52%	98.52%
VGG19	PLD	augmented all	99.75%	99.72%	99.75%	99.75%
VGG19	PLD	none	99.01%	99.06%	99.01%	99.01%

Table 6. Performance of Transformer-based Models on the PLDDUE Dataset with Different Augmentation Techniques.

Model	Dataset	Augmentation	Accuracy	F1-Score	Recall	Precision
ViT	PLDDUE	augmented all	99.50%	99.51%	99.50%	99.50%
ViT	PLDDUE	none	94.90%	94.85%	94.90%	94.89%
Transformer	PLDDUE	augmented all	99.40%	99.41%	99.40%	99.40%
Transformer	PLDDUE	none	94.80%	94.75%	94.80%	94.79%

Table 7. Performance of Transformer-based Models on the PLD Dataset with Different Augmentation Techniques.

Model	Dataset	Augmentation	Accuracy	F1-Score	Recall	Precision
ViT	PLD	augmented all	99.85%	99.80%	99.85%	99.86%
ViT	PLD	none	98.00%	97.98%	98.00%	98.02%
Transformer	PLD	augmented all	99.80%	99.75%	99.80%	99.81%
Transformer	PLD	none	97.98%	97.92%	97.98%	97.99%

Table 8. Average Test Metrics per Augmentation and Dataset for CNNs-based and Transformer-based Models.

Model	Augmentation	Dataset	Accuracy	F1-Score	Recall	Precision
CNNs-based	Augmentation all	PLD	99.55%	99.54%	99.55%	99.55%
	No Augmentation	PLD	99.11%	99.07%	99.11%	99.11%
	Augmentation all	PLDDUE	97.31%	97.06%	97.31%	97.37%
	No Augmentation	PLDDUE	95.67%	94.78%	95.47%	95.79%
Transformer-based	Augmentation all	PLD	99.01%	99.02%	99.01%	99.03%
	No Augmentation	PLD	98.52%	98.51%	98.50%	98.53%
	Augmentation all	PLDDUE	98.95%	98.96%	98.95%	98.97%
	No Augmentation	PLDDUE	98.02%	98.01%	98.01%	98.03%

combining multiple augmentation techniques. MobileNetV3Large and VGG19 also reach their best performance (99.75%) with this strategy. VGG19's accuracy drops to 99.01% without augmentation, demonstrating its reliance on augmentation. ResNet50V2 peaks at 99.26% with "augmented all" but falls to 98.52% without, showing its dependency on augmentation for optimal performance. Overall, the "augmented all" strategy is the most effective, enhancing accuracy and robustness, with models like DenseNet121 and EfficientNetV2B0 benefiting the most. Single or no augmentation results in noticeable performance drops.

The results in Tables 6 and 7 demonstrate the strong impact of augmentation techniques on transformer-based models. In Table 6, Augmentation all (transform, color, noise) boosts the ViT and Transformer models to 99.50% and 99.40% accuracy, respectively, compared to 94.90% and 94.80% without augmentation. Table 10 shows similar trends on the PLD dataset, with the ViT at 99.85%. Models without augmentation drop significantly, with accuracies of 98.00% for ViT and 97.98% for the Transformer. These results underscore the importance of data diversity and combined augmentation for improving accuracy and generalization in transformer-based models.

The analysis in Table 8, compares the performance of CNN-based and Transformer-based models across two datasets: PLD (lab) and PLDDUE (field). For CNN models, Augmentation all performed best on PLD and PLDDUE, achieving 99.55% and 97.31% respectively. Transformer models consistently performed better with Augmentation all achieving 99.01% on PLD and 98.95% on PLDDUE, outperforming CNNs on the field dataset. CNNs showed an advantage in the lab (99.55% vs. 99.01% for Transformers), while Transformers excelled in the field (98.95% vs. 97.31% for CNNs). These results suggest that CNNs are more suited to lab environments, while Transformer models are more robust in real-world conditions, particularly with Augmentation. This highlights the need for tailored models and augmentation strategies based on the deployment environment.

VI. CONCLUSION

In this study, we demonstrate that data augmentation techniques significantly improve the performance of both CNN and Transformer-based models for potato leaf disease classification. Augmentation strategies, including color, transformation, and noise-based methods, enhanced the robustness and accuracy of pre-trained models such as MobileNetV3, DenseNet121, EfficientNetV2B0, ResNet50V2, VGG19, transformer, and ViT. In controlled laboratory datasets (PLD), augmentation successfully captured disease-specific features, achieving high accuracy and precision. However, the true challenge arose in field datasets (PLDDUE), where environmental variability necessitated effective model generalization. Transformer-based models outperformed CNN-based models, particularly on the PLDDUE dataset. On the PLD dataset, Transformer models achieved an accuracy of 99.01% with combined augmentation, behind CNNs at 99.70%. In contrast, on the PLDDUE dataset, Transformer models reached an accuracy of 98.95%, surpassing CNNs, which peaked at 97.31%. Our findings highlight that models trained on datasets that were enhanced with advanced augmentation strategies, provide a more robust solution for potato leaf disease detection, particularly in real-world field conditions. The improved performance in the field dataset emphasizes the adaptability of models trained with augmentation enhanced data for agricultural applications. Future research could explore advanced techniques like GANs for synthetic data generation, domain-specific augmentations, and integrating additional data sources such as environmental conditions and soil health.

REFERENCES

[1] Strange, R. N., & Scott, P. R. (2005). "Plant disease: a threat to global food security," *Annu. Rev. Phytopathol.*, vol. *43, no.* 1, pp. 83-116.

[2] Li, L., Zhang, S., & Wang, B. (2021). "Plant disease detection and classification by deep learning—a review," *IEEE Access*, *9*, pp. 56683-56698.

[3] Alexey, D. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2010.11929.

[4] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. *147*, pp. 70-90.

[5] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. arXiv preprint arXiv:2204.08610.

[6] Eunice, J., Popescu, D. E., Chowdary, M. K., & Hemanth, J. (2022). "Deep learning-based leaf disease detection in crops using images for agricultural applications," *Agronomy*, *12*(10), 2395.

[7] Fulle, B. A., Ma, C., Shi, X., Zhu, W., Yemataw, Z., & Assefa, E. (2023, June). "Efficient Early Warning System in Identifying Enset Bacterial Wilt Disease using Transfer Learning," In *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8.

[8] Paiva-Peredo, E. (2022, December). "Deep learning for the classification of cassava leaf diseases in unbalanced field data set," In *International Conference on Advanced Network Technologies and Intelligent Computing*, pp. 101-114.

[9] Lanjewar, M. G., Morajkar, P., & P, P. (2024). "Modified transfer learning frameworks to identify potato leaf diseases," *Multimedia Tools and Applications*, *83*(17), 50401-50423.

[10] Ramya, R., & Kumar, P. (2023). "High-performance deep transfer learning model with batch normalization based on multiscale feature fusion for tomato plant disease identification and categorization," *Environmental Research Communications*, *5*(12), 125015.

[11] Salam, A., Naznine, M., Jahan, N., Nahid, E., Nahiduzzaman, M., & Chowdhury, M. E. (2024). "Mulberry Leaf Disease Detection Using CNN-based Smart Android Application," *IEEE Access*.

[12] Shah, S. K., Kumbhar, V., & Singh, T. P. (2024). "Grape (Vitis Vinifera) Leaf Disease Detection and Classification Using Deep Learning Techniques: A Study on Real-Time Grape Leaf Image Dataset in India," *International Journal of Engineering*, *37*(8), 1522-1533.

[13] Rashid, J., Khan, I., Ali, G., Almotiri, S. H., AlGhamdi, M. A., & Masood, K. (2021). "Multi-level deep learning model for potato leaf disease recognition," *Electronics*, *10*(17), 2064.

[14] Saeed, M. R., & Rashid, J. (2021). *Potato Disease Leaf Dataset (PLD)*. Kaggle.

[15] Shabrina, N. H., Indarti, S., Maharani, R., Kristiyanti, D. A., & Prastomo, N. (2024). "A novel dataset of potato leaf disease in uncontrolled environment," *Data in Brief*, *52*, 109955.

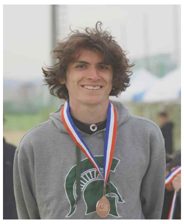
[16] Shabrina, N. H., Indarti, S., Maharani, R., Kristiyanti, D. A., Irmawati, I., Prastomo, N., & M, T. A. (2023). "*Potato leaf disease dataset in uncontrolled*

environment (Version 1) [Data set]," Mendeley Data.
[17] Richter, D. J., & Kim, K. (2024). Assessing the Performance of Domain-Specific Models for Plant Leaf Disease Classification: A Comprehensive Benchmark of Transfer-Learning on Open Datasets.
[18] CNU DNS Lab. (2025). *PotatoAugmentation* [Computer software]. GitHub. https://github.com/CNU-DNS-Lab/PotatoAugmentation (accessed Jan., 10, 2025).
[19] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). "Searching for mobilenetv3," In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314-1324.
[20] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep residual learning for image recognition," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778.
[21] Simonyan, K., & Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556.
[22] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). "Densely connected convolutional networks," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708.
[23] Tan, M., & Le, Q. (2021, July). "Efficientnetv2: Smaller models and faster training," In *International conference on machine learning*, pp. 10096-10106.
[24] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Identity mappings in deep residual networks," In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pp. 630-645.
[25] Vaswani, A. (2017). "Attention is all you need," *Advances in Neural Information Processing Systems*.
[26] Dosovitskiy, A. (2020). "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929.

Authors

MD Ilias Bappi

He received his B.S. degree in Computer Science and Engineering from Daffodil International University, in 2017. He worked as a Software Engineer at Samsung Electronics, for five years until 2023. Currently, he is pursuing a master' s degree in AI Convergence at Chonnam National University. His research interests include machine learning, deep learning, continual learning in agriculture and healthcare.



David J. Richter

He received the B. S. degree in computer science from Kempten University of Applied Sciences, in 201 9, and the M. S. degree in computer information technology from Purdue University Northwest, in 2021 . He is currently pursuing a master' s degree in AI convergence with Chonnam National University. His research interests include AI, machine and deep learning, reinforcement learning, genetic programming.

Kyungbaek Kim

He received the B. S., M. S., and Ph. D. degrees in electrical engineering and computer science from Korea Advanced Institute of Science and Technology (KAIST), in 1999, 2001, and 2007, respectively. He is currently a Professor with the Department of AI Convergence, Chonnam National University. Previously, he was a Postdoc toral Researcher with the Department of Computer Sciences, University of California, Irvine. His research interests include intelligent distributed systems, SDN/SDI, big data platform, GRID/cloud systems, social networking systems, AI applied cyber-physical systems, blockchain, and other issues of distribute