Sciences Methods and Technologies International Journal (SciMeTech)

ISSN: 3085-5284, journal homepage: www.scimetech.com

Lightweight Damascena Rose Maturity Classification via Wavelet Scattering

Mohamed Ohamouddou[1], Rafik Lasri[1], Suhail odeh[2]

[1]*TED:AEEP Research Group, FPL, Abdelmalek Essaâdi University, Tetouan, Morocco*

[2]*Software Engineering Department, ASTE Faculty, Bethlehem University, Bethlehem P125, Palestine*

How to cite:

Mohamed Ohamouddou, Rafik Lasri, Suhail odeh, "Lightweight Damascena Rose Maturity Classification via Wavelet Scattering", Sciences Methods and Technologies International Journal (SciMeTech), Vol 2, Issue 2, p 197-202

Keywords:

Rosa damascena
Rose harvesting
Wavelet scattering
Precision agriculture
Lightweight neural network

Abstract

Visual assessment of rose maturity is a routine but labour intensive task in cut rose production, where harvest timing directly affects flower quality and shelf life. This paper proposes a lightweight image classifier for binary rose maturity recognition (immature versus mature) that combines a frozen two dimensional Morlet wavelet scattering transform with a small convolutional head. Because the scattering transform uses fixed wavelet filters, only 61,410 head parameters are trained from data. The model is evaluated on a rose dataset of 623 test images and compared, under an identical training protocol, with six ImageNet-pretrained baselines: SqueezeNet 1.1, MobileNetV3-Small, ShuffleNetV2-x0.5, MobileNetV2, EfficientNet-B0, and ResNet50. The proposed model reaches 96.0% test accuracy and an F1-score of 0.96, with 5.6 to 383 times fewer trainable parameters than the baselines. The accuracy gap to the best baseline is 2.2 percentage points, exchanged for a substantial reduction in trainable capacity. These results indicate that a fixed wavelet scattering front-end is a practical alternative to deep learned backbones for on-device deployment in field conditions.

I.Introduction

Rosa damascena, commonly known as the Damascus rose, is one of the most widely cultivated rose species for natural essential oil and rose water. In Morocco, its cultivation is concentrated in the Dades Valley around Kelaa M'Gouna, where the crop has long been a defining part of the local economy and where an annual Rose Festival is held at the end of the harvest. The harvest is short and time-sensitive: petals must be picked at a specific stage of bud opening and before the morning sun warms the field, because the volatile aromatic compounds evaporate quickly as the temperature rises. Sorting is done by hand, and pickers judge each bud visually under variable light. An automated visual classifier on a hand-held device could improve consistency and reduce labour cost, but only if the underlying model is small enough for the hardware available in the field.

Convolutional neural networks (CNNs) are now a standard tool in agricultural image analysis [20]. Mohanty et al. [13] fine-tuned AlexNet [15] and GoogLeNet [16] on the PlantVillage dataset to identify 26 diseases across 14 plant species, and the surveys of Kamilaris and Prenafeta-Boldu [14] and Liakos et al. [19] document many further applications in crop monitoring, weed detection, fruit recognition, and yield estimation. Most systems start from a large ImageNet [18] pretrained backbone such as VGG [17] or ResNet [12] and fine-tune it on the target dataset. The recipe transfers to rose maturity, but the resulting models are heavy: a fine-tuned ResNet50 has 23.5 million trainable parameters and produces a 90 MB checkpoint at full precision, well beyond the budget of low-power devices.

To address this, several families of compact CNNs have been developed. SqueezeNet [6] reaches AlexNet-level accuracy with about fifty times fewer parameters using fire modules. The MobileNet family [7, 8, 9] uses depthwise separable convolutions, with inverted residuals in MobileNetV2 [8] and a hardware-aware architecture search in MobileNetV3 [9]. ShuffleNetV2 [10] introduces channel shuffling and design rules tuned for measured wall-clock speed. EfficientNet [11] proposes a compound scaling rule for depth, width, and input resolution. All of these architectures still learn the entire feature stack end-to-end and assume access to a large pretraining corpus.

A complementary line of work asks whether the early visual features need to be learned at all. The wavelet scattering transform [1, 2] cascades wavelet convolutions with a modulus non-linearity to produce a representation that is locally translation-invariant and Lipschitz stable to small deformations. Because the wavelets are fixed Morlet filters, the transform contributes zero trainable parameters. Scattering features have been combined with shallow classifiers and small CNN heads on standard image benchmarks [4, 5], where they perform competitively in the small-data regime. The Kymatio library [3] provides an efficient PyTorch implementation, which is the one used in this work.

In agricultural settings, where datasets are typically small and on-device deployment is required, the inductive bias of fixed wavelet features can compensate for limited supervision while keeping the gradient and parameter footprint small. To the best of our knowledge, this design has not yet been evaluated on a floriculture task. The contributions of this paper are: (i) a binary rose maturity classifier that uses a 2D Morlet wavelet scattering transform ($J = 3$, $L = 8$) as a frozen front-end followed by a 61,410-parameter convolutional head; (ii) a benchmark against six ImageNet-pretrained baselines (SqueezeNet 1.1, MobileNetV3-Small, ShuffleNetV2-x0.5, MobileNetV2, EfficientNet-B0, and ResNet50) trained under an identical protocol; and (iii) a trade-off analysis showing 96.0% accuracy and an F1-score of 0.96 with 5.6 to 383 times fewer trainable parameters than the baselines, at the cost of a 2.2 percentage-point accuracy gap.

II. Methods

The dataset consists of 3,114 RGB photographs of Rosa damascena buds collected in the field (In Kelaa M'gouna, Morocco) and organised into two classes: immature (closed buds, prior to peak essential-oil concentration) and mature (buds at the picking stage). The data were partitioned image-wise into a training set of 2,491 images and a test set of 623 images (275 immature and 348 mature) with no overlap to prevent data leakage. All images are resized to 224×224 pixels and normalised with the standard ImageNet statistics. Training images are augmented with horizontal flip, rotation up to 15°, colour jitter (brightness 0.2, contrast 0.2, saturation 0.2, hue 0.1), and random affine translation (0.1) and scale (0.9-1.1). The test split is evaluated without augmentation (Figure 1).

The two-dimensional wavelet scattering transform [1, 2] computes, at each spatial location, an iterated cascade of complex-valued wavelet convolutions followed by a modulus non-linearity and a low-pass averaging. With Morlet wavelets at J scales and L orientations, the transform retained up to second order produces $1 + LJ + L^2J(J-1)/2$ coefficients per channel. The transform is implemented with the Kymatio library [3] and contributes zero trainable parameters.

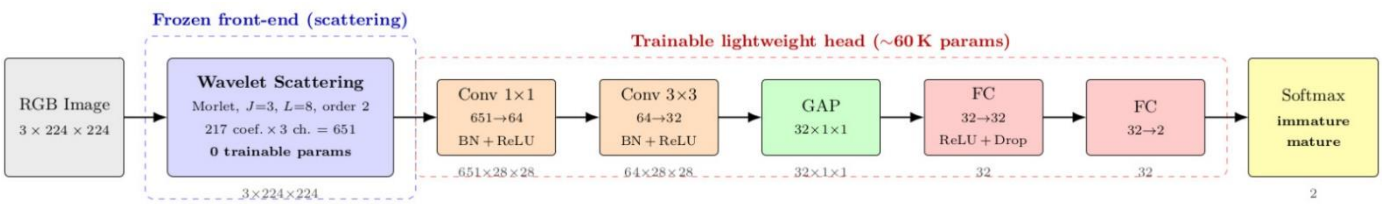

Figure 1: Model architecture

The architecture (Figure 1) places a small CNN head on top of the frozen scattering front-end. The head consists of a 1×1 convolution from 651 to 64 channels with batch normalisation and ReLU; a 3×3 convolution from 64 to 32 channels with batch normalisation and ReLU; global average pooling to a 32-dimensional vector; a fully-connected layer 32 → 32 with ReLU and dropout ($p = 0.3$); and a final fully-connected classifier 32 → 2. The head has 61,410 trainable parameters.

We compare the proposed model with six ImageNet-pretrained torchvision backbones in increasing order of trainable parameter count: ShuffleNetV2-x0.5 [10] (343,842 parameters), SqueezeNet 1.1 [6] (723,522), MobileNetV3-Small [9] (1,519,906), MobileNetV2 [8] (2,226,434), EfficientNet-B0 [11] (4,010,110), and ResNet50 [12] (23,512,130). For each baseline, the final classifier is replaced with a two-class linear layer and all parameters are fine-tuned.

All models are trained for 60 epochs with the Adam optimiser (learning rate 10^{-3}, weight decay 10^{-4}), batch size 32, cross-entropy loss, and a cosine-annealing learning-rate schedule. The random seed is fixed to 42. We track loss, accuracy, precision, recall, and F1-score on both splits at every epoch, and retain the checkpoint at the epoch of best test F1.

III. Results and discussion

Table 1 reports the test-set metrics for all seven models. The proposed model reaches 96.0% accuracy and an F1-score of 0.96 with 61,410 trainable parameters. Three baselines tie at the highest accuracy (98.4%): ShuffleNetV2-x0.5, MobileNetV2, and ResNet50. The proposed model is 5.6 times smaller than the smallest baseline and 383 times smaller than the largest, and the accuracy gap to the best baseline is 2.2 percentage points (ΔF1 = 0.023). The full FP32 checkpoint of the proposed model occupies 8.71 MB, but the trainable head accounts for only 240 KB; the size is dominated by the pre-computed Morlet wavelet buffers, which can be re-generated at load time from the configuration (J, L, image size)

.

Table 1. Test-set comparison of the proposed model with six ImageNet-pretrained baselines on the Damascena rose maturity dataset.

Model	Accuracy	Precision	Recall	F1-score	Params (M)	Size (KB)
Proposed (ours)	0.96	0.95	0.96	0.96	0.06	8712.0
SqueezeNet 1.1	0.97	0.97	0.98	0.97	0.72	2826.3
MobileNetV3-Small	0.98	0.97	0.98	0.98	1.52	5984.7
ShuffleNetV2-x0.5	0.98	0.98	0.98	0.98	0.34	1374.6
MobileNetV2	0.98	0.98	0.98	0.98	2.23	8830.7
EfficientNet-B0	0.98	0.97	0.98	0.98	4.01	15829.0
ResNet50	0.98	0.98	0.98	0.98	23.51	92052.2

Figure 2 shows the training and test of the proposed model. Both decrease rapidly during the first ten epochs and stabilise around epoch 25. The test loss exhibits occasional spikes (the largest at epoch 29, with test loss 0.69) that recover within one epoch. The lowest test loss is reached at epoch 59 (0.103), and the best test F1 is reached at epoch 45 (0.96. The test metrics sit above the train metrics for most epochs because the train metrics are computed on heavily augmented inputs while the test split is unaugmented; this is the expected pattern under strong online augmentation rather than evidence of label leakage.

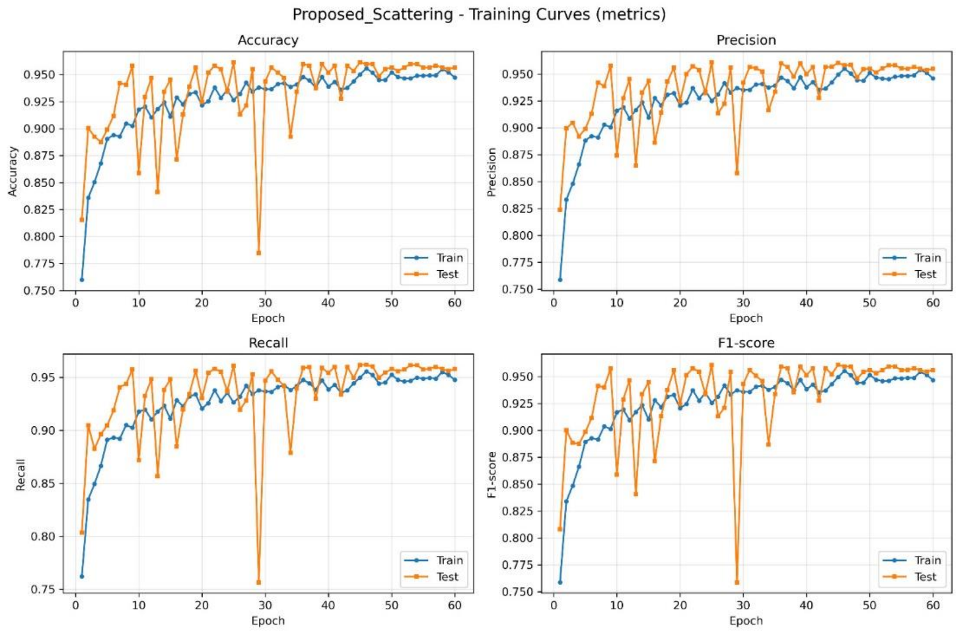

Figure 2 : Training and test cross-entropy loss of the proposed model across 60 epochs

Figure 3 compares the test-set trajectories of all seven models. The proposed model converges more slowly than the pretrained baselines, which start near 0.93 F1 from the first epoch thanks to ImageNet initialisation. The asymptotic gap, however, is small: by epoch 60, the proposed model's F1 is within 0.025 of the best baseline.

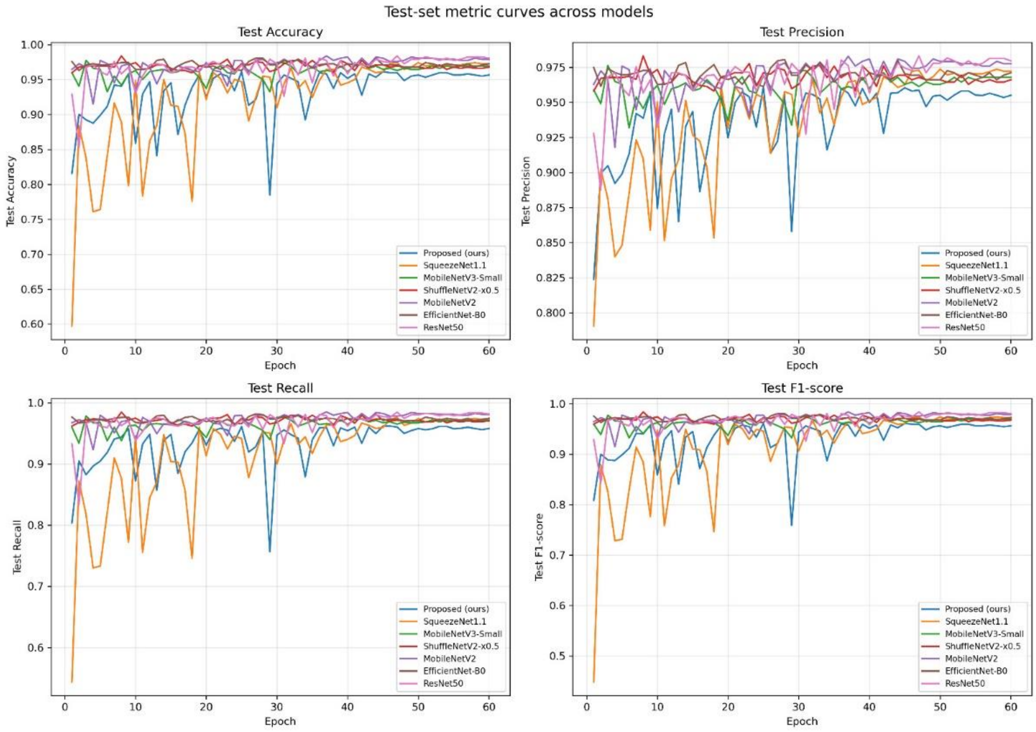

Figure 3 : Test-set metric trajectories of the proposed model and the six ImageNet-pretrained baselines on the Damascena rose maturity dataset.

Figure 4 shows the confusion matrix of the proposed model on the test set. Of 275 immature buds, 265 are correctly classified and 10 are confused with mature (96.4% per-class accuracy); of 348 mature buds, 334 are correctly classified and 14 are confused with immature (96.0%). The two error directions are roughly balanced.

Figure 5 presents eight test-set predictions: six correct and two misclassifications. Correct predictions span a range of confidences from 87% to 99.7%. The misclassified examples are cases in which the bud is small in the frame, partially occluded by leaves, or surrounded by similarly textured foliage, where the local scattering statistics resemble those of the opposite class.

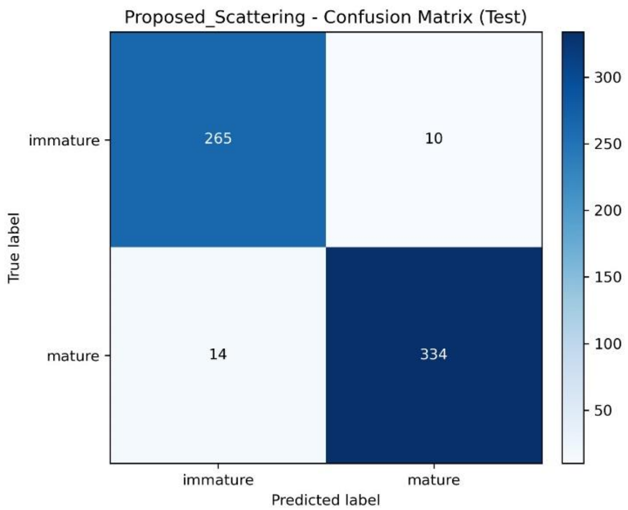

Figure 4 : Confusion matrix of the proposed model on the test set.

The results support a clear trade-off. Pretrained CNN backbones with millions of parameters reach 98% accuracy with little tuning, but their parameter count is difficult to justify for a binary classification problem on a small, single-domain dataset. A 61 K-parameter head over a frozen wavelet scattering front-end loses 2.2 percentage points in exchange for at least an order-of-magnitude reduction in trainable capacity and a deployable artefact whose head weights occupy 240 KB at FP32 (approximately 60 KB at INT8). For an in-field tool used during the Kelaa M'Gouna harvest, the second profile is the more practical: training on a single farmer's annotations is feasible, the model fits comfortably on a Raspberry Pi or smartphone, and the inference path requires no ImageNet pretraining.

Several limitations should be noted. The evaluation is on a single dataset and a single random seed, and the comparison reflects one training regime rather than per-baseline hyper-parameter tuning. The two error directions are balanced, but their economic cost may differ in the field: an immature bud sent to distillation contributes little oil, while a mature bud left on the plant is lost outright. A cost-sensitive decision threshold should be chosen with input from local growers. Robust deployment also requires evaluation under varying illumination, motion blur, and a broader range of camera angles than the present dataset covers.

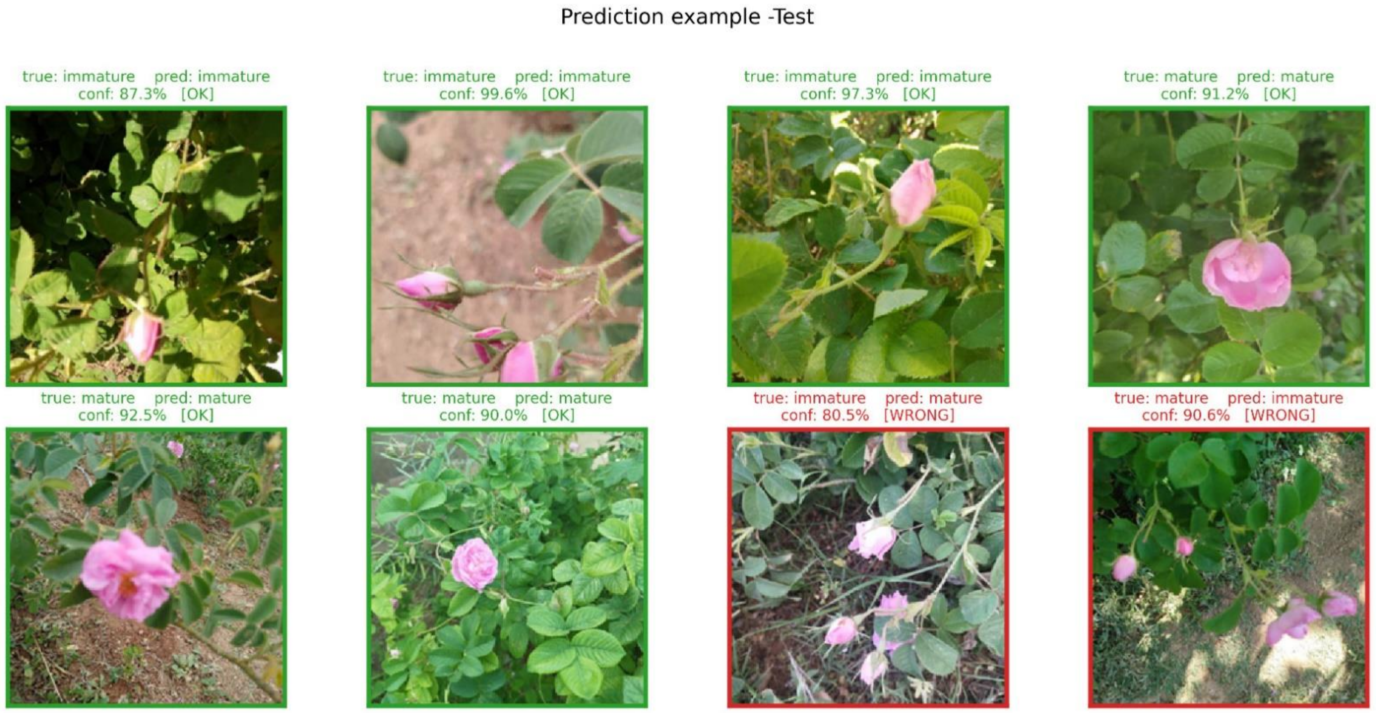

Figure 5 : prediction examples from the test set.

IV.Conclusion

This paper presented a parameter-efficient image classifier for binary *Rosa damascena* maturity recognition that combines a frozen two-dimensional Morlet wavelet scattering front-end with a small CNN head of 61,410 trainable parameters. On a Damascena rose dataset of 623 test images, and against six ImageNet-pretrained baselines trained under the same protocol, the proposed model reaches 96.0% accuracy and an F1-score of 0.96 with 5.6 to 383 times fewer trainable parameters than the baselines. The 2.2 percentage-point accuracy gap to the best baseline is exchanged for a substantially smaller deployable footprint, which makes the model a practical option for on-device use during the Kelaa M'Gouna harvest. Future work will extend the evaluation to a larger, multi-season dataset, study INT8 quantisation and on-device latency on a Raspberry Pi, and develop a cost-sensitive decision threshold with input from local growers.

References

[1] Mallat, S. (2012). Group invariant scattering. Communications on Pure and Applied Mathematics, 65(10), 1331-1398.

[2] Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1872-1886.

[3] Andreux, M., Angles, T., Exarchakis, G., Leonarduzzi, R., Rochette, G., Thiry, L., Zarka, J., Mallat, S., Anden, J., Belilovsky, E., Bruna, J., Lostanlen, V., Chaudhary, M., Hirn, M. J., Oyallon, E., Zhang, S., Cella, C., & Eickenberg, M. (2020). Kymatio: Scattering transforms in Python. Journal of Machine Learning Research, 21(60), 1-6.

[4] Oyallon, E., & Mallat, S. (2015). Deep roto-translation scattering for object classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2865-2873.

[5] Oyallon, E., Belilovsky, E., Zagoruyko, S., & Valko, M. (2019). Scattering networks for hybrid representation learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(9), 2208-2221.

[6] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint arXiv:1602.07360.

[7] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4510-4520.
[9] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 1314-1324.
[10] Ma, N., Zhang, X., Zheng, H.-T., & Sun, J. (2018). ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), 116-131.
[11] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Vol. 97, 6105-6114.
[12] 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 (CVPR), 770-778.
[13] Mohanty, S. P., Hughes, D. P., & Salathe, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.
[14] Kamilaris, A., & Prenafeta-Boldu, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70-90.
[15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (NeurIPS), 25, 1097-1105.
[16] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-9.
[17] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (ICLR).
[18] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 248-255.
[19] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.
[20] Ohamouddou, M., Ohamouddou, S., El Afia, A., & Lasri, R. (2025). ATMS-KD: Adaptive temperature and mixed sample knowledge distillation for a lightweight residual CNN in agricultural embedded systems. Smart Agricultural Technology, 12, 101617.