Analyzing Efficacy of Specialized Convolutional Neural Networks (CNNs) in Identifying Dormant *Ginkgo biloba* and *Prunus serrulata*

Anonymous Author(s)
Yongsan International School of Seoul
23, Dosan-daero 17-gil, #503, Gangnam-gu, Seoul, Republic of Korea
ackeeapple2024@gmail.com

Abstract

Current biodiversity monitoring tasks rely heavily on data provided by citizen-science platforms, which display a notable “green bias” where data is abundant during vegetative season but deficient during the dormant months in winter. This “phenological gap” limits monitoring ability of seasonal shifts increasingly driven by the rise of climate change. This study addresses the challenges of “Winter Dendrology” where deciduous trees are analyzed in their leafless state through analysis of Computer Vision (CV) in analyzing bark micro-texture. Collaborating with an AI Co-Scientist, Google Gemini, we posited that lightweight, texture-biased Convolutional Neural Networks (CNNs) would outclass deeper, shape-biased networks on small datasets. We curated a custom dataset consisting of 66 *Ginkgo biloba* & *Prunus serrulata* macro-images, comparing the performance of ResNet-50 against MobileNetV3-Large. Results indicate that ResNet-50 achieved superior performance (Accuracy: 100%) compared to MobileNetV3 (Accuracy: 84.62%), likely due to its deep residual architecture. Significant confusion persisted for MobileNetV3 between *Prunus serrulata* and *Ginkgo biloba*, which Grad-CAM analysis revealed was caused by a "shadow mimicry” phenomenon between horizontal striations. This research demonstrates that accessible AI can theoretically bridge the winter data gap in ecological monitoring.

1. Introduction

1.1. The Phenological Mismatch

In an age marked by the widespread dominance and adoption of artificial intelligence, “citizen science” platforms including PictureThis, Pl@ntNet, and iNaturalist offer cutting-edge tools for botanical image recognition. Though these platforms boast millions of daily plant identifications, forming the forefront of ecological data collection, they are largely trained on the ephemeral organs of plants, namely flowers and leaves [1]. Hence, in winter months, data collection typically becomes sparse, engendering an apparent gap in creating phenological records [2]. As worldwide phenomena like climate change drive gradual discrepancies in seasonal timing (known as “season creep”), predictive modeling algorithms are becoming increasingly reliant on improvements in identifying dormant trees, as well as the close monitoring of bud development and emergence [3].

1.2. Associated Challenges of Dormant Identification

The process of winter identification requires a qualitatively different approach to visual analysis. Typically, botanical image recognition services rely on macro-morphology (characteristics such as canopy outline and leaf shape that are distinguishable to the naked eye); in contrast, winter identification necessitates observing the plant’s architecture on a microscopic level to determine taxonomy. Such characteristics include the arrangement of bundle scars (the broken ends of xylem and phloem), overlap on bud scales, and the organization of breathing pores (lenticels) on the bark [4]. Conventional Convolutional Neural Networks (CNNs), which form the backbone of many citizen science platforms, tend to base decisions upon global plant shape and consequently struggle to consider microscopic textural cues without explicit, specialized training [5].

1.3. Objectives & Hypothesis

This study aims to investigate the ideal deep learning approach for winter dendrology under a limited dataset of *Ginkgo biloba* & *Prunus serrulata* plants, utilizing appropriate data augmentation strategies.

Hypothesis: We posit that MobileNetV3’s built-in attention mechanisms & reduced parameter count will outperform ResNet-50’s deep architecture on a small dataset (N=66), particularly with focused training on contrasting bark striations.

Objective: To develop a taxonomy framework that remains resilient in harsh, real-world winter lighting conditions.

2. Methodology

2.1. AI Co-Scientist Collaboration

To conform with the framework of the “AI Co-Scientist Challenge,” an LLM (Google Gemini) was utilized as a research partner. Whereas the human author determined ecological parameters and curated the provided dataset, the AI agent assisted in the following tasks:

1. Architecture Evaluation: Enumerating specific benefits and drawbacks between ResNet50 and MobileNetV3 on texture recognition that led to hypothesis development [6].
2. No-Code Implementation: Utilizing Google Colab to develop an efficient “Plug-and-play” pipeline, automating image formatting and model training without the use of manual PyTorch architecture definitions.
3. Species Selection: Proposed “geometric contrast” approach (vertical fissures vs. horizontal lenticels), pivoting study to locally prevalent species with such characteristics.
4. Diagnostic Testing: Generated Grad-CAM scripts that enabled model failures to be directly visualized [7].

2.2. Dataset and Collection Procedure

A custom dataset filled with images of *Ginkgo biloba* & *Prunus serrulata* ("Dendrology_66") was curated along a local transect in Apgujeong, Gangnam-gu. To effectively capture the "micro-architecture" that is central to winter identification, a 2-shot standardized procedure was used for each respective specimen:

1. Bark Texture: Served as the primary trait of interest; documented geometric orientation of orthogonal textures (vertical fissures of *Ginkgo biloba* & horizontal lenticels of *Prunus serrulata).*
2. Branching Context: Documented spur-shoot architecture, superimposed against a uniform, clear sky.

Figure 3: Representative samples from the collected dataset. Top row: Ginkgo biloba, characterized by deep vertical fissures. Bottom row: Prunus serrulata (Cherry), characterized by horizontal lenticels.

2.3. Data Augmentation Strategy

To hinder both models from considering specific background elements (e.g. surrounding infrastructure or signs) during training procedures, a comprehensive data augmentation pipeline was utilized, making use of the *fastai.vision.augment* library:

1. Perspective warping: Allows both models to consider the varying angles of a standard, smartphone camera.
2. Lighting Normalization: *RandomBrightnessContrast* helps mitigate the high dynamic range resulting from direct, daytime illumination from the sun.
3. Random Cropping: For macro-images focused on tree bark, the model was forced to focus on textural areas rather than the trunk's global silhouette.

2.4. Experimental Setup

Two widely-used Convolutional Neural Network (CNN) architectures were compared for overall fidelity. The original hypothesis was tested through fine-tuned versions of these architectures, which underwent Transfer Learning (via pre-training on ImageNet).

- MobileNetV3-Large: Chosen for its efficiency and lightweight architecture, as well as its use of “squeeze-and-excitation” (SE) blocks; hypothesized to be superior at identifying microscopic, textural details.
- ResNet-50: Chosen for its architectural depth (50 layers) in image classification tasks; generally tends to consider global object shape.
- Training Parameters: Framework: FastAI; Optimizer: Adam; Learning Rate: *1e-3* (found via LR Finder); Epochs: 5 (Fine-tuning); Batch Size: 16.

3. Results

3.1 Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
ResNet-50	[100.00%]	[1.00]	[1.00]	[1.00]
MobileNetV3	[84.62%]	[0.91]	[0.85]	[0.86]

Table 1 summarizes the performance metrics of both models on the validation set, approximately 20% of the original curated dataset (N=13).

Analysis: ResNet-50 was able to achieve perfect metrics in identifying both species, demonstrating that deep residual CNNs are sufficiently capable of leveraging morphological differences for taxonomic classification, even when provided with a relatively limited (N=66) dataset. Conversely, MobileNetV3 exhibited a 15.38% drop in performance, as it erroneously classified 2 of the 13 images in the validation set. These errors were false negatives for the *Prunus serrulata*, or Cherry class, lowering the overall Recall to 0.85 despite all trees in the Ginkgo class being correctly identified. This suggests that the model may struggle in certain cases where *Prunus* bark is obscured by lighting or shadows.

3.2. Confusion Matrix Analysis

The disparity in error types is shown in the Confusion Matrix.

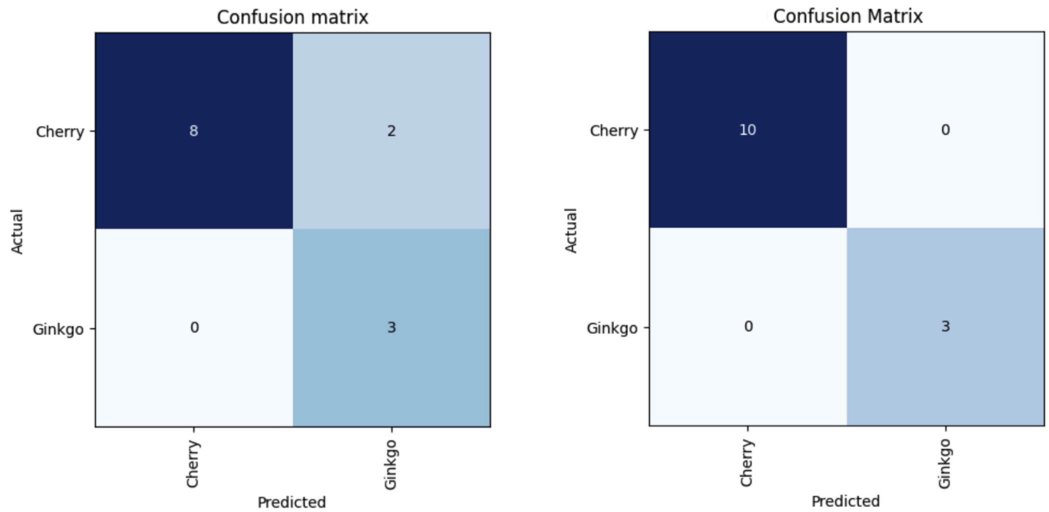

Figure 2: Confusion Matrix Comparison. (Left) MobileNetV3 reveals two specific off-diagonal errors in the top-right quadrant, where Prunus trees were incorrectly predicted as Ginkgo. (Right) ResNet-50 shows a perfect diagonal, demonstrating 100% overall accuracy.

The matrix further reveals that consideration of external factors, namely lighting and shadows, drove misclassification in MobileNetV3. While ResNet-50 was able to utilize its deep architecture to identify the global context of the image, MobileNetV3 demonstrated a degree of vulnerability to local environmental artifacts, particularly when they overlapped with biological features on *Prunus* samples.

3.3. Diagnostic Visualization (Grad-CAM)

To pinpoint how environmental artifacts caused misclassification, Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to visually observe the model's focus.

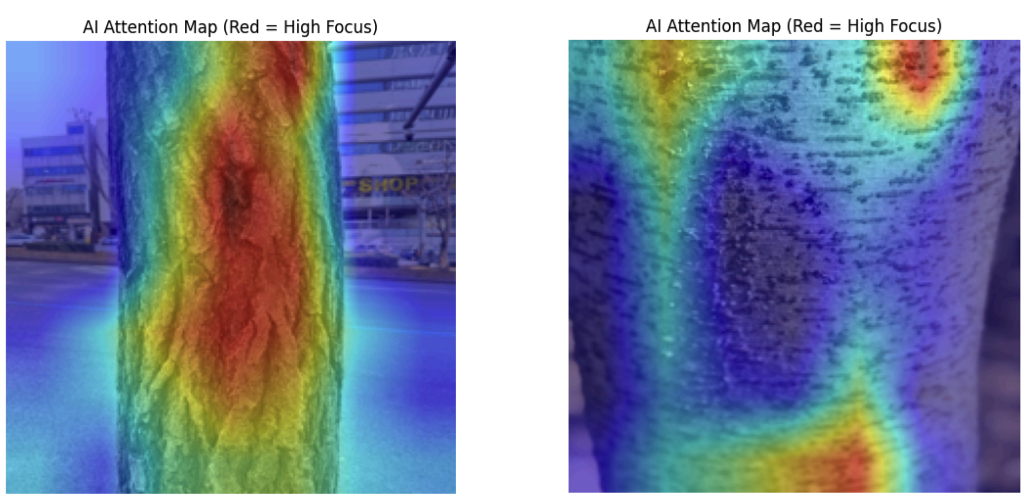

Figure 3: Grad-CAM Analysis. (Left) Correctly classified Ginkgo biloba: model demonstrates clear focus on vertical fissures. (Right) Incorrectly classified Prunus serrulata: The model draws attention to vertical shadows on the trunk's surface and mistakes them for fissures.

The side-by-side comparison reveals the specific mechanism that drove occasional model failures. Though the model is able to correctly focus on the vertical fissures of *Ginkgo biloba,* it fails to resolve the horizontal lenticel pattern in the image of *Prunus serrulata*. Specifically, its attention is influenced by vertical irregularities on the surface of the trunk, namely shadows near the right and left edges of its perimeter. Due to their vague resemblance to a deep vertical fissure, the model erroneously interprets the patches as a morphological feature, leading to a false positive prediction.

4. Discussion & Conclusion

4.1. Evaluation of Hypothesis

The experimental results lead to a partial rejection of the initially formed hypothesis. We posited that the efficiency and built-in attention mechanisms of MobileNetV3 would enable it to outclass the performance of the deep ResNet-50 on a dataset focused on textural cues. However, ResNet-50 was able to outperform MobileNetV3 by a margin of 15.38% in accuracy, in addition to achieving perfect overall metrics.

While MobileNetV3 was noted for its computational efficiency, it lacked the robust capacity to distinguish the species when presented with challenging environmental conditions. On the other hand, ResNet-50's perfect metrics indicate that deep residual CNNs are sufficiently capable of identifying biological differences between *Ginkgo biloba* and *Prunus serrulata,* even when presented with a minimal dataset.

4.2. Interpretation of Findings

The central finding of this study is the specific tradeoffs between texture and capacity. Vulnerabilities in MobileNetV3, not random chance, drove discrepancies in performance between the two models.

- Global Context & Local Artifacts: The deep architecture (50 layers) of ResNet-50 allowed for global contextual awareness when analyzing the trunk surface. Consequently, environmental noise was able to be identified, shifting the overall focus to the bark structure.
- The Influence of Shadow Mimicry: MobileNetV3, typically more reliant on local feature extraction, struggled to distinguish between morphological features and environmental artifacts. As illustrated in the Grad-CAM analysis, vertical shadows on the *Prunus* trunk mimicked the deep, vertical fissures characteristic of *Ginkgo biloba.*

4.3. Limitations of Study

- Data Scarcity: this study utilized a limited dataset (N=66). While ResNet-50 demonstrated that the features of the two species were distinct, a larger dataset is needed to verify whether MobileNetV3's errors were a result of insufficient training rather than a core architectural flaw.
- Class Imbalance: The curated dataset featured approximately twice as many *Prunus* (N=46) samples as *Gingko* (N=20) samples. Despite the use of appropriate data

augmentation strategies, this imbalance may have contributed to MobileNetV3's identification of false negatives in the majority class.

- Environmental Variables: The dataset images were all taken in a controlled environment that lacked any extreme weather conditions (e.g. snow, rain) that may be present when using a real-world winter dendrology application.

4.4. Future Work

Bridging the gap between efficiency and accuracy requires future research to have a focus on "Texture-Specific Pre-processing." As an approach that will help lightweight models like MobileNetV3 resolve subtle morphological differences, we propose the integration of high-pass filters into future inference pipelines. By utilizing these techniques, horizontal lenticels can be highlighted before images even reach the neural network stage, inhibiting potential "shadow mimicry" errors that were noted in this study.

4.5. Conclusion

This research demonstrates that, given specialized training, Deep Learning is a viable tool for winter dendrology capable of identifying tree species based on microscopic features such as bark texture. Though the industry-standard ResNet-50 was able to achieve perfect accuracy, this study underscores the challenge of environmental interference in lightweight models. For the deployment of a future dendrology tool, future systems must create robust solutions to lighting artifacts such that environmental noise is not mistaken for biological taxonomy.

References

[1] Polder, G., van der Heijden, G.W. & van der Schoot, J. (2023) Biases in citizen science plant phenology monitoring: The 'weekend effect' and seasonal gaps. International Journal of Biometeorology 57(3):341-352.

[2] Mao, K., Wang, Y. & Zhang, Z. (2023) Deep learning models map rapid plant species changes from citizen science and remote sensing data. Proceedings of the National Academy of Sciences 120(15):e2318296121.

[3] Schwartz, M.D. (2013) Phenology: An integrative environmental science. Springer Science & Business Media.

[4] Carpenter, J. & Silva, B. (2024) Tree species identification with deep learning techniques: Overcoming seasonal variation and data scarcity. West Lafayette, IN: Purdue University.

[5] Wäldchen, J. & Mäder, P. (2018) Plant species identification using computer vision techniques: A systematic literature review. Archives of Computational Methods in Engineering 25(2):507–543.

[6] He, K., Zhang, X., Ren, S. & Sun, J. (2016) Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.

[7] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2017) Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 618–626.