LIGHTWEIGHT TRANSFER LEARNING MODELS FOR KERANGAS LANDSCAPE CLASSIFICATION ON EDGE DEVICES

Dwi Ahmad Dzulhijjah[1,2,6], Kusrini Kusrini[2], Afdhal Afdhal[3], Noor Akhmad Setiawan[4], Adi Wibowo[5]

[1]Program Profesi Insinyur, Institut Teknologi Indonesia, Tangerang, Indonesia
[2]Computer Science Faculty, Universitas AMIKOM Yogyakarta, Indonesia
[3]Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia
[4]Department of Electrical and Information Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia
[5]Department of Informatics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia
[6]Department of Informatics and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Indonesia
Emails: dwiahmaddzulhijjah@iti.ac.id, kusrini@amikom.ac.id, afdhal@usk.ac.id, noorwewe@ugm.ac.id, bowo.adi@live.undip.ac.id

Abstract:

Kerangas forests (tropical heath forests) are nutrient-poor, fire-prone ecosystems threatened by deforestation and land-use change. Monitoring these post-disturbance landscapes is crucial for ecological restoration but remains challenging due to field inaccessibility and the high resource demands of conventional deep learning. This study evaluates lightweight CNNs (MobileNetV1/V2) for classifying Kerangas imagery into three ecological succession stages. Using transfer learning and domain-specific data, twelve RGB and grayscale models were assessed by accuracy, latency, RAM, and flash usage for edge deployment. MobileNetV2 with 160×160 RGB and $\alpha = 0.75$ reached 95.8% accuracy, matched by a 96×96 grayscale model with $\alpha = 0.35$—which reduced memory by over 97% and achieved sub-second inference time.

Keywords:

Convolutional neural networks, ecological monitoring, edge computing, Kerangas forest, transfer learning.

1 Introduction

Kerangas forests are ecologically distinct, with acidic, nutrient-poor soils and sparse vegetation [1], increasingly at risk from deforestation, fires, and land-use change in Borneo and Southeast Asia [2]. Accurate post-disturbance monitoring is vital for restoration.

CNNs have advanced visual classification [3], but their high resource needs [4] hinder use in constrained settings. Satellite data often lack sufficient resolution, and UAVs face operational limits. Edge devices offer a practical, scalable alternative, especially in degraded, hard-to-access Kerangas zones. These devices can support low-power CNN models for continuous monitoring.

We investigate lightweight transfer learning models for Kerangas classification, evaluating MobileNetV1/V2 variants with different width multipliers on domain-specific images. Our contributions include: (1) lightweight CNN development, (2) comparative evaluation of performance and efficiency, (3) a reproducible classification pipeline, and (4) deployment insights for low-resource environments.

2 Related Work

Recent studies have demonstrated the effectiveness of lightweight convolutional neural networks (CNNs) combined with transfer learning (TL) for ecological and agricultural image classification under resource constraints. In agriculture, TL-based models such as SqueezeNet, InceptionV3, and VGG variants achieved 96.4% accuracy in rice leaf disease detection [5], while EfficientNet-Lite4 reached over 90% accuracy in classifying endemic plant species in Colombia's Santurbán paramo [6]. In forest ecosystems, Sentinel-2 imagery has been used with DenseNet, MobileNetV3, and ShuffleNet for tree species classification [7], and SqueezeNet demonstrated high interpretability in noctilucent cloud detection [8].

Beyond agriculture and forestry, lightweight CNNs have proven effective in edge-focused domains such as biomedical

imaging and embedded systems. MobileNet variants have been used for breast cancer histopathology [9] and COVID-19 detection [10]. These models have also been deployed on ultra-low-power platforms like Arduino Nano BLE and nRF52840 for real-time vehicle logo classification [11] and haze detection [12]. In wireless communication and energy monitoring, MobileViT and EfficientNet architectures were adapted for signal classification and GAF-based energy profiling [13]. In recent developments, PCNet—based on EfficientNetV2—has been proposed for insect pest classification [14], while LightSoilNet, a 2.7M-parameter CNN, outperformed other compact models for soil classification [15].

Despite these advances, Kerangas forest succession remains underexplored, particularly with high-resolution, field-acquired imagery. Moreover, prior studies rarely report resource metrics such as memory usage or inference latency—key aspects for deployment on edge devices. Few also analyze how MobileNet configurations (e.g., input resolution, width multiplier α, and modality) influence the trade-off between accuracy and efficiency. This motivates our investigation into grayscale-optimized MobileNet variants for on-device classification of Kerangas ecological succession.

3 Methodology

3.1 Data Collection and Annotation

A total of 162 images were collected through direct fieldwork across 30 locations within the districts of Kubu, Kumai, and Kotawaringin Barat Regency. These observations were conducted during 12 separate field sessions between May and November, each guided by ecological protocols to ensure ground-truth relevance to tropical heath forest (Kerangas) conditions. Three visually distinguishable classes were defined based on post-disturbance ecological succession stages. The Mid class represents intermediate recovery, characterized by medium vegetation density with partially established undergrowth and shrubs. The Late class denotes well-recovered heath forest, marked by dense vegetation cover, mature shrubs, and closed canopy features. Meanwhile, the No Recovery class corresponds to severely degraded or recently disturbed landscapes—often caused by fire or land conversion—characterized by minimal vegetation cover and exposed soil.

Image labeling was conducted collaboratively by a team of three annotators. One of the annotators holds a formal background in biology and ecological studies, providing ecological guidance during the labeling process. The classification relied on observable indicators—such as canopy density, shrub presence, and soil exposure—defined in alignment with ecological succession literature [16]. Annotations were further validated against field notes and photographic documentation to ensure consistency. Figure 1 illustrates representative samples from each class. These visual examples help clarify the distinguishing ecological characteristics used during the annotation process. The dataset was split using stratified sampling into 78% training and 22% testing subsets to maintain class balance across experiments.

FIGURE 1. Representative field images for each ecological class. Left: No Recovery; Center: Mid; Right: Late.

3.2 Model Architecture and Configuration

To evaluate performance under resource-constrained scenarios, we implemented lightweight CNNs using MobileNetV1 and V2 architectures, optimized for 96×96 inputs [17]. Width multipliers (α) were applied to scale model size: V1 with $\alpha = \{0.25, 0.20, 0.10\}$ required 105.9–53.2 KB RAM, while V2 with $\alpha = \{0.35, 0.10, 0.05\}$ used 296.8–265.3 KB; flash memory ranged from 101 to 575.2 KB. To ensure consistent training and avoid domain shifts, models were trained separately for RGB and grayscale inputs. All models shared a common architecture with $96 \times 96 \times 3$ input (grayscale replicated across channels), a truncated MobileNetV1/V2 backbone, followed by global average pooling, dropout (rate = 0.2), and a softmax classifier for three ecological classes. This modular design enabled reliable benchmarking across model sizes and input types.

3.3 Training Procedure

All models were trained using a unified pipeline for consistency, differing only in backbone architecture (MobileNetV1/V2 with various width multipliers) and input modality (RGB or grayscale). Each model used pretrained weights with input size $96 \times 96 \times 3$; grayscale images were replicated across channels. The convolutional base was initially frozen, with a lightweight classification head comprising dropout, flattening, and a softmax dense layer. Training was performed in two stages: (1) 80 epochs with Adam optimizer (learning rate =

0.0005) and categorical cross-entropy loss, and (2) fine-tuning for 10 additional epochs after unfreezing 65% of the base layers with a reduced learning rate (0.000045). All models used batch size 32, deterministic mode (when enabled), prefetch buffering, and callbacks for logging and early stopping. The best checkpoint from the initial phase was reloaded before fine-tuning.

3.4 Experimental Design

To evaluate the influence of model architecture and input modality, we conducted a series of controlled experiments involving twelve model configurations: six MobileNetV1 and six MobileNetV2 variants, each trained separately on RGB and grayscale datasets. The width multipliers (α) varied across values of 0.1, 0.2, and 0.25 for MobileNetV1, and 0.05, 0.1, and 0.35 for MobileNetV2. This design allowed us to systematically assess the trade-offs between input complexity, model size, and classification performance across different resource budgets. A summary of the experimental setup is provided in Table 1.

TABLE 1. Experimental Model Configurations

Model	Input Type	Width Multiplier	Notation
MobileNetV1	RGB	0.25	V1-RGB-0.25
MobileNetV1	RGB	0.20	V1-RGB-0.20
MobileNetV1	RGB	0.10	V1-RGB-0.10
MobileNetV1	Grayscale	0.25	V1-GS-0.25
MobileNetV1	Grayscale	0.20	V1-GS-0.20
MobileNetV1	Grayscale	0.10	V1-GS-0.10
MobileNetV2	RGB	0.35	V2-RGB-0.35
MobileNetV2	RGB	0.10	V2-RGB-0.10
MobileNetV2	RGB	0.05	V2-RGB-0.05
MobileNetV2	Grayscale	0.35	V2-GS-0.35
MobileNetV2	Grayscale	0.10	V2-GS-0.10
MobileNetV2	Grayscale	0.05	V2-GS-0.05

4 Results and Discussion

This section presents the performance evaluation of the lightweight CNN models trained for classifying Kerangas landscape imagery. The evaluation focuses on classification accuracy, confusion matrices, and resource utilization metrics. Additionally, the impact of input modalities (RGB vs. grayscale) and model architectures (MobileNetV1 vs. MobileNetV2) is discussed to understand their practical suitability for edge deployment.

4.1 Data Overview and Feature Space Visualization

As detailed in Section III-A, we applied a t-distributed Stochastic Neighbor Embedding (t-SNE) projection to the test set embeddings extracted from the penultimate layer of the best-performing model [18]. Figure 2 presents the 2D feature space, where circles represent ground-truth labels and squares denote predicted classes. The axes are labeled as t-SNE 1 and t-SNE 2. The plot reveals three well-formed and relatively separable clusters corresponding to the ecological succession stages: *Late*, *Mid*, and *No Recovery*. Minor overlaps, particularly involving *Mid*, are expected due to the transitional nature of these classes.

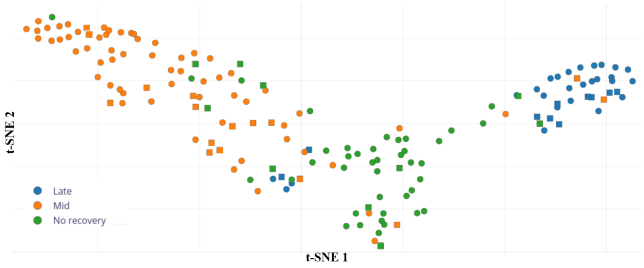

FIGURE 2. t-SNE visualization of test samples across three classes. Circles denote ground-truth labels; squares denote predicted classes. Axes represent the two t-SNE projection dimensions.

4.2 Classification Performance

All models were evaluated on a held-out test set using FLOAT32 precision to assess their generalization capabilities. Tables 2 and 3 present the performance of selected RGB and grayscale models, respectively.

TABLE 2. Test Set Performance of Selected RGB Models (FLOAT32)

Model	Accuracy (%)	F1-score	ROC AUC
MobileNetV2 160x160 $\alpha = 0.75$	95.83	0.9581	0.9694
MobileNetV2 160x160 $\alpha = 0.50$	91.67	0.9185	0.9930
MobileNetV2 160x160 $\alpha = 0.35$	91.67	0.9250	0.9977
MobileNetV2 96x96 $\alpha = 0.10$	87.50	0.8750	0.9807
MobileNetV1 96x96 $\alpha = 0.35$	83.33	0.8481	0.9778

RGB models generally outperform grayscale models, particularly in lower-capacity configurations. However, MobileNetV2 96×96 with $\alpha = 0.35$ (Grayscale) achieved the same accuracy and F1-score as the best RGB model, demonstrating its robustness and potential in constrained environments.

TABLE 3. Test Set Performance of Selected Grayscale Models (FLOAT32)

Model	Accuracy (%)	F1-score	ROC AUC
MobileNetV2 96×96 $\alpha = 0.35$ (Gray)	95.83	0.9581	0.9906
MobileNetV1 96×96 $\alpha = 0.25$ (Gray)	91.67	0.9155	0.9634
MobileNetV2 96×96 $\alpha = 0.10$ (Gray)	83.33	0.8328	0.9513
MobileNetV2 96×96 $\alpha = 0.05$ (Gray)	83.33	0.8447	0.9444
MobileNetV1 96×96 $\alpha = 0.20$ (Gray)	79.17	0.7873	0.9446

4.3 Resource Utilization

Resource consumption is crucial for edge applications. Table 4 lists inference time, RAM, and Flash usage for all models. The most accurate RGB model (V2-160-0.75) is computationally heavy, requiring over 5 MB of Flash and taking 36 seconds per inference—rendering it impractical for real-time edge deployment. In contrast, MobileNetV2 96×96 $\alpha = 0.35$ (Grayscale) achieves equivalent accuracy with less than 1 second inference time and memory usage under 600 KB, making it an optimal trade-off. Ultra-lightweight models such as MobileNetV2 96×96 $\alpha = 0.05$ (Gray) offer sub-second inference and minimal memory usage, suitable for latency-sensitive applications with relaxed accuracy requirements.

TABLE 4. Inference Efficiency and Memory Footprint of Selected Models

Model	Input	Inference Time (ms)	Peak RAM (KB)	Flash (KB)
MobileNetV2 160×160 $\alpha = 0.75$	RGB	36353	860.8	5300.0
MobileNetV2 160×160 $\alpha = 0.50$	RGB	19295	501.7	2800.0
MobileNetV2 160×160 $\alpha = 0.35$	RGB	6751	441.6	1700.0
MobileNetV1 96×96 $\alpha = 0.35$	RGB	1798	232.9	587.4
MobileNetV2 96×96 $\alpha = 0.10$	RGB	1001	169.2	224.9
MobileNetV2 96×96 $\alpha = 0.35$	Gray	933	214.6	587.1
MobileNetV1 96×96 $\alpha = 0.25$	Gray	8078	202.3	872.1
MobileNetV2 96×96 $\alpha = 0.10$	Gray	3414	106.9	192.4
MobileNetV2 96×96 $\alpha = 0.05$	Gray	828	159.2	174.8
MobileNetV1 96×96 $\alpha = 0.20$	Gray	6372	164.4	580.5

4.4 Learning Curve Analysis

To better understand model convergence and learning dynamics, training and validation loss/accuracy curves were analyzed for the best-performing models in RGB and grayscale settings. Figures 3 and 4 show the respective learning curves.The RGB model, *MobileNetV2 160×160* $\alpha = 0.75$, achieves near-perfect training accuracy within the first 10 epochs and maintains stable validation accuracy around 95%. Both training and validation losses converge quickly and remain low, indicating strong generalization without overfitting. This model exhibits the highest overall performance but requires substantial memory and inference time. The grayscale counterpart, *MobileNetV2 96×96* $\alpha = 0.35$, demonstrates slower but steady learning. Despite limited input modality, it eventually reaches the same accuracy (95.8%) as the RGB model, with slightly higher validation loss and a more gradual convergence. Its training curve remains stable with minimal overfitting, suggesting that spatial and textural features alone can drive reliable classification. These learning dynamics confirm that MobileNetV2 96×96 $\alpha = 0.35$ (Grayscale) is the most efficient model in terms of resource-performance trade-off, while MobileNetV2 160×160 $\alpha = 0.75$ (RGB) remains the most accurate, though less practical for edge deployment. While RGB inputs generally enhance accuracy due to richer information, grayscale models—particularly MobileNetV2 $\alpha = 0.35$—achieve comparable performance with significantly reduced resource usage, making them suitable for low-bandwidth and embedded scenarios.

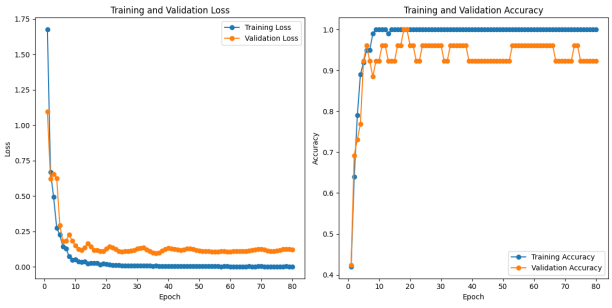

FIGURE 3. Training and validation curves of MobileNetV2 160×160 $\alpha = 0.75$ (RGB).

5 Discussion and Limitations

Our findings highlight that model architecture, width multiplier (α), and input modality jointly influence both accuracy and deployment efficiency. Across all tested configurations,

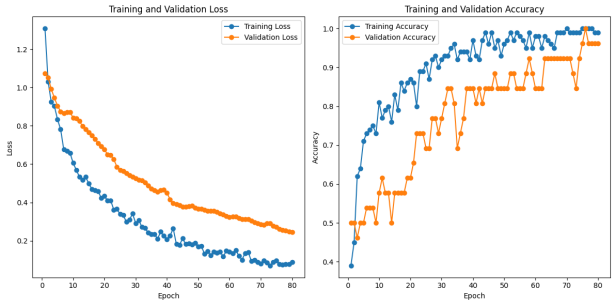

FIGURE 4. Training and validation curves of MobileNetV2 96×96 $\alpha = 0.35$ (Grayscale).

MobileNetV2 consistently outperformed MobileNetV1 due to its depthwise separable convolutions and optimized residual connections. Although RGB inputs deliver richer information and accelerate convergence, grayscale inputs demonstrate strong generalization with drastically reduced memory and latency costs.

Remarkably, MobileNetV2 96×96 ($\alpha = 0.35$, Grayscale) achieves comparable test accuracy and F1-score to the RGB-based MobileNetV2 160×160 ($\alpha = 0.75$), while consuming over 90% less memory and inference time. This suggests that grayscale-optimized models are highly suitable for real-time ecological monitoring on embedded edge devices.

However, training dynamics differ: RGB models converge more rapidly and stably, whereas grayscale models exhibit a wider training–validation gap and slower convergence—suggesting a higher overfitting risk during early training due to lower input entropy. Nonetheless, the grayscale model's final performance remains robust.

t-SNE visualizations further validate the model's ability to distinguish between succession stages [19], though performance variability was observed in samples captured under non-uniform lighting (e.g., late afternoon), potentially introducing noise [20]. Standardizing image acquisition—especially capture time—could improve data consistency in future work.

Succession class definitions were tailored to Kerangas conditions: *No Recovery* (bare or exposed post-disturbance soil), *Mid* (intermediate regrowth dominated by *Acacia mangium*), and *Late* (dense mature vegetation including *Cratoxylum arborescens*). These were assigned pragmatically rather than rigid ecological succession stages. Field data were collected in the 2024 dry season without destructive sampling.

EfficientNet variants were initially considered, but excluded due to structural complexity and suboptimal performance in early tests. The final focus on MobileNet aligns with real-world deployment practices for TinyML and edge AI. All configurations were logged for reproducibility, though only representative results are reported; non-converging or suboptimal models are omitted without affecting the study's integrity.

6 Conclusion

This study demonstrates that lightweight CNNs—particularly MobileNetV2 variants—are effective for Kerangas landscape classification in resource-constrained settings. Through comparative evaluation of twelve configurations under RGB and grayscale modalities, we identified MobileNetV2 96×96 with $\alpha = 0.35$ (Grayscale) as the optimal trade-off model, achieving 95.8% accuracy with peak RAM usage under 215 KB and inference time below 1 second.

The model's generalization ability, along with its minimal memory footprint, positions it as a practical solution for real-time ecological monitoring on embedded edge devices. These findings support the adoption of grayscale-optimized deep learning models in domains where computational resources, energy efficiency, and latency are critical. Future work may explore domain adaptation, temporal ensembling, or lightweight augmentation pipelines to enhance robustness under dynamic field conditions.

Acknowledgements

This research was funded by the IEEE Student Branch of Universitas AMIKOM Yogyakarta under the IEEE ASB Research Grant program for the 2024 fiscal year.

References

[1] I. S. Mohd Ikbal, S. Md Jaafar, N. Ahmad, and R. S. Sukri, "Contrasting environmental drivers of tree community variation within heath forests in Brunei Darussalam, Borneo," *Biodivers. Data J.*, vol. 12, Dec. 2024, Art. no. e127919, doi: 10.3897/BDJ.12.e127919.

[2] R. M. Nisaa, U. K. Sari, and Y. B. Sulistioadi, "Heath forest identification using remote sensing in the surrounding area of the new capital city of Indonesia," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1266, no. 1, p. 012068, 2023, doi: 10.1088/1755-1315/1266/1/012068.

[3] C. Sothe, et al., "A comparison of machine and deep-learning algorithms applied to multisource data for a subtropical forest area classification," *Int. J. Remote Sens.*, vol. 41, no. 5, pp. 1943–1969, 2020, doi: 10.1080/01431161.2019.1681600.

[4] J. Wang, M. Bretz, M. A. A. Dewan, and M. A. Delavar, "Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects," *Sci. Total Environ.*, vol. 822, 2022, Art. no. 153559, doi: 10.1016/j.scitotenv.2022.153559.

[5] V. Gautam, N. K. Trivedi, A. Singh, H. G. Mohamed, I. D. Noya, P. Kaur, and N. Goyal, "A Transfer Learning-Based Artificial Intelligence Model for Leaf Disease Assessment," *Sustainability*, vol. 14, no. 20, p. 13610, 2022, doi: 10.3390/su142013610.

[6] M. Flórez, O. Becerra, E. Carrillo, M. Villa, Y. Álvarez, J. Suárez, and F. Mendes, "Deep Learning Application for Biodiversity Conservation and Educational Tourism in Natural Reserves," *ISPRS Int. J. Geo-Inf.*, vol. 13, p. 358, 2024, doi: 10.3390/ijgi13100358.

[7] T. He, H. Zhou, C. Xu, J. Hu, X. Xue, L. Xu, X. Lou, K. Zeng, and Q. Wang, "Deep Learning in Forest Tree Species Classification Using Sentinel-2 on Google Earth Engine: A Case Study of Qingyuan County," *Sustainability*, vol. 15, no. 3, p. 2741, 2023, doi: 10.3390/su15032741.

[8] R. Sapkota, P. Sharma, and I. Mann, "Comparison of Deep Learning Models for the Classification of Noctilucent Cloud Images," *Remote Sensing*, vol. 14, no. 10, p. 2306, 2022, doi: 10.3390/rs14102306.

[9] R. O. Ogundokun, S. Misra, A. O. Akinrotimi, and H. Ogul, "MobileNet-SVM: A Lightweight Deep Transfer Learning Model to Diagnose BCH Scans for IoMT-Based Imaging Sensors," *Sensors*, vol. 23, no. 2, p. 656, 2023, doi: 10.3390/s23020656.

[10] M. Ragab, S. Alshehri, G. A. Azim, H. M. Aldawsari, A. Noor, J. Alyami, and S. Abdel-Khalek, "COVID-19 Identification System Using Transfer Learning Technique With Mobile-NetV2 and Chest X-Ray Images," *Front. Public Health*, vol. 10, p. 819156, 2022, doi: 10.3389/fpubh.2022.819156.

[11] D. Abhinay, S. V. Vighnesh, L. K. Durgam, and R. K. Jatoth, "Real-time Classification of Vehicle Logos on Arduino Nano BLE using Edge Impulse," in *Proc. 4th Int. Conf. Signal Process. Commun. (ICSPC)*, Coimbatore, India, 2023, pp. 316–320, doi: 10.1109/ICSPC57692.2023.10126068.

[12] R. G. Bhamidipati, R. K. Jatoth, M. Naresh, and S. P. Surepally, "Real-time Classification of Haze and Non-Haze Images on Arduino Nano BLE Using Edge Impulse," in *Proc. 4th Int. Conf. Comput. Commun. Syst. (I3CS)*, Shillong, India, 2023, pp. 1–7, doi: 10.1109/I3CS58314.2023.10127312.

[13] Q. Zheng, S. Saponara, X. Tian, et al., "A Real-Time Constellation Image Classification Method of Wireless Communication Signals Based on the Lightweight Network MobileViT," *Cogn. Neurodyn.*, vol. 18, pp. 659–671, 2024, doi: 10.1007/s11571-023-10015-7.

[14] T. Zheng, X. Yang, J. Lv, M. Li, S. Wang, and W. Li, "An Efficient Mobile Model for Insect Image Classification in the Field Pest Management," *Eng. Sci. Technol. Int. J.*, vol. 39, p. 101335, 2023, doi: 10.1016/j.jestch.2023.101335.

[15] D. N. K. Pandiri, R. Murugan, and T. Goel, "Smart Soil Image Classification System Using Lightweight Convolutional Neural Network," *Expert Syst. Appl.*, vol. 238, p. 122185, 2024, doi: 10.1016/j.eswa.2023.122185.

[16] L. Poorter, M. T. van der Sande, L. Amissah, F. Bongers, I. Hordijk, J. Kok, S. G. W. Laurance, *et al.*, "A Comprehensive Framework for Vegetation Succession," *Ecosphere*, vol. 15, no. 4, p. e4794, 2024, doi: 10.1002/ecs2.4794.

[17] A. Mellit, N. Blasuttigh, S. Pastore, M. Zennaro, and A. M. Pavan, "TinyML for Fault Diagnosis of Photovoltaic Modules Using Edge Impulse Platform and IR Thermography Images," *IEEE Trans. Ind. Appl.*, early access, 2025, doi: 10.1109/TIA.2025.3556792.

[18] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008. https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

[19] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed., Amsterdam, Netherlands: Academic Press, 2009.

[20] Y. Hold-Geoffroy, A. Athawale, and J.-F. Lalonde, "Deep sky modeling for single image outdoor lighting estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 6926–6935.