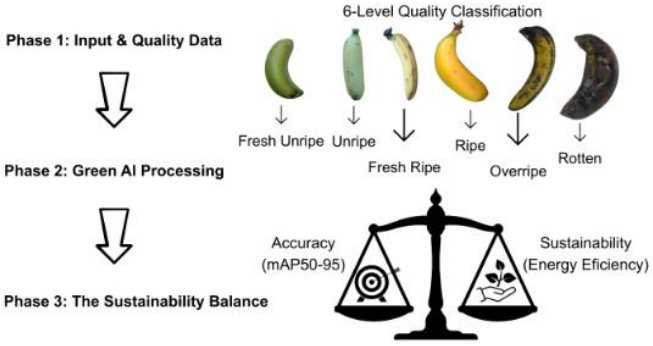
ENERGY-AWARE BENCHMARKING OF LIGHTWEIGHT YOLO MODELS FOR SUSTAINABLE BANANA QUALITY GRADING

NGOC BAO HO[1*], NGUYET THI MINH NGUYEN[2]
[1]*Department of Food Technology, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao, Ward 4, Go Vap District, Ho Chi Minh City, Vietnam*
[2]*Department of Nutrition and Food Science, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao, Ward 4, Go Vap District, Ho Chi Minh City, Vietnam*
** Corresponding author: ngochb26091@pgr.iuh.edu.vn*

Abstract. Post-harvest losses present a critical challenge in global agricultural supply chains, where inconsistent manual grading drives significant food waste and economic instability. While AI-based computer vision offers a promising automated solution, the environmental cost of deploying deep learning models is frequently ignored. Lightweight networks are often widely considered to be energy-efficient, yet this requires strict empirical validation. This study investigates the intersection of accuracy and sustainability by benchmarking four YOLO architectures on over 11,000 annotated banana images. Our empirical results challenge the assumption that legacy compact models guarantee environmental efficiency: YOLOv3-tiny consumed the highest energy (0.2036 kWh) while delivering the lowest accuracy (mAP50-95 = 0.6418). Conversely, modern nano-scale architectures achieve an improved balance. YOLOv8n minimized energy consumption to just 0.1711 kWh while maintaining strong accuracy (mAP50-95 = 0.7084), and YOLO11n maximized precision (mAP50-95 = 0.7112) with only a modest computational increase. Crucially, cost-benefit analysis reveals that the carbon emissions generated during the training of these optimized models are minimal compared to the considerable greenhouse gases avoided by drastically reducing post-harvest fruit waste. This evidence-based framework demonstrates that deploying Green AI in agriculture can secure high productivity while strictly ensuring long-term ecological preservation.

Keywords. Green AI, banana quality grading, YOLO, energy efficiency, error analysis, sustainable agriculture.



1 INTRODUCTION

Over the past decade, artificial intelligence (AI) has progressed from experimental systems to practical applications across healthcare, food industries, and agricultural supply chains [1]. Within the context of Agriculture 4.0, the integration of AI, the Internet of Things (IoT), and edge computing is enabling intelligent production systems oriented toward food security and sustainable development [2]. Bananas constitute a strategic fruit commodity for Vietnam, providing nutritional value and contributing substantially to export revenues, with an estimated export turnover of approximately USD 372 million in 2024 [3]. However, current manual grading practices often result in substantial inconsistency, thereby increasing post-harvest losses. Modern computer-vision models such as YOLO, Faster R-CNN and DETR have demonstrated strong capability in addressing these limitations [4, 5]. Recent research trends further emphasize lightweight architectures designed for deployment devices [6, 7] and hardware-aware

© 2026 Industrial University of Ho Chi Minh City

compression techniques [8]. Beyond overall grading accuracy, understanding the nature of detection errors is essential in post-harvest systems, where fruit deformation, surface defects, and class ambiguity can directly affect grading reliability.

Despite this progress, most existing studies continue to prioritize accuracy or speed without adequately quantifying energy cost and emissions core dimensions of "Green AI" [9]. A largely untested paradox remains: do so-called "tiny" legacy models actually consume less energy than more advanced anchor-free architectures. The absence of empirical evaluation of real electricity consumption (kWh) introduces deployment risks, particularly when models are intended for resource-constrained edge platforms. This challenge is increasingly relevant as Vietnam has committed to agricultural emission reductions and participation in the Global Methane Pledge [10]. With post-harvest losses estimated at 20–25%, each kilogram of wasted bananas may generate approximately 0.5–1.1 kg of CO_2 equivalent per kilogram according to recent life-cycle assessments [11, 12]. Consequently, an AI-based grading system must not only achieve high classification accuracy but also maintain a low carbon footprint, ensuring a positive net environmental balance.

Building on this context, the present study conducts a comprehensive assessment of accuracy energy trade-offs among lightweight YOLO models (v3-tiny, v5n, v8n, v11n) for six-level banana quality classification. An extended Cost–Benefit Analysis (CBA) framework is applied to compare technical performance alongside life-cycle emissions. The objective is to identify the model achieving the optimal equilibrium point, thereby illustrating the transition toward "Green AI" in modern architectures and supporting Vietnam's Net-Zero roadmap and sustainable agricultural development strategies [13].

2 MATERIAL AND METHOD

2.1 Data, sampling strategy, hardware, and model architectures

The dataset consists of banana images manually annotated into six quality categories: fresh ripe, fresh unripe, ripe, overripe, unripe, and rotten. The data were intentionally partitioned to ensure balance and prevent information leakage across splits: training (9,592 images), validation (816 images), and an independent test set (414 images). The independent test split enables an objective assessment of model generalization. The models evaluated include YOLOv3-tiny, YOLOv5n, YOLOv8n, and YOLO11n all lightweight variants designed for deployment on constrained devices while maintaining acceptable accuracy levels. All models were trained using the Ultralytics YOLO framework, which integrates standardized pipelines for training, inference, and evaluation, thereby reducing implementation bias [14].

Experiments were conducted under a consistent hardware configuration: Intel Core i5-12450HX (12 cores), NVIDIA GeForce RTX 3050 (6 GB), ~20 GB RAM, Windows 11, and Python 3.11.9. Maintaining a fixed configuration ensures fairness in cross-model comparison.

2.2 Training procedure

The training process was designed to balance speed and accuracy, following optimization recommendations from YOLOv4 [15]. Core hyperparameters included: 50 epochs, batch size 16, image resolution 640×640, SGD optimizer (learning rate = 0.01), cosine decay for learning-rate scheduling, momentum 0.937, and early stopping (patience = 10). The pipeline comprised: (i) loading pretrained weights; (ii) training on the training set; (iii) monitoring performance on the validation set; (iv) saving the best.pt checkpoint based on peak mAP; (v) logging metrics into result.csv. To enhance reproducibility, the random seed was fixed and computational settings were kept stable across runs.

2.3 Energy and emission tracker

Alongside performance evaluation, energy consumption and emissions were tracked using CodeCarbon, which records hardware power draw, execution time, and region-specific emission factors [16, 17].

Energy consumption was estimated as:

$$E = \sum_{t} P_{total,t} \times \Delta t \tag{1}$$

where:

$$P(t) = P_{CPU} + P_{GPU} + P_{RAM} \quad (2)$$

Estimated CO_2-equivalent emissions were computed as:

$$C_{emissions} = E \times I_{carbon} \times PUE \quad (3)$$

Where I_{carbon} denotes the carbon intensity of the grid (approximately 0.716 kgCO_2e/kWh for the Vietnamese grid mix), and $PUE \approx 1$ personal computing environments. All measurements were stored in emissions.csv, enabling comparison of models from a Green-AI perspective.

2.4 Model evaluation

While standard detection metrics were employed for quantitative evaluation, greater emphasis was placed on system-level relevance for post-harvest grading under energy and infrastructure constraints. To ensure robustness and reproducibility, each model was trained and evaluated across five independent runs using different random initializations, and performance results were reported as mean values with associated variability. This approach reduces the influence of stochastic effects and supports reliable comparison between architectures.

Detection errors were further examined using the TIDE toolkit, which decomposes performance degradation into interpretable error categories, including localization, classification, background, duplicate, and missing detections, as well as false positives and false negatives. By aggregating error patterns across repeated runs, this analysis identifies consistent sources of grading unreliability and provides practical guidance for improving post-harvest quality assessment systems [18]. Statistical significance was determined using one-way Analysis of Variance (ANOVA) followed by Tukey's HSD post-hoc test ($p < 0.05$) using Statgraphics Centurion XV.

3 RESULTS AND DISCUSSION

3.1 Comparing the results of the training process

The training processes of the four models (YOLOv3-tiny, YOLOv5n, YOLOv8n, and YOLO11n) were evaluated across three dimensions: (i) convergence behavior (loss dynamics), (ii) recognition performance (mAP), and (iii) energy consumption. Experimental data extracted from result.csv and emissions.csv reveal pronounced performance differentiation, reflecting the paradigm shift from anchor-based to anchor-free network design.

The loss curves indicate that YOLOv3-tiny exhibits substantial oscillations and the slowest convergence rate, particularly in box_loss. This observation is consistent with the discussion by Redmon and Farhadi (2018) [19], who highlight inherent limitations of anchor-based mechanisms, where the model must continuously refine pre-defined anchor templates. Such structural rigidity constrains the network when dealing with objects characterized by curved geometry and irregular deformation, such as bananas. In contrast, YOLOv8n and YOLO11n demonstrate faster and more stable convergence. This improvement is attributed to the adoption of anchor-free detection and the incorporation of a decoupled detection head, which separates classification and localization objectives, thereby mitigating optimization conflicts [20]. Superior mAP50–95 scores relative to YOLOv3-tiny further indicate the benefits of backbone refinements. In YOLOv5n, CSPNet alleviates gradient redundancy and enhances learning efficiency [21], whereas in YOLOv8n, C2f blocks inspired by ELAN facilitate more effective gradient flow control [22]. Moreover, YOLO11n leverages Spatial Attention (C2PSA), analogous to CBAM [23], enabling the model to emphasize defect regions while suppressing background noise. Across repeated training runs, modern lightweight architectures exhibited lower performance variability compared with legacy compact models.

Lower variability across repeated training runs indicates reproducible model behavior, which is essential for maintaining consistent grading performance in post-harvest handling environments.

A quantitative analysis of convergence dynamics, defined as the epoch required to reach 95% of peak performance, further substantiates these architectural advantages. YOLOv8n demonstrates the most rapid feature assimilation, stabilizing its mAP50 metric at epoch 19, which is approximately 17% faster than the legacy YOLOv3-tiny (epoch 23). This acceleration indicates that the anchor-free head and Task-Aligned Assigner significantly reduce the search space for the optimizer [14, 22]. Furthermore, regarding localization precision (box_loss), YOLO11n exhibits superior optimization efficiency by converging at epoch 31, outperforming YOLOv8n (epoch 34) and significantly surpassing YOLOv5n (epoch 41). This finding suggests that the integration of spatial attention mechanisms (C2PSA) allows the network to accurately localize to object boundaries more effectively, thereby reducing the computational budget required to reach optimal performance [16, 24].

Energy monitoring results highlight an additional contrast. Despite being labeled "tiny," YOLOv3-tiny records the highest energy consumption and CO_2 emissions due to reliance on standard convolutions with

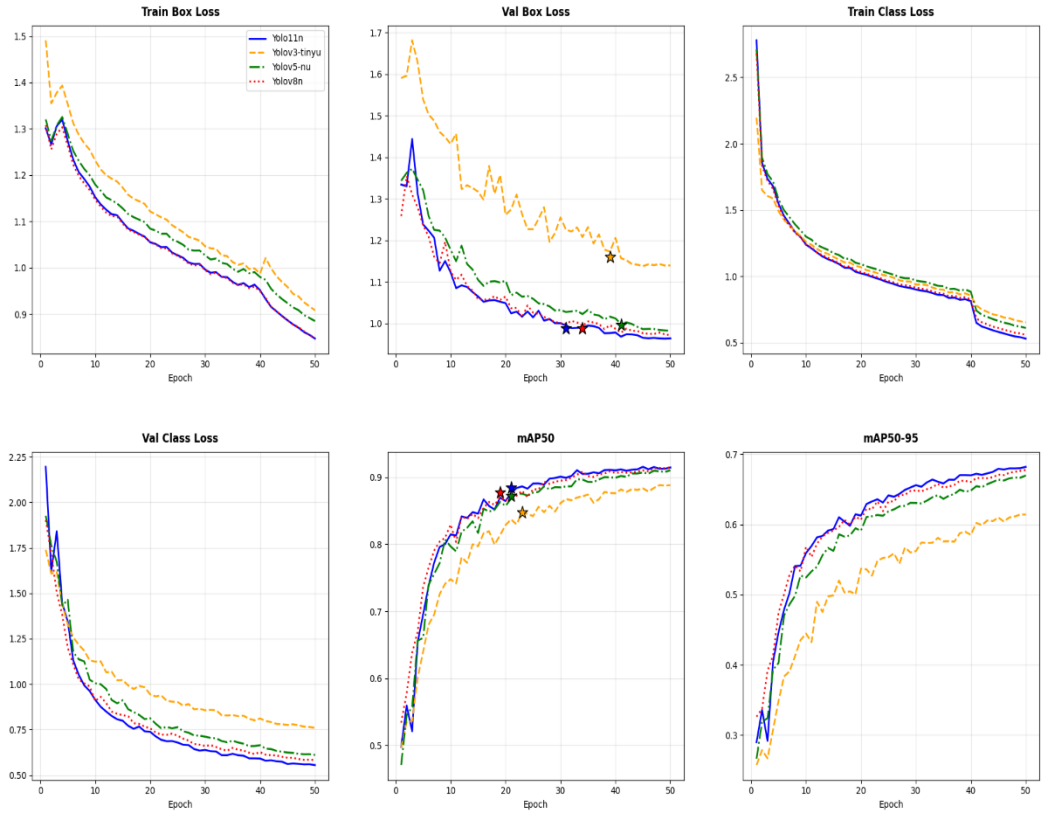


Figure 1: Comparison of YOLO Models Training Metrics.

a relatively large parameter count but low representational efficiency an example of "Red AI" inefficiency [24]. Conversely, YOLOv8n and YOLO11n achieve lower consumption by reducing FLOPs through architectural optimization, aligning with deployment requirements on energy-constrained edge devices in smart agriculture [16].

Table 1. Resource Consumption & Emissions Comparison (CodeCarbon)

Model	YOLOv3-tiny	YOLOv5n	YOLOv8n	YOLO11n
Energy Consumption (kWh)	0.2036 ± 0.0005^{a}	0.1723 ± 0.0001^{c}	0.1711 ± 0.0001^{d}	0.1820 ± 0.0006^{b}
CO_2 Emission (Kg)	0.0968 ± 0.0002^{a}	0.0819 ± 0.0001^{c}	0.0814 ± 0.0000^{d}	0.0865 ± 0.0003^{b}

Data are presented as Mean ± SD. Different letters (e.g., a, b, c) indicate statistically significant differences ($p < 0.05$) determined by one-way ANOVA followed by Tukey's HSD post-hoc test using Statgraphics Centurion XV.

From an agricultural systems perspective, the reduced performance variability across repeated training runs supports the robustness of lightweight models for routine post-harvest operations, where stable grading behavior is often prioritized over marginal accuracy gains. Overall, the superior performance of YOLOv8n and YOLO11n is primarily attributable to: (i) removal of anchor constraints [19, 20], (ii) improved gradient utilization via CSPNet/ELAN [21, 22], and (iii) enhanced representational capacity through attention mechanisms [23]. While YOLOv3-tiny shows clear limitations in accuracy and performance variability, YOLOv8n emerges as a balanced option for scenarios requiring sustainable deployment. Specifically, the lower variability across training runs indicates reproducible model behavior, which is critical for consistent grading performance in post-harvest environments. Furthermore, the observed stable convergence behavior suggests that the model can be trained reliably without extensive hyperparameter tuning, supporting deployment in resource-limited agricultural settings.

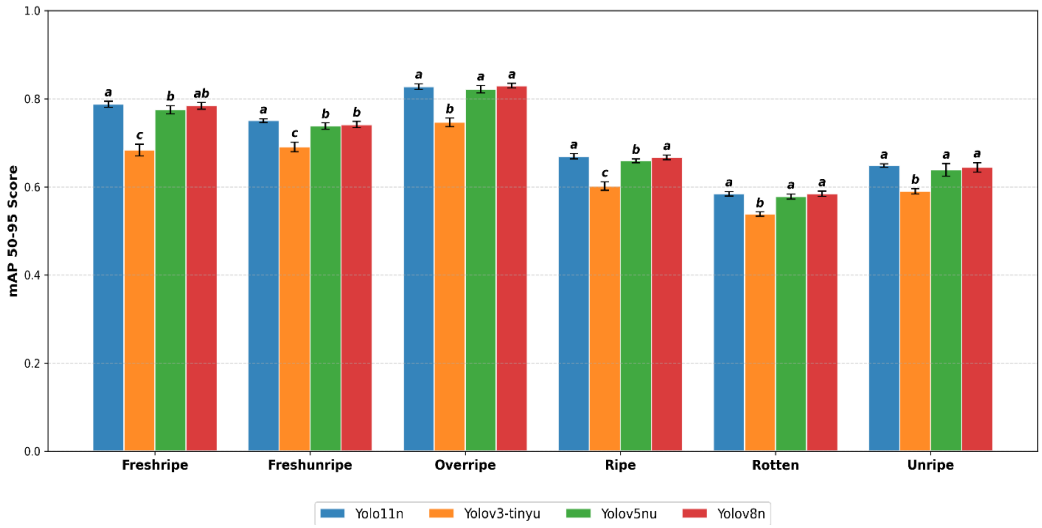


Figure 2. Detection Performance per Banana Class (mAP 50-95). Data are presented as Mean ± SD. Different letters (e.g., a, b, c) indicate statistically significant differences ($p < 0.05$) determined by one-way ANOVA followed by Tukey's HSD post-hoc test using Statgraphics

3.2 Accuracy comparison and error analysis

Experiments were conducted on a NVIDIA GeForce RTX 3050, evaluating both detection accuracy and computational efficiency. As shown in Table 2, YOLO11n achieves the highest precision with an mAP50 of 0.9483 ± 0.0031. However, substantial divergence appears when examining mAP50–95. While visually salient categories (overripe, freshripe) achieve mAP50 > 0.948, the rotten class deteriorates sharply at mAP50–95 (0.5841 ± 0.0051), indicating a specific weakness in detecting decay with high precision.

Table 2. Overall Performance Comparison (mAP50 vs. mAP50–95).

Model	YOLOv3-tiny	YOLOv5n	YOLOv8n	YOLO11n
mAP50	0.9228 ± 0.0054^{b}	0.9454 ± 0.0010^{a}	0.9457 ± 0.0029^{a}	0.9483 ± 0.0031^{a}
mAP95	0.6418 ± 0.0044^{c}	0.7017 ± 0.0035^{b}	0.7084 ± 0.0021^{a}	0.7112 ± 0.0028^{a}

Data are presented as Mean ± SD. Different letters (e.g., a, b, c) indicate statistically significant differences ($p < 0.05$) determined by one-way ANOVA followed by Tukey's HSD post-hoc test using Statgraphics Centurion XV.

Biological & Geometric Factors: Three primary factors drive the performance drop in the rotten class. First, high inter-class similarity between late-ripening and decay stages leads to fine-grained decision boundaries that are difficult to separate [25]. Second, large intra-class variability limits generalization capacity [26].

Third, heavily damaged fruit exhibit non-rigid geometric deformation that conventional convolutional networks struggle to model [27]. To address the localization challenges in the rotten class, future work should incorporate targeted data augmentation strategies, acquire images from more diverse viewing angles, and include additional samples of geometrically deformed decayed fruits.

Localization Error: TIDE analysis reveals that localization dominates mAP loss across all models. This reflects the geometric mismatch between axis-aligned bounding boxes and the intrinsically curved geometry of bananas [28].

Architectural Evolution: Classification errors differ sharply across generations. YOLOv3-tiny exhibits substantially higher classification error (dAP_cls = 2.1420 ± 0.2691), attributable to its anchor-based design and shallow depth [19]. In contrast, newer Nano variants benefit from anchor-free decoupled heads that mitigate optimization conflicts [20]. Furthermore, the reliance on LeakyReLU and absence of attention mechanisms in older models bias them toward local textures, increasing false detections in cluttered contexts [29].

Training Dynamics: YOLOv3-tiny converges rapidly but saturates prematurely, consistent with representational limits in shallow networks [30]. Conversely, modern models (e.g., YOLOv8n) show high false-positive rates combined with exceptionally low miss rates (0.0497 ± 0.1110), indicating recall-oriented optimization driven by modern loss functions [31, 32].

From a post-harvest operational perspective, these metrics dictate system viability. The high throughput and stable convergence of YOLOv8n and YOLO11n support real-time grading on high-speed conveyor belts. However, the prevalence of localization errors exacerbated by the curved geometry of the fruit suggests challenges for robotic grasping mechanisms, which require precise coordinate estimation to avoid mechanical damage. Additionally, the specific difficulty in separating decay stages implies that strictly "quality control" lines may require stricter confidence thresholds or multi-modal sensing to prevent spoiled fruit from entering the supply chain.

Table 3. Quantitative error decomposition using the TIDE framework across four YOLO models.

Model	YOLOv3-tinyu	YOLOv5nu	YOLOv8n	YOLO11n
Cls	2.1420±0.2691[a]	1.3551±0.1949[b]	1.3140±0.1869[b]	1.2496±0.1726[b]
Loc	2.4827±0.2452[a]	1.7897±0.3728[b]	1.8293±0.1218[b]	1.8199±0.2790[b]
Both	0.6165±0.0802[a]	0.3736±0.0636[bc]	0.4243±0.0758[b]	0.3180±0.0635[c]
Dupe	0.1438±0.0322[a]	0.0897±0.0181[b]	0.1314±0.0554[ab]	0.1160±0.0374[ab]
Bkg	0.5760±0.0640[a]	0.4972±0.0529[a]	0.5313±0.0462[a]	0.5679±0.0591[a]
Miss	0.3440±0.2267[a]	0.1108±0.0953[ab]	0.0497±0.1110[b]	0.1852±0.2001[ab]

Data are presented as Mean ± SD. Different letters (e.g., a, b, c) indicate statistically significant differences ($p < 0.05$) determined by one-way ANOVA followed by Tukey's HSD post-hoc test using Statgraphics Centurion XV.

3.3 Overall comparison of the four models

The experimental results on the banana dataset indicate a clear evolutionary trajectory across YOLO generations. Contrary to the common assumption that legacy "tiny" models offer the best efficiency, our findings demonstrate that modern architectures (YOLOv8n, YOLO11n) achieve superior performance in terms of accuracy, resource efficiency, and environmental impact.

Table 4. Multi-dimensional sustainability assessment of the proposed AI solution for post-harvest banana management.

Factor	Cost / Negative Impact	Benefit / Positive Impact	Supporting Evidence
Environmental	($E_{training} + E_{inference}$): Carbon emissions generated during AI model training and inference. (Training YOLOv8n/11n emits approx 0.081–0.086 kg CO_2e).	($E_{avoided}$): Emissions avoided through reduced waste. 1 kg of spoiled bananas generates approx 0.5–1.1 kg CO_2e via production and decomposition [12].	Preventing the waste of a single kilogram of bananas is sufficient to offset the entire training emissions of one modern YOLO model [10].

Economic	Hardware costs (Raspberry Pi approx $80, Jetson Nano approx $150) and operational electricity costs.	Agricultural value preserved. Vietnam's banana exports reached approx USD 372 million in 2024 [3]. A 1% reduction in post-harvest loss can save millions of USD.	Current post-harvest losses in Vietnam are 20–25%, indicating substantial potential for financial recovery [10].
Social	Requirement for digital-skill training for farmers (initial adoption barrier).	Enhances agricultural value, increases farmer income, and aligns with the National Digital Transformation Program (Decision 749/QĐ-TTg) [34, 35].	Supports sustainable rural economic development and modernization targets.

YOLO11n achieved the highest detection accuracy (mAP50–95 = 0.7112 ± 0.0028), maintaining a compact parameter count (2.59M). However, this precision comes with a trade-off in speed (100.65 ± 3.22 FPS). YOLOv8n emerged as the most balanced candidate for sustainable agriculture, providing an optimal compromise: it rivals the legacy YOLOv3-tiny in speed (138.87 ± 4.57 FPS vs. 140.82 ± 2.15 FPS) but consumes significantly less energy. Specifically, YOLOv8n recorded the lowest energy consumption (0.1711 ± 0.0001 kWh) and CO_2 emissions (0.0814 ± 0.0000 kg).

Notably, YOLOv3-tiny, despite its "tiny" designation, exhibits the largest number of parameters (12.14M) and the highest energy consumption (0.2036 ± 0.0005 kWh) while delivering the lowest accuracy (0.6418 ± 0.0044). This reflects the structural constraints of legacy anchor-based architectures compared to the optimized anchor-free mechanisms and C2f/C3k2 blocks found in newer iterations [14, 33].

Overall, AI-enabled detection systems contribute not only to productivity gains but also to greenhouse-gas mitigation. The analysis confirms that YOLOv8n offers the most "Green AI" compliant solution, balancing high throughput with minimal carbon footprint, while YOLO11n remains the superior choice for precision-critical applications. Crucially, the observed training behavior challenges the assumption that higher model complexity inherently leads to more reliable post-harvest grading outcomes.

Table 5. Summary of statistical benchmark results comparing detection performance, computational efficiency, and environmental impact across YOLO architectures.

Model	mAP50-95	Parameters (M)	FPS	Energy (kWh)	CO_2 Emission (Kg)
YOLOv3-tiny	0.6418 ± 0.0044^{c}	12.14	140.82 ± 2.15^{a}	0.2036 ± 0.0005^{a}	0.0968 ± 0.0002^{a}
YOLOv5n	0.7017 ± 0.0035^{b}	2.51	119.56 ± 10.30^{b}	0.1723 ± 0.0001^{c}	0.0819 ± 0.0001^{c}
YOLOv8n	0.7084 ± 0.0021^{a}	3.01	138.87 ± 4.57^{a}	0.1711 ± 0.0001^{d}	0.0814 ± 0.0000^{d}
YOLO11n	0.7112 ± 0.0028^{a}	2.59	100.65 ± 3.22^{c}	0.1820 ± 0.0006^{b}	0.0865 ± 0.0003^{b}

Data are presented as Mean ± SD. Different letters (e.g., a, b, c) indicate statistically significant differences ($p < 0.05$) determined by one-way ANOVA followed by Tukey's HSD post-hoc test using Statgraphics Centurion XV.

4 CONCLUSION

This study evaluates the trade-off between detection accuracy and environmental efficiency in AI-based banana quality grading by benchmarking four lightweight YOLO architectures. Challenging the assumption that compact models are inherently efficient, the results demonstrate that architectural design is more critical than parameter size alone. Specifically, the legacy YOLOv3-tiny exhibited the highest energy consumption and lowest accuracy, whereas modern nano-scale models achieved an improved balance. YOLOv8n emerged as the most balanced solution, offering minimal energy use and high processing speed, while YOLO11n delivered the highest accuracy for applications requiring strict quality control. Furthermore, error analysis revealed that detection failures primarily stemmed from localization difficulties caused by the banana's curved geometry and surface degradation, rather than fundamental classification confusion. From a sustainability perspective, empirical energy measurement is essential, though the carbon footprint of training these models remains minimal compared to the significant environmental benefits of reducing post-harvest food waste. It is important to emphasize that the current energy evaluations are strictly limited to the training phase. Because the initial experiments utilized a single dataset under controlled hardware conditions, future research must validate these models in diverse field environments and directly measure inference-phase energy consumption on embedded edge platforms. Ultimately, by highlighting the need for geometry-aware detection methods, this benchmark provides a comprehensive foundation for deploying Green AI solutions in sustainable agriculture.

REFERENCES

[1] X. Min, Y. Ye, S. Xiong, and X. Chen, "Computer Vision Meets Generative Models in Agriculture: Technological Advances, Challenges and Opportunities," *Applied Sciences*, vol. 15, no. 14, p. 7663, 2025.
[2] Z. Ali, A. Muhammad, N. Lee, M. Waqar, and S. W. Lee, "Artificial Intelligence for sustainable agriculture: a comprehensive review of AI-driven technologies in crop production," *Sustainability*, vol. 17, no. 5, p. 2281, 2025.
[3] Vietnam News Agency. (2025). *Việt Nam targets $1 billion in banana exports*. Available: https://vietnamnews.vn/economy/1731834/viet-nam-targets-1-billion-in-banana-exports.html

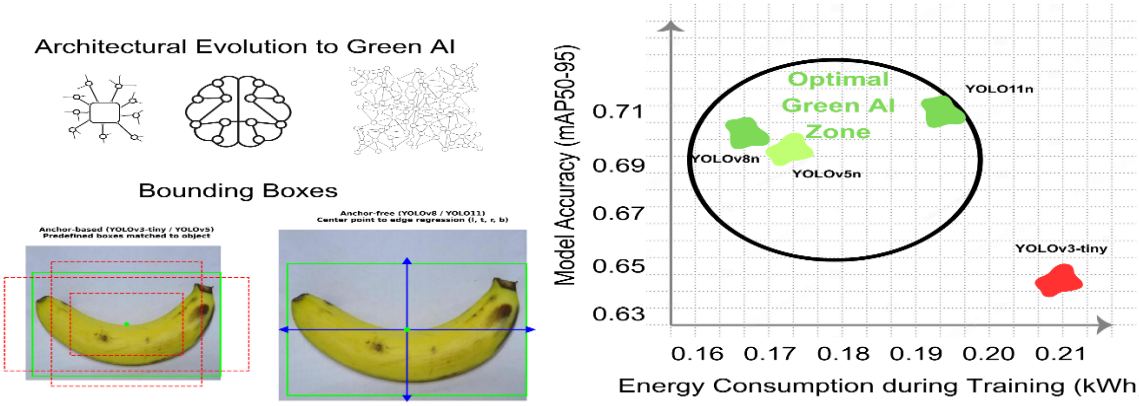


Figure 4. Architectural evolution and performance benchmarking of YOLO models. The left panel illustrates the transition from legacy anchor-based mechanisms to modern anchor-free architectures. The right panel presents the trade-off between localization accuracy (mAP50-95) and training energy consumption (kWh), highlighting the optimal Green AI zone.

[4] A. O. Saltık, A. Allmendinger, and A. Stein, "Comparative analysis of yolov9, yolov10 and rt-detr for real-time weed detection," in *European Conference on Computer Vision*, 2024, pp. 177-193: Springer.
[5] Z. Khan, Y. Shen, and H. Liu, "ObjectDetection in Agriculture: A Comprehensive Review of Methods, Applications, Challenges, and Future Directions," *Agriculture*, vol. 15, no. 13, p. 1351, 2025.
[6] H. Joshi, "Edge-AI for agriculture: lightweight vision models for disease detection in resource-limited settings," *arXiv:2412.18635*, 2024.

[7] J. Kim, G. Kim, R. Yoshitoshi, and K. Tokuda, "Real-time object detection for edge computing-based agricultural automation: A case study comparing the YOLOX and YOLOv12 architectures and their performance in potato harvesting systems," *Sensors*, vol. 25, no. 15, p. 4586, 2025.
[8] C. S. Kouzinopoulos and Y. Manna, "Hardware-Aware YOLO Compression for Low-Power Edge AI on STM32U5 for Weeds Detection in Digital Agriculture," *arXiv:2511.07990*, 2025.
[9] D. Patterson *et al.*, "Carbon emissions and large neural network training," *arXiv:2104.10350*, 2021.
[10] (2022). *150/QĐ-TTg, Approval of the Strategy for Sustainable Agricultural and Rural Development for the period 2021-2030, with a vision to 2050.*
[11] N. Suppen-Reynaga, A. B. Guerrero, E. R. Dominguez, E. Sacayón, and A. Solano, "Life cycle assessment of bananas, melons, and watermelons from Costa Rica," *Cleaner Circular Bioeconomy*, vol. 9, p. 100120, 2024.
[12] E. Svanes and A. K. Aronsson, "Carbon footprint of a Cavendish banana supply chain," *The International Journal of Life Cycle Assessment*, vol. 18, no. 8, pp. 1450-1464, 2013.
[13] R. Vinuesa *et al.*, "The role of artificial intelligence in achieving the Sustainable Development Goals," *Nature communications*, vol. 11, no. 1, p. 233, 2020.
[14] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," ed: Ultralytics, 2023.
[15] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv:2004.10934*, 2020.
[16] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the carbon emissions of machine learning," *arXiv:1910.09700*, 2019.
[17] CodeCarbon, "Methodology ", ed, 2024.
[18] D. Bolya, S. Foley, J. Hays, and J. Hoffman, "Tide: A general toolbox for identifying object detection errors," *European Conference on Computer Vision*, pp. 558-573, 2020.
[19] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv:1804.02767*, 2018.
[20] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv:2107.08430*, 2021.
[21] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 390-391, 2020.
[22] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7464-7475, 2023.
[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," *Proceedings of the European conference on computer vision (ECCV)*, pp. 3-19, 2018.
[24] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "GREEN AI," *Communications of the ACM*, vol. 63, no. 12, pp. 54-63, 2020.
[25] A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," *Journal of King Saud University-Computer Information Sciences*, vol. 33, no. 3, pp. 243-257, 2021.
[26] L. Liu *et al.*, "Deep learning for generic object detection: A survey," *International journal of computer vision*, vol. 128, no. 2, pp. 261-318, 2020.
[27] J. Dai *et al.*, "Deformable convolutional networks," *Proceedings of the IEEE international conference on computer vision*, pp. 764-773, 2017.
[28] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 658-666, 2019.
[29] A. Araujo, W. Norris, and J. Sim, "Computing receptive fields of convolutional neural networks," *Distill*, vol. 4, no. 11, p. e21, 2019.
[30] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning* (no. 2). MIT press Cambridge, 2016.
[31] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in *European conference on computer vision*, 2012, pp. 340-353: Springer.
[32] H. Zhang, Y. Wang, F. Dayoub, and N. Sunderhauf, "Varifocalnet: An iou-aware dense object detector," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8514-8523, 2021.
[33] G. Jocher and J. Qiu, "Ultralytics YOLO11," ed: Ultralytics, 2024.
[34] (2020). *749/QĐ-TTg, Decision 749/QD-TTg approves the "National Digital Transformation Program until 2025, with orientation to 2030".* Available: https://chinhphu.vn/default.aspx?pageid=27160&docid=200163
[35] (2022). *431/QĐ-TTg, Decision 431/QD-TTg approves the Project for the Development of Key Fruit Trees until 2025 and 2030.* Available: https://vanban.chinhphu.vn/?pageid=27160&docid=200163

Received on August 20 – 2026
Revised on October 08 – 2026