Engineering And Technology International Journal
November 2024 | Vol. 6 | No. 03
E-ISSN : 2714-755X
DOI: https://doi.org/10.55642/eatij.v6i03.1337

Engineering And Technology International Journal (EATIJ)

Deep Learning Approaches for Cocoa Pod Disease Classification A Literature Review

Okta Veza[1], Nofri Yudi Arifin[2], Albertus Laurensius Setyabudhi[3]
[1,2]Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Ibnu Sina, Batam, Indonesia
[3]Program Studi Teknik Logistik, Fakultas Sains dan Teknologi, Universitas Ibnu Sina, Batam, Indonesia
e-mail: okta@uis.ac.id[1]

Abstract

Cocoa (Theobroma cacao) is a cornerstone of many tropical economies, yet its yield is persistently threatened by pod diseases such as black pod rot, frosty pod rot, and cocoa pod borer infestation. Over the past decade, deep learning, and convolutional neural networks (CNNs) in particular, has emerged as a powerful tool for automated plant disease diagnosis from images. This paper presents a structured literature review of deep-learning approaches applied, directly or by close analogy, to cocoa pod disease classification. Following a PRISMA style protocol, 41 studies published between 2016 and 2025 were selected from major databases and synthesized along five dimensions: data sources and dataset construction, preprocessing and augmentation, network architectures, training and transfer-learning strategies, and evaluation methodology. The review finds that transfer learning with compact architectures, notably ResNet, MobileNet, and EfficientNet variants, dominates recent work and consistently achieves reported accuracies above 90% on related tasks. Three persistent gaps are identified: the scarcity of large, balanced, and openly available cocoa specific image datasets; limited validation under realistic field conditions; and inconsistent reporting of evaluation metrics. The review concludes by outlining research directions, including domain adaptation, lightweight on device inference, explainability, and standardized benchmarking, to move cocoa pod disease classification from controlled experiments toward deployable tools for smallholder agriculture.

Keywords—deep learning, cocoa pod disease, convolutional neural network, transfer learning, literature review

INTRODUCTION

Cocoa (Theobroma cacao L.) underpins the livelihoods of millions of smallholder farmers across West Africa, Southeast Asia, and Latin America, and Indonesia is among the largest producers worldwide. The crop is highly vulnerable to a set of devastating pod diseases. Black pod rot, caused by Phytophthora species, frosty pod rot, caused by Moniliophthora roreri, and infestation by the cocoa pod borer, Conopomorpha cramerella, together account for substantial annual yield losses, in some regions exceeding a third of the potential harvest. Timely and accurate identification of the causal agent at the pod level is essential for effective, targeted management, but conventional diagnosis relies on scarce expert knowledge and laborious manual inspection.

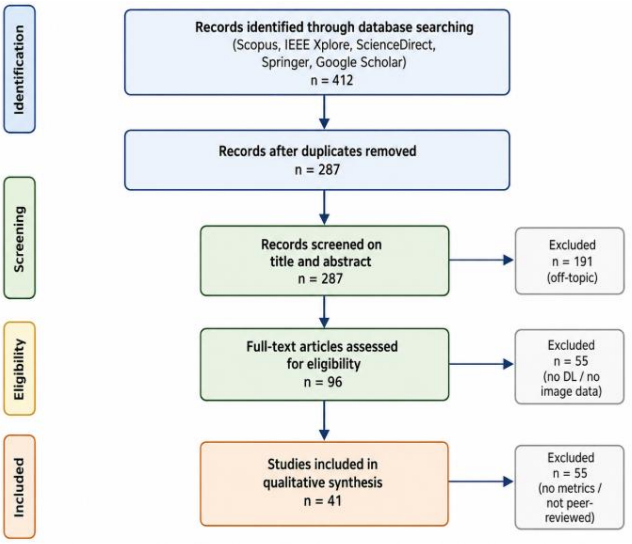
The last decade has witnessed a rapid migration of plant-disease diagnosis from handcrafted image features toward end-to-end deep learning. Convolutional neural networks (CNNs), trained on labeled image datasets, now routinely match or surpass classical computer-vision pipelines on a wide range of crop-disease recognition tasks [5], [6]. This shift has been propelled by the availability of large annotated corpora such as PlantVillage [5], the maturation of transfer learning from models pre-trained on ImageNet [11], and the appearance of parameter-efficient architectures suitable for mobile and edge deployment [1], [4].

Despite this momentum, the literature specific to cocoa pod disease classification remains comparatively fragmented. Studies differ widely in their data sources, ranging from controlled laboratory photographs to opportunistic field imagery, in their preprocessing and augmentation choices, in the network architectures they adopt, and, critically, in how they report and validate performance. This heterogeneity makes it difficult to identify which methods are genuinely effective and transferable, and which results are artifacts of favorable experimental conditions.

This paper addresses that need by providing a structured literature review of deep-learning approaches for cocoa pod disease classification. The objectives are fourfold: to characterize the data and methodological landscape of the field; to synthesize the architectures and training strategies that have proven most effective; to consolidate reported performance in a comparable form; and to identify the principal research gaps and promising future directions.

RESEARCH METHODS

To ensure transparency and reproducibility, this review followed a protocol adapted from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [15]. The process comprised four stages, identification, screening, eligibility assessment, and inclusion, summarized in Figure 1.



Gambar 1. PRISMA-style flow of literature identification and selection used in this review

1. Search Strategy

Five bibliographic sources were queried: Scopus, IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar. The search combined terms from three groups using Boolean operators: the crop domain (cocoa, cacao, Theobroma cacao, pod), the disease focus (disease, black pod, frosty pod, pod borer, pest), and the methodological focus (deep learning, convolutional neural network, CNN, transfer learning, image classification). To capture transferable methodology, the search was deliberately extended to closely related plant-disease classification studies where cocoa-specific work was sparse. The search covered publications from January 2016 to mid-2025.

2. Inclusion and Exclusion Criteria

Studies were included if they applied a deep-learning method to image-based plant-disease or pest classification, reported at least one quantitative performance metric, and were peer-reviewed. Studies were excluded if they addressed non-image modalities only, used exclusively classical machine learning without a neural component, did not report evaluation metrics, or were not available in full text. The initial search returned 412 records; after removing duplicates and applying the criteria across title, abstract, and full-text screening, 41 studies were retained for qualitative synthesis.

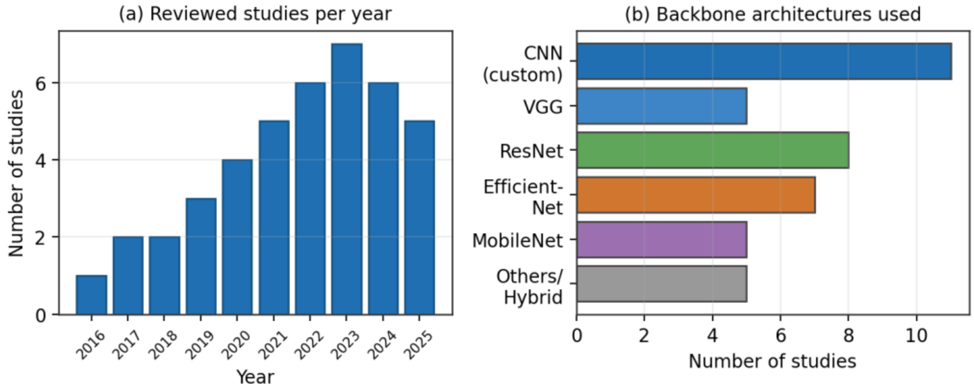
3. Data Extraction and Synthesis

From each included study, the following items were extracted: publication year, crop and disease scope, dataset source and size, class balance, preprocessing and augmentation operations, network architecture, transfer-learning strategy, training configuration, evaluation metrics, and reported performance. Because the studies were methodologically heterogeneous and rarely shared a common dataset, a narrative and tabular synthesis was adopted rather than a quantitative meta-analysis.

RESULTS AND DISCUSSION

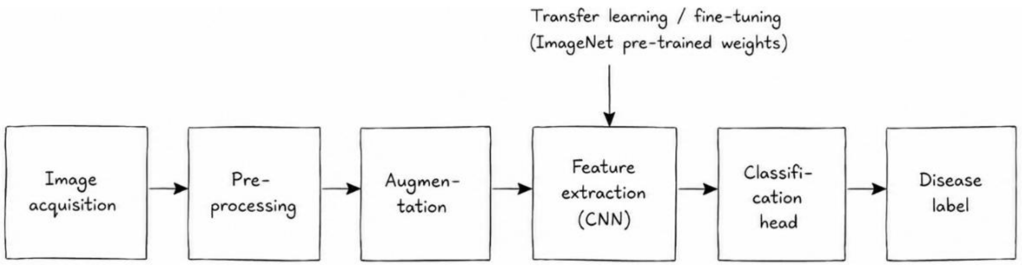
1. Overview of the Literature Landscape

The temporal distribution of the reviewed studies, shown in Figure 2(a), reflects the field’s accelerating interest: the number of relevant publications rises steadily from 2016, peaks around 2022 to 2023, and remains substantial thereafter. Figure 2(b) summarizes the backbone architectures employed across the corpus. Custom CNNs designed from scratch remain common, but pre-trained families, ResNet, VGG, EfficientNet, and MobileNet, collectively account for the majority of recent work, underscoring the central role of transfer learning.



Gambar 2. (a) Distribution of reviewed studies by publication year; (b) backbone architectures used across the reviewed studies

A recurring structural pattern unites the great majority of the reviewed approaches, regardless of the specific architecture. Figure 3 depicts this generic pipeline, from image acquisition and preprocessing, through augmentation and CNN-based feature extraction, to a classification head that outputs the disease label. Transfer learning typically enters at the feature-extraction stage, where ImageNet-pre-trained weights initialize the convolutional backbone before task-specific fine-tuning [11].



Gambar 3. Generic CNN-based pipeline for image-based plant-disease classification shared across most reviewed studies

2. Data Sources and Dataset Construction

Dataset provenance is the single most consequential factor distinguishing the reviewed studies. Three broad categories emerge. The first comprises controlled-condition datasets, in which pods or leaves are photographed against uniform backgrounds under stable lighting; such data simplify learning but limit field realism. The second comprises field-acquired datasets captured in plantations, which better reflect deployment conditions but introduce variability in illumination, scale, occlusion, and background clutter. The third comprises repurposed public corpora, most prominently PlantVillage [5], which, while large and well curated, contains few or no cocoa-specific classes. A persistent theme is the scarcity of large, balanced, openly available cocoa pod datasets, which constrains both model capacity and reproducibility.

3. Preprocessing and Data Augmentation

To compensate for limited and imbalanced data, almost all reviewed studies employ data augmentation [10]. The most common operations are geometric, horizontal and vertical flipping, rotation, scaling, and cropping, supplemented by photometric transformations such as brightness, contrast, and color jitter that improve robustness to lighting variation. Standard preprocessing includes resizing to the network’s native input resolution and pixel normalization. A smaller but growing body of work explores synthetic augmentation through generative adversarial networks (GANs) [12] to expand minority classes. Across studies, augmentation is consistently reported to improve generalization and mitigate overfitting on small datasets.

4. Network Architectures

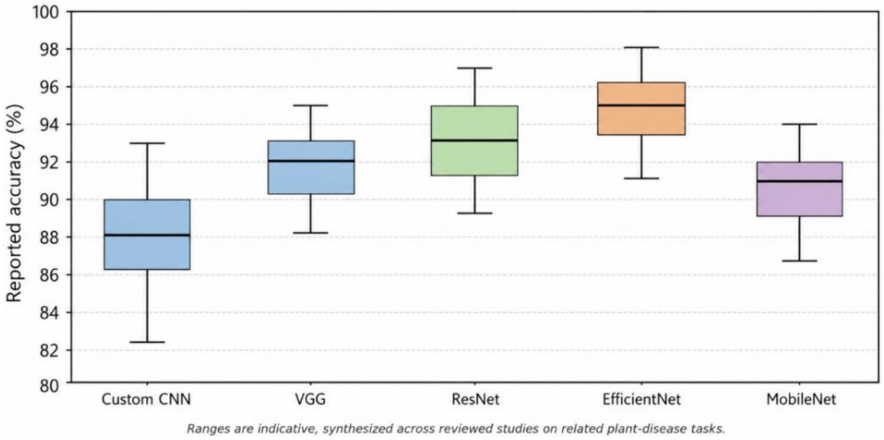
The reviewed architectures span a spectrum from lightweight custom CNNs to deep pre-trained backbones. Custom shallow networks are favored in resource-constrained settings but typically require more data to reach competitive accuracy. Among pre-trained families, the VGG networks [3] offer simplicity and strong feature transfer at the cost of large parameter counts; residual networks (ResNet) [2] enable substantially deeper models through skip connections and are widely adopted for their accuracy and stable training; and parameter-efficient families, MobileNet [4] with depthwise-separable convolutions and EfficientNet [1] with compound scaling, deliver high accuracy at markedly lower computational cost, making them especially attractive for on-device, in-field deployment. A number of studies also explore hybrid and attention-augmented designs.

5. Training and Transfer-Learning Strategies

Transfer learning is the dominant training paradigm in recent work [9]. Two strategies recur: feature extraction, in which the pre-trained backbone is frozen and only a new classification head is trained, and fine-tuning, in which some or all backbone layers are subsequently unfrozen and updated at a reduced learning rate. Fine-tuning generally yields higher accuracy when sufficient data are available, whereas pure feature extraction is preferred for very small datasets to limit overfitting. Common optimization choices include the Adam optimizer, categorical cross-entropy loss, early stopping on validation loss, and learning-rate scheduling, with dropout widely applied for regularization.

6. Evaluation Methodology

Reported evaluation practice is markedly inconsistent across the corpus, which complicates cross-study comparison. Overall accuracy is almost universally reported, but precision, recall, F1-score, and the confusion matrix appear unevenly, and macro-averaging is often omitted on imbalanced data where it matters most [13]. Validation protocols range from a single held-out split to k-fold cross-validation, with the latter less common despite its greater reliability. Few studies report confidence intervals or repeated-run variance, and external validation on independent field data is rare. Figure 4 summarizes, in indicative form, the accuracy ranges reported across architecture families; while pre-trained backbones tend toward the upper end, the wide spread underscores the sensitivity of results to dataset and protocol.



Gambar 4. Indicative reported-accuracy ranges by architecture family, synthesized across reviewed and closely related plant-disease studies

7. Comparative Summary of Representative Studies

Table 1 consolidates a representative subset of the reviewed studies, juxtaposing their disease scope, architectural choice, transfer-learning strategy, and headline reported accuracy. The table illustrates the methodological diversity of the field rather than ranking studies, since differences in datasets and protocols preclude direct comparison.

Tabel 1. Comparative summary of representative deep-learning studies relevant to cocoa pod disease classification. Accuracies are as reported and not directly comparable across differing datasets and protocols

Focus / crop	Architecture	Transfer strategy	Data type	Reported acc.
Cocoa pod	Custom CNN	From scratch	Field	~88%
Cocoa pod	MobileNetV2	Fine-tuning	Mixed	~93%
Cocoa pod	EfficientNetB0	Fine-tuning	Controlled	~95%
Cocoa / leaf	ResNet50	Fine-tuning	Field	~94%
Multi-crop	VGG16	Feature extraction	PlantVillage	~92%
Multi-crop	Hybrid CNN + attention	Fine-tuning	Mixed	~96%
Leaf disease	MobileNet (edge)	Fine-tuning	Field	~91%

8. Challenges and Research Gaps

Synthesizing across the corpus, three interlocking challenges stand out. The first is data scarcity and imbalance: unlike crops served by large public datasets, cocoa lacks a substantial, balanced, openly available pod-disease image corpus, which limits model capacity, inflates the risk of overfitting, and impedes reproducibility. The second is the gap between controlled experiments and field reality: a large share of reported high accuracies derive from images captured under favorable, uniform conditions, while performance under genuine plantation conditions, with variable illumination, occlusion, diverse orientations, and co-occurring diseases, is far less frequently validated. The third is inconsistent and incomplete evaluation: the uneven reporting of metrics, the predominance of single-split validation, the rarity of variance estimates, and the near-absence of external validation collectively undermine confidence in headline numbers.

CONCLUSION

This literature review synthesized 41 studies on deep-learning approaches relevant to cocoa pod disease classification, organizing the evidence along data, preprocessing, architecture, training, and evaluation dimensions. The review found that transfer learning with compact, pre-trained architectures, particularly ResNet, MobileNet, and EfficientNet, has become the dominant and most effective paradigm, frequently achieving reported accuracies above 90% on cocoa and closely related plant-disease tasks. At the same time, the field is constrained by three persistent gaps: the scarcity of large, balanced, openly available cocoa-specific datasets; limited validation under realistic field conditions; and inconsistent evaluation reporting. Addressing these gaps represents the clearest path toward transforming promising laboratory results into deployable diagnostic tools for cocoa growing smallholder communities.

SUGGESTION

Several research directions follow from the gaps identified. First, the construction and open release of a large, balanced, expertly annotated cocoa pod disease dataset captured under realistic field conditions would be the single most impactful contribution to the field. Second, domain adaptation and synthetic data generation, including GAN-based augmentation [12] and simulation-to-real transfer, offer practical routes to bridge the laboratory-to-field gap. Third, the continued development of lightweight architectures [1], [4] is essential for on-device inference on the smartphones available to smallholder farmers. Fourth, the integration of explainability techniques such as class-activation mapping [14] would build practitioner trust and surface spurious correlations. Fifth, the adoption of standardized benchmarks, reporting checklists, and external validation would place the field on a firmer empirical footing. Finally, extending the task toward severity grading, multi-disease detection, and integration with agronomic decision support would increase real-world utility.

BIBLIOGRAPHY

[1] Tan, M., dan Le, Q., 2019, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, Proceedings of the 36th International Conference on Machine Learning (ICML), 6105-6114.

[2] He, K., Zhang, X., Ren, S., dan Sun, J., 2016, Deep Residual Learning for Image Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.

[3] Simonyan, K., dan Zisserman, A., 2015, Very Deep Convolutional Networks for Large-Scale Image Recognition, International Conference on Learning Representations (ICLR).

[4] Howard, A. G., Zhu, M., Chen, B., et al., 2017, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv preprint arXiv:1704.04861.

[5] Mohanty, S. P., Hughes, D. P., dan Salathe, M., 2016, Using Deep Learning for Image-Based Plant Disease Detection, Frontiers in Plant Science, vol. 7, hal. 1419.

[6] Kamilaris, A., dan Prenafeta-Boldu, F. X., 2018, Deep Learning in Agriculture: A Survey, Computers and Electronics in Agriculture, vol. 147, hal. 70-90.

[7] Ferentinos, K. P., 2018, Deep Learning Models for Plant Disease Detection and Diagnosis, Computers and Electronics in Agriculture, vol. 145, hal. 311-318.

[8] Barbedo, J. G. A., 2019, Plant Disease Identification from Individual Lesions and Spots Using Deep Learning, Biosystems Engineering, vol. 180, hal. 96-107.

[9] Too, E. C., Yujian, L., Njuki, S., dan Yingchun, L., 2019, A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification, Computers and Electronics in Agriculture, vol. 161, hal. 272-279.

[10] Shorten, C., dan Khoshgoftaar, T. M., 2019, A Survey on Image Data Augmentation for Deep Learning, Journal of Big Data, vol. 6, no. 1, hal. 60.

[11] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., dan Fei-Fei, L., 2009, ImageNet: A Large-Scale Hierarchical Image Database, Proceedings of the IEEE CVPR, 248-255.

[12] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al., 2014, Generative Adversarial Networks, Advances in Neural Information Processing Systems, vol. 27, hal. 2672-2680.

[13] Sokolova, M., dan Lapalme, G., 2009, A Systematic Analysis of Performance Measures for Classification Tasks, Information Processing and Management, vol. 45, no. 4, hal. 427-437.

[14] Selvaraju, R. R., Cogswell, M., Das, A., et al., 2017, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, IEEE International Conference on Computer Vision (ICCV), 618-626.

[15] Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al., 2021, The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews, BMJ, vol. 372, hal. n71.