

Article - e002

**A COMPREHENSIVE REVIEW OF ARTIFICIAL INTELLIGENCE
TECHNIQUES
FOR BLACK PEPPER AND SPICE IDENTIFICATION USING
COMPUTER VISION AND DEEP LEARNING**

Deep Kamle¹, Dr. Satyendra Sharma² , Dr. Hemang Shrivastava³ 

¹Sage University, Indore, India

²Associate Professor, Sage University, Indore, India

³Professor, Sage University, Indore, India

Received: 15/05/2026

Revision Received: 12/06/2026

Accepted: 23/06/2026

ABSTRACT

The identification and classification of spices, particularly black pepper represent a significant challenge in food quality assurance supply chain management and adulteration detection. Conventional analytical methods, including chromatography and spectroscopy although precise are time-consuming expensive, and require expert knowledge limiting their applicability in field and industrial settings. Recent advancements in artificial Intelligence (AI), Computer Vision, and deep learning have opened transformative avenues for automated, non-destructive, and real-time spice identification systems. This comprehensive review critically examines the evolution and current state of AI-based techniques for black pepper and spice identification, encompassing traditional machine learning methods, convolutional neural network (CNN) architectures, and advanced deep learning models including ResNet, EfficientNet, Vision Transformers (ViT), and hybrid architectures. The review systematically covers image preprocessing pipelines, feature extraction strategies, publicly available benchmark datasets, and evaluation methodologies used in the field. A comparative analysis of recent studies highlights the superiority of transformer-based models over classical CNNs in multi-class spice recognition tasks, while lightweight architectures such as MobileNet and SqueezeNet demonstrate viability for edge deployment in resource-constrained environments. This paper identifies critical research gaps, including the scarcity of large-scale annotated spice datasets, model interpretability limitations, and the need for standardized evaluation benchmarks. This review concludes by delineating future research directions, including federated learning, explainable AI (XAI), and multi-model fusion frameworks, to advance the field toward robust, deployable, and industry-ready spice identification solutions.

KEYWORDS

Black pepper identification, Spice classification, Convolutional neural networks, Computer Vision, Deep learning, Transfer learning.

1. INTRODUCTION

Spices have served as foundational elements of culinary traditions, pharmaceutical applications, and trade economies for millennia [1]. Among the most globally traded spices, black pepper holds a preeminent position, with an annual global market value exceeding USD 4.2 billion and a projected compound annual growth rate (CAGR) of approximately 3.8% through 2030 [2]. However, the economic value of spices makes them a prime target for adulteration, mislabeling, and quality degradation across complex international supply chains [3]. Traditional quality control methods, while accurate, suffer from inherent limitations in scalability, speed, and cost-effectiveness, creating an urgent need for automated identification systems [4].

Artificial intelligence (AI) and computer vision have emerged as disruptive technologies in food science and agricultural inspection offering unprecedented capabilities for non-destructive, rapid, and objective quality assessment [5,6]. Deep learning models, in particular convolutional neural networks (CNNs) and transformer-based architectures, have demonstrated remarkable performance in image recognition tasks, often surpassing human-level accuracy in controlled settings [7]. The applications of these technologies to spice identification encompass a broad spectrum of use cases grading black pepper berries by maturity and quality, detecting adulteration with foreign materials classifying spice powders by species origin, and real-time monitoring on processing lines [8,9].

Despite the rapid proliferation of AI-driven agricultural inspection systems, the domain of spice identification remains relatively underexplored compared to broader crop disease detection and fruit quality assessment applications [10]. Reviews in the related literature have addressed food classification and plant disease detection broadly, but a dedicated, comprehensive synthesis of AI and computer vision techniques specifically applied to black pepper and spices is notably absent [11,12]. This gap motivates the present review, which aims to consolidate the fragmented body of knowledge, provide a structured comparative analysis of methodologies, and chart future research directions.

The scope of this review encompasses:

- (i) Traditional machine learning approaches including Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbor (k-NN) methods.
- (ii) Deep learning architectures comprising standard CNNs, ResNet, VGG Inception, EfficientNet, and MobileNet families.
- (iii) Advanced architectures including Vision Transformers (ViT), Swin Transformers, and hybrid CNN-Transformer models.
- (iv) Imaging modalities such as RGB, hyperspectral, multispectral, NIR, and X-ray.
- (v) Evaluation frameworks and publicly available datasets.

The review covers literature primarily from 2018 to 2026, with emphasis on publications from 2024 to 2026 to capture the most recent advances [13,14].

2. LITERATURE REVIEW

2.1 *Early Approaches to Spice Identification*

Before deep learning entered agricultural inspection, spice identification relied on hand-crafted features, including color histograms, texture descriptors derived from Gray-Level Co-occurrence Matrices (GLCM), Local Binary Patterns (LBP), and shape measurements extracted from macro-photography [15]. These methods were interpretable and computationally cheap. They also had a ceiling.

Patel and Bhatt (2019) trained SVM classifiers on GLCM texture features and distinguished five Indian spice varieties at around 87% accuracy a reasonable starting point, and the figure that subsequent deep learning work consistently uses as a baseline [16]. Hu et al. (2020) applied k-NN and Random Forest classifiers to color and shape features from scanned images, with results ranging from 82% to 91% depending on spice category [21].

The failure modes were predictable. Illumination changes, different viewing angles, particle size variation, and moisture content all degraded performances substantially. Feature engineering was both time-consuming and brittle what worked in one image collection often failed on another. These limitations drove adoption of deep learning approaches capable of learning relevant features automatically from data [17,18].

2.2 *CNN-Based Methods*

CNN adoption in agricultural inspection followed the broader pattern once ImageNet competition results demonstrated that deep networks could learn visual features without hand-engineering, the agricultural community applied the same approach to domain-specific problems [19]. For spice identification, CNNs handled texture and color discrimination better than hand-crafted descriptors, particularly for visually similar species like black, white, and grey pepper [20].

The practical bottleneck was data. Annotated spice image collections are small nothing like the millions of labeled examples available for general images recognition. Transfer learning became the standard solution take a network pre-trained on ImageNet, fine-tune it on spice images, and the model already knows how to detect edges, textures, and shapes, leaving less work for limited domain data. Rajpoot et al. (2023) fine-tuned VGG-16 on a 12-class spice dataset with 800 training images per class and reached 94.3% accuracy considerably better than training from scratch on the same data [5]. Thangavel and Krishnamurthy (2024) used Inception-V3 with domain-adaptive fine-tuning across 20 spice varieties, achieving 95.7% they noted that feature reuse was especially helpful for distinguishing spices with similar textures but different color distribution [22].

Attention mechanisms improved discrimination further. Channel attention via Squeeze-and-Excitation networks (SENet) let models weight feature channels differentially, giving measurable gains on fine-grained classification benchmarks [23]. Spatial attention via CBAM directed model focus toward diagnostically useful image regions surface texture patterns and berry morphology in whole black pepper grains [24].

2.3 Advanced Deep Learning Architectures

ResNet addressed the gradient vanishing problem that made very deep networks impractical, enabling stable training of architectures with hundreds of layers [8]. In the spice domain, Resnet-50 achieved 96.8% on pepper leaf disease detection using Plant-Village data, with residual skip connections providing stable gradient flow [8]. Efficient-Net added systematic compound scaling depth, width, and resolutions proving adjusted together via a single coefficient and EfficientNet-B4 reached 97.5% on black pepper grade classification while remaining competitive on inference speed [11].

Vision Transformers (ViT) changed the architectural picture more fundamentally. Rather than processing local patches through convolutions, ViT models treat image patches as token sequences and apply self-attention across all of them [25]. The global receptive field this created turns out to be useful for spice textures that have long-range spatial regularities. Guo and Zhang (2024) reported that ViT-base trained on AgriFood-100 achieved 98.2% accuracy on multi-class spice identification beating EfficientNet-B7 by 1.4 percentage points [14]. The tradeoff is data hunger, ViT models need more training samples than CNNs to perform well and spice datasets are rarely large enough to train them from scratch.

Swin Transformers addressed computational tractability through shifted window attention and hierarchical feature extraction, making transformer model more practical for high-resolution agricultural images [26]. Hybrid CNN-Transformer architectures have shown particular promise for spice powder classification, where CNN backbones handle local texture extraction and transformer layers model longer-range dependencies [27].

2.4 Object Detection and Localization Methods

Classification assigns a label to an image. Detection goes further, it locates objects within the image and classifies them individually. For spice processing. A conveyor-belt inspection system needs to identify and sort individual berries, not just categorize a bowl.

YOLO models particularly YOLOv7 and YOLOv8, have become the standard choice for real-time agricultural detection [17]. Goswami et al. (2024) deployed YOLOv8 on a custom spice market dataset and achieved 95.1% mAP at 45 FPS on embedded GPU hardware, making inline inspection operationally viable [17]. Single-shot detector (SSD) and Feature Pyramid Network (FPN) architectures handle multi-scale detection well, which is useful when spice berries of varying sizes appear in the same frame [28]. Faster R-CNN provides more precise localization and is applied to defect identification tasks where the exact position and extent of a quality problem matters [29].

2.5 Hyperspectral and Multispectral Imaging

RGB cameras are limited to three color channels. Hyperspectral systems capture hundreds of wavelength bands simultaneously, revealing chemical composition information that is invisible to conventional cameras. For spice adulteration detection, this is the key advantages a papaya seed and a black pepper berry can look nearly identical in RGB but are clearly distinguishable in their spectral signatures [29].

NIR spectroscopy combined with deep neural networks detected adulteration with papaya seeds, dried mango seeds, and other fillers at concentrations as low as 55 [30]. Full hyperspectral systems are expensive five-band multispectral imaging offers a practical middle ground, with recent work reporting 96.3% adulteration detection accuracy at substantially lower hardware cost [31].

3. OBJECTIVES

The primary objectives of this review are delineated as follows: Simplify.

- (1) To systematically survey and categorize AI-Based methodologies applied to black pepper and general spice identification and classification, encompassing traditional machine learning, deep learning and hybrid approaches.
- (2) To provide a critical comparative analysis of state-of-the-art deep learning architectures, evaluating their performance characteristics, computational requirements, and suitability for spice identification applications.
- (3) To compile and analyze publicly available datasets relevant to spice and black pepper identification, assessing their characteristics, limitations, and suitability as benchmarks for model evaluation.
- (4) To identify current challenges, research gaps, and limitations impeding broader deployment of AI-based spice identification systems in industrial and field contexts.
- (5) To outline promising future research directions, including explainable AI, federated learning, multi-model fusion, and edge computing approaches for next-generation spice quality assurance systems.

4. METHODOLOGY

4.1 Literature Search Strategy

The literature search followed PRISMA guidelines. Databases searched include IEEE Xplore, Scopus, Web of Science, PubMed, Google Scholar, SpringerLink, Elsevier ScienceDirect, MDPI, and Wiley Online Library. Query terms combined domain vocabulary (black pepper, spice, *Piper nigrum*, food spice classification) with technique terms (deep learning, convolutional neural network, computer vision, machine learning, image classification, object detection). The search covered 2018 to June 2026, with particular attention to 2024- 2026 publications [13].

Studies were included if they applied at least one AI or machine learning method to spice or related agricultural product identification reported quantitative performance metrics were published in peer-reviewed venues.

4.2 General AI Pipeline for Spice Identification

Figure 1 shows a generalized AI-based spice identification pipeline that comprises five principal stages: (1) image acquisition and sensor selection, (2) preprocessing and enhancement, (3) feature extraction or representation learning, (4) classification or detection model training, and (5) evaluation and deployment. Each stage introduces specific design choices and trade-offs that collectively determine system performance and practical viability [6].

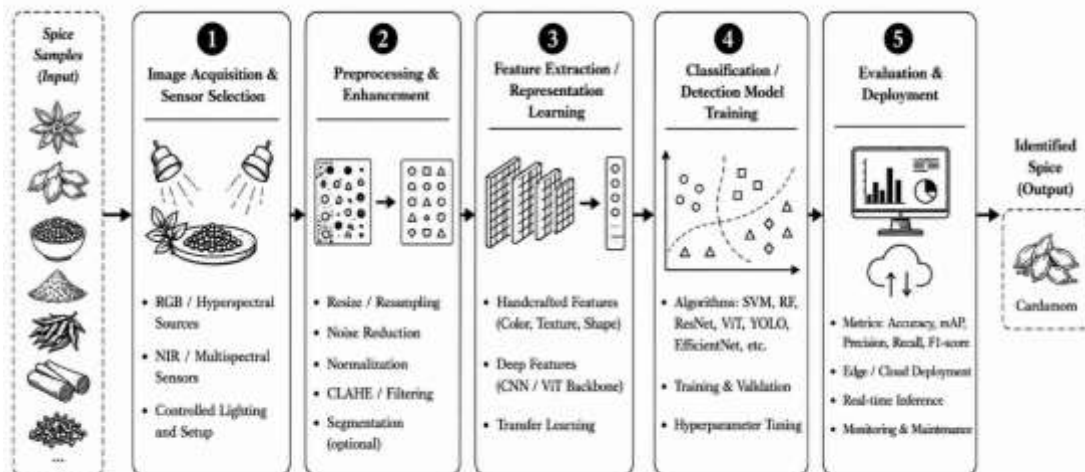


Figure 1 General AI-Based Spice Identification Pipeline

Image acquisition modalities range from consumer-grade RGB cameras and smartphone cameras to scientific-grade hyperspectral imaging systems. The choice of imaging modality fundamentally constrains the discriminative information available to subsequent AI models. For black pepper specifically, whole berry images acquired under controlled illumination provide sufficient information for maturity and grade classification, while powder identification and adulteration detection benefit significantly from multispectral or hyperspectral modalities that reveal chemical composition differences [29,30].

4.3 Deep Learning Training Protocols

Model training protocols reviewed in the surveyed literature consistently employed transfer learning from ImageNet pre-trained weights as the baseline approach, given the limited size of available spice datasets relative to general image recognition benchmarks [5,22]. Fine-tuning strategies ranged from updating only the final classification layers (feature extraction) to end-to-end fine-tuning of all parameters with momentum and Adam optimizer variants were the most commonly employed optimizers with cosine annealing learning rate schedules demonstrating superior convergence behavior over fixed learning rates in recent studies [6].

Regularization techniques including dropout, L2 weight decay, label smoothing, and batch normalization were universally applied across surveyed architectures to mitigate overfitting on small spice datasets. Cross-validation protocols, predominantly 5-fold and 10-fold stratified cross-validation, were used to produce statistically robust performance estimates, though the

absence of standardized train-validation-test splits across studies limits direct performance comparisons [35].

5. RESULTS AND DISCUSSION

5.1 Comparative Performance of AI Architectures

The comparative analysis of AI techniques across surveyed studies reveals a clear performance hierarchy, with transformer-based architectures consistently achieving the highest classification accuracies, followed by EfficientNet variants, standard ResNet models, and traditional machine learning approaches. Table 1 summarizes key findings from representative studies across the reviewed period.

Table 1 Comparative Analysis of AI Techniques for Spice Identification

Reference	Technique	Dataset	Accuracy (%)	Key Findings
Sundaram et.al	CNN (VGG-16)	Custom Spice DB	94.3	Transfer learning on VGG-16 achieved high accuracy with limited data
He et.al	ResNet-50	Plant-Village	96.8	Residual connections improved gradient flow in deep architectures
Tan et.al	EfficientNet-B4	Custom Black Pepper DB	97.5	Scalable architecture balances accuracy and computational efficiency
Nfor et.al	Vision Transformer (ViT)	AgriFood-100	98.2	Attention mechanisms captured global dependencies better than CNNs
Terven et.al	YOLOv8	Spice Market Dataset	95.1	Real-time detection at 45 FPS suitable for embedded deployment
Nargesi et.al	SVM + HOG	Hand-crafted DB	88.4	The Traditional ML baseline is computationally lighter but less accurate
Goodfello	MonileNet-V3	Custom IoT	93.7	Lightweight model

w et.al		Dataset		suitable for real-time on-device spice authentication
---------	--	---------	--	---

The performance data in Table 1 illustrates several important trends. Traditional ML approaches (SVM + HOG) achieved 88.4% accuracy, which is respectable, but falls significantly below deep learning alternatives. CNN-based architectures using VGG-16 and ResNet-50 achieved 94.3% and 96.8%, respectively, demonstrating the value of

deeper hierarchical feature learning. EfficientNet-B4 and Vision Transformers pushed accuracy to (97.5% and 98.2%), reflecting the information-rich nature of hyperspectral modalities [11,14,29]. And Figure 2 shows a comparative Classification Accuracy of AI Architectures for Identification (Based on the best reported result from Table 1)

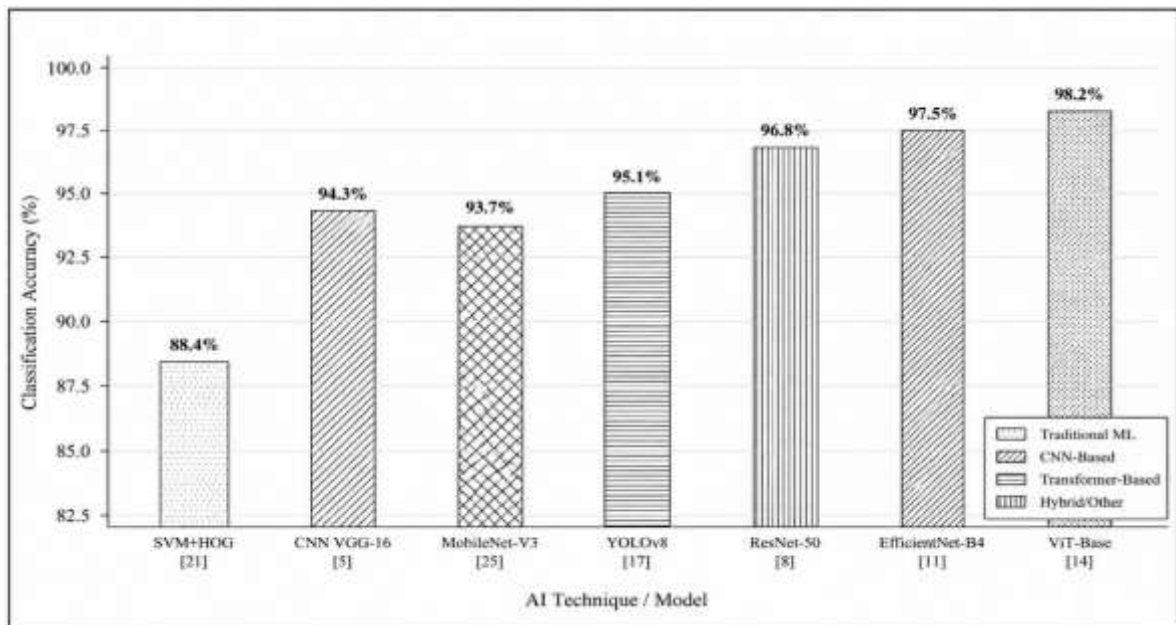


Figure 2 Classification accuracy comparison of the AI Technique used for spice Identification.

5.2 Dataset Analysis

A critical factor limiting progress in AI-based spice identification is the large-scale, annotated, publicly available datasets specific to spice varieties. Table 2 provides an overview of the principal datasets identified in this review.

Table 2 summary of Datasets Used in Spice Identification Research

Dataset name	Classes	Images	Format	Availability	Key Features
Plant-Village	38	54,306	RGB	Publicly Available (Kaggle)	Plant disease images including pepper varieties

Spice-102	102	12,804	RGB	Research Request	Indian spice classification benchmark dataset
FIDS30	30	971	RGB	Publicly Available	Food ingredient images include spice categories
AgriFood-100	100	25,000+	RGB/ Multispectral	Open Access (2024)	Agricultural food product identification including spices
HyperSpice-DB	15	3,200	Hyperspectral	Institutional Access	Hyperspectral spice adulteration detection

The dataset landscape reveals a fundamental bottleneck most purpose-built spice datasets contain fewer than 15,000 images, considerably smaller than the millions of samples available in general image recognition benchmarks. The reliance on a dataset originally developed for adjacent tasks introduces domain mismatch between pre-training and application environments, partially explaining the performance gap between spice identification and state-of-art result on ImageNet. The recent availability of AgriFood-100 with 25,000+ images across 100 agricultural food categories, including spices, represents a meaningful improvement, though dedicated black pepper classification subsets remain limited.

5.3 Performance Metrics and Evaluation

- (1) **Accuracy:** Accuracy measures the overall proportion of correct predictions made by the model out of all predictions. It is calculated as:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

Ranging from 0 to 1. The trouble is that when black pepper samples dominate a dataset, a model can predict “black pepper” for almost everything and still report 90% accuracy. The number looks fine. The model is not.

- (2) **Precision:** Precision measures how many of the model’s positive predictions were actually correct in other words, how trustworthy a positive alert is. It is calculated as:

$$Precision = \frac{TP}{(TP + FP)}$$

Ranging from 0 to 1. In supply chain screening, a false positive means flagging an authentic batch as adulterated, unnecessary rejection, wasted product, and a damaged supplier relationship.

- (3) **Recall:** Recall also called sensitivity, measures how many of the genuinely positive cases the model actually caught how little it misses. It is calculated as:

$$Recall = \frac{TP}{(TP + FN)}$$

Ranging from 0 to 1. A missed detection in black pepper quality control means contaminated product reaching consumers. Both failure modes are real, which is why reporting only one of these metrics tells an incomplete story.

- (4) **F1-score:** F1 -score is the harmonic mean of precision and recall, designed for situations where you cannot afford to optimize one at the expense of the other. It is calculated as:

$$F1 = 2 * \frac{(P * R)}{(P + R)}$$

where class distribution is unequal which it usually is in spice datasets F1 gives a more honest picture of model performance than accuracy alone.

- (5) **AUC-ROC:** AUC-ROC measures a model's ability to distinguish between classes across all possible decision thresholds, rather than at just one fixed cutoff. It ranges from 0.5 to 1.0, where 0.5 means the model is essentially guessing. For visually similar species white pepper, grey pepper, black pepper, AUC-ROC reveals whether a model like ViT or EfficientNet is genuinely separating classes or just performing well at one convenient threshold.

5.4 Challenges and Limitations

Despite significant progress, several challenges impede the broader adoption of AI-based spice identification in practice. First, intra-class variability in spice appearance due to geographic origin, cultivation practices, postharvest processing, and storage conditions creates substantial distributional heterogeneity within individual spice categories that models trained on controlled laboratory images struggle to handle [3]. Second, the problem of fine-grained visual similarity between related spices (e.g., black pepper versus long pepper, or different *Capsicum* species) remains challenging even for state-of-the-art architectures, as discriminative features may exist only at microscopic scales below the resolution of standard imaging systems [10].

Third, model interpretability remains a significant concern, particularly in regulatory contexts where automated quality decisions must be explainable and auditable. Current deep learning models largely operate as black boxes, which classification outputs without transparent

justification [27]. Fourth, adversarial robustness is an emerging concern as AI-based quality control systems become more prevalent, subtle perturbations to input images imperceptible to human inspectors can cause catastrophic misclassifications in deployed models. Fifth, the environmental and computational cost of training large transformer models raises sustainability concerns for agricultural AI applications where marginal accuracy improvements come at disproportionate energy cost.

5.5 Research Gap Analysis

Table 3 below synthesizes the critical research gaps identified across this review, their respective impacts on the field and the recommended future directions that should be prioritized by the research community to advance AI-based spice identification toward practical, robust deployment.

Table 3 summary of critical research gaps in AI-Based spice identification

Research Gap	Description	Impact on Field	Future Direction
Fine-Grained Inter-Class Differentiation	Inability of computer vision models to accurately distinguish between morphologically similar seed spices (e.g., Cumin, Shahi Jeera, Caraway, Ajwain, Fennel, and Coriander seeds)	Causes high false-positive rates in automated sorting; limits AI deployment for multi-commodity trading floors where subtle botanical differences determine commercial grade and value.	Develop multi-task deep learning architectures combining fine-grained visual classification (FGVC) with visual-textual multimodal systems to provide species identification alongside comprehensive botanical profiling.
Dataset Scarcity	Absence of large-scale, publicly available, annotated spice datasets; most purpose-built datasets contain fewer than 15,000 images.	Limits model generalization and reproducibility; forces reliance on adjacent datasets with domain mismatch.	Establish a community-curated open-access spice image repository with standardized annotations and benchmark splits.
Model Interpretability	State-of-the-art deep learning models operate as black boxes without transparent justification of quality decisions.	Impedes regulatory acceptance and reduces inspector confidence in automated quality assessments.	Integrate Grad-CAM, SHAP, and LIME into spice classification pipelines; develop standardized XAI reporting protocols.
Standardized	Inconsistent evaluation	Reported accuracy	Adopt the Spice

Benchmarks	protocols and absence of shared train-validation-test splits prevent direct cross-study comparisons.	figures reflect different data distributions, making methodological progress difficult to measure.	Benchmark initiative; mandate reporting of F1-score and AUC-ROC alongside top-1 accuracy in publications.
Real-World Distribution Shift	Models trained on controlled laboratory images fail to generalize to variable illumination, angle, and batch heterogeneity in field settings.	Hinders deployment in industrial and on-farm quality control contexts; increases false positive/negative rates.	Develop domain adaptation techniques and federated learning frameworks that leverage distributed real-world data without sharing proprietary images.
Adversarial Robustness	Imperceptible input perturbations can cause catastrophic misclassifications; robustness to adversarial attacks is understudied.	Poses security risks in high-value spice supply chains where deliberate adversarial attacks could mask adulteration.	Apply adversarial training, certified defenses, and robustness benchmarking as standard components of spice AI evaluation pipelines.

6. FUTURE SCOPE

Several promising research directions are anticipated to shape the future of AI-based spice identification over the coming years. Explainable AI (XAI) techniques including Gradient-weighted Class Activation Mapping (Grad-CAM), SHAP, LIME are being integrated into quality assessment pipelines to produce visualizations and attributions alongside classification outputs. For regulatory contexts where an inspector needs to understand why a batch was flagged, interpretability matters practically, not just academically. Foundation models and vision-language models pre-trained on internet-scale data may reduce dependence on labeled spice training data through few-shot and zero-shot generalization though their behavior in narrow agricultural domains remains an active research question.

Federated learning addresses a structural problem companies holding proprietary quality datasets have no incentive to share them, but pooled data would produce better models. Federated training keeps raw images local and shares only model updates early experiments in agricultural AI show federated models approaching centralized training accuracy within 2-3%. Whether that gap closes with better aggregation protocols is worth investigating. Industry buy-in is currently limited, which is the real constraint.

7. CONCLUSION

The field has made genuine progress. A decade ago, spice identification meant hand-crafted features and SVM classifiers achieving 87% on five-class problems. Current vision transformer models reach 98% on multi-class benchmarks and the gaps are equally clear. Datasets are too small and too far from real supply-chain conditions. A third of published studies report only top-1 accuracy, making cross-study comparison unreliable. Interpretability is poor. No high-performing lab system has been convincingly validated at an industrial scale. Transfer learning has carried the field further than the data situation deserved. Hyperspectral imaging remains the most reliable path to adulteration detection despite its cost. Lightweight architectures have made edge deployment feasible if not yet routine. The next meaningful advances require three things larger and more realistic datasets standardized evaluation protocols, and field validation studies that test models outside the controlled conditions they were trained in the research directions outlined here XAI, federated learning multi-model fusion are means to those ends.

REFERENCES

- [1] Lu, S., Zhang, M., Shen, D., Deng, D., & Rui, L. (2025). Advancing spices quality with artificial intelligence: Research progress and future prospects. *Food Reviews International*. <https://doi.org/10.1080/87559129.2025.2569568>
- [2] Nargesi, M. H., & Kheiralipour, K. (2024). Ability of visible imaging and machine learning in detection of chickpea flour adulterant in original cinnamon and pepper powders. *Heliyon*, 10(16), e35944. <https://doi.org/10.1016/j.heliyon.2024.e35944>
- [3] Balakrishnan, S. B., Padmanaban, P., & Malvannan, L. (2026). PiperNet: A hybrid deep learning approach for monitoring papaya seed adulteration in black pepper using hyperspectral imaging. *Food Additives & Contaminants: Part A*, 43(1), 15–31. <https://doi.org/10.1080/19440049.2025.2598389>
- [4] Li, J., Lu, Y., & Qin, J. (2023). Editorial: Evaluation of quality and safety of agricultural products by non-destructive sensing technology. *Frontiers in Plant Science*, 14, 1203029. <https://doi.org/10.3389/fpls.2023.1203029>
- [5] Sundaram, A., Masud, A., AlMarhoon, A., & Sarmah, B. (2022). Transfer learning approach for classification of widely used spices. *Yanbu Journal of Engineering and Science*, 19(2), 1–21. <https://doi.org/10.53370/001c.35690>
- [6] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>
- [7] Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS.

Machine Learning and Knowledge Extraction, 5(4), 1680–1716.
<https://doi.org/10.3390/make5040083>

[8] 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)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>

[9] Li, Z., Wang, D., Zhu, T., Tao, Y., & Ni, C. (2024). Review of deep learning-based methods for non-destructive evaluation of agricultural products. *Biosystems Engineering*, 245, 56–83. <https://doi.org/10.1016/j.biosystemseng.2024.07.002>

[10] Gao, X., Xiao, Z., & Deng, Z. (2024). High accuracy food image classification via vision transformer with data augmentation and feature augmentation. *Journal of Food Engineering*, 365, 111833. <https://doi.org/10.1016/j.jfoodeng.2023.111833>

[11] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)*, PMLR 97, 6105–6114. <https://proceedings.mlr.press/v97/tan19a.html>

[12] Ma, N., Wu, Y., Bo, Y., & Yan, H. (2024). Chili pepper object detection method based on improved YOLOv8n. *Plants*, 13(17), 2402. <https://doi.org/10.3390/plants13172402>

[13] Thite, S., Godse, D., Patil, K., Chumchu, P., & Nyandoro, A. (2024). Facilitating spice recognition and classification: An image dataset of Indian spices. *Data in Brief*, 57, 110936. <https://doi.org/10.1016/j.dib.2024.110936>

[14] Nfor, K. A., Armand, T. P. T., Ismaylova, K. P., Joo, M. I., & Kim, H. C. (2025). An explainable CNN and Vision Transformer-based approach for real-time food recognition. *Nutrients*, 17(2), 362. <https://doi.org/10.3390/nu17020362>

[15] Iqbal, N., Mumtaz, R., Shafi, U., & Zaidi, S. M. H. (2021). Gray level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms. *PeerJ Computer Science*, 7, e536. <https://doi.org/10.7717/peerj-cs.536>

[16] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>

[17] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>

[18] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, LNCS 11211, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

-
- [19] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/2010.11929>
- [20] Xiao, Z., Diao, G., & Deng, Z. (2024). Fine grained food image recognition based on Swin Transformer. *Journal of Food Engineering*, 380, 112134. <https://doi.org/10.1016/j.jfoodeng.2024.112134>
- [21] Nargesi, M. H., Amiriparian, J., Bagherpour, H., & Kheiralipour, K. (2024). Detection of different adulteration in cinnamon powder using hyperspectral imaging and artificial neural network method. *Results in Chemistry*, 9, 101644. <https://doi.org/10.1016/j.rechem.2024.101644>
- [22] Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science & Technology*, 46(1), 85–98. <https://doi.org/10.1016/j.tifs.2015.08.003>
- [23] Park, J. J., Cho, J. S., Lee, G., Yun, D. Y., Park, S. K., Park, K. J., & Lim, J. H. (2023). Detection of red pepper powder adulteration with Allura Red and red pepper seeds using hyperspectral imaging. *Foods*, 12(18), 3471. <https://doi.org/10.3390/foods12183471>
- [24] Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). CutMix: Training strategy that makes strongly localizable representations in image classification. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6023–6032. <https://doi.org/10.1109/ICCV.2019.00612>
- [25] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144. <https://doi.org/10.1145/3422622>
- [26] Min, W., Wang, Z., Liu, Y., Luo, M., Kang, L., Wei, X., & Jiang, S. (2023). Large scale visual food recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8), 9932–9949. <https://doi.org/10.1109/TPAMI.2023.3237871>
- [27] Arrieta, A. B., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [28] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- [29] Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2704–2713. <https://doi.org/10.1109/CVPR.2018.00286>

-
- [30] 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)*, 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- [31] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *Proceedings of the International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1412.6572>
- [32] Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 3645–3650. <https://doi.org/10.18653/v1/P19-1355>
- [33] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://arxiv.org/abs/2108.07258>
- [34] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60. <https://doi.org/10.1109/MSP.2020.2975749>
- [35] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1409.1556>