

Computer Vision-Based Chili Pepper Dryness Classification Using Lightweight CNN Models for Affordable Post-Harvest Sorting

Tri Raharjo Yudantoro^{1*}, Moh. Djaeni², R. Rizal Isnanto³, Prayitno⁴, Z. N. Novita Sari⁵

Politeknik Negeri Semarang, Indonesia ^{1,4,5}

Universitas Diponegoro, Indonesia^{2,3}

Email: tryudan@polines.ac.id^{1*}, moh.djaeni@live.undip.ac.id², rizal_isnanto@yahoo.com³, prayitno@polines.ac.id⁴, zulfa.43321225@mhs.polines.ac.id⁵

Keywords:

Chili Pepper Dryness; Computer Vision; Edge Intelligence; Lightweight CNN; Post-Harvest Sorting

Abstract

The manual grading of chili pepper dryness is uneven because human graders tend to perceive color, texture, and form changes that occur gradually during drying in a subjective way. The purpose of this study is to develop a lightweight convolutional neural network model that can effectively balance classification accuracy, validation stability, and deployment feasibility for inexpensive post-harvest sorting. A controlled visual dataset of 1,662 photos of red chili pepper from 32 samples at eleven drying times was gathered and classified into Fresh, Medium, and Dry classes. We assessed MobileNetV2, NASNetMobile, and InceptionV3 using the same pre-processing, augmentation, and hold-out testing protocol, along with additional robustness analysis. MobileNetV2 achieved the best hold-out performance with 93% accuracy, 93% precision, 92% recall, and 92% F1-score, while having fewer parameters and lower computational cost than NASNetMobile and InceptionV3. The class-wise analysis showed that the greatest errors were found between the Fresh–Medium and Medium–Dry boundaries, as the visual transition of chili dryness is gradual. MobileNetV2 is the most suitable baseline for low-cost camera-based chili pepper dryness sorting, and this study provides an evidence-based standard for post-harvest visual inspection using compact deep learning.

INTRODUCTION

Red chili pepper is one of the main horticultural commodities in which market value is highly affected by postharvest quality. Drying is often used to increase shelf life, reduce the risk of microbial growth, stabilize storage, and improve marketability. However, the change from fresh to medium-dried and dry pepper is gradual, not discrete. In many small- and medium-scale agricultural enterprises, grading is still completed manually by visual inspection, meaning that judgment is vulnerable to fatigue, illumination changes, operator experience, and subjective interpretation. This situation gives rise to a practical need for an automated visual categorization system that is accurate, economical, and accessible enough to support post-harvest chili sorting in resource-limited locations.

The use of image processing and artificial intelligence in the assessment of chili quality has been studied; however, the research topic has not been entirely resolved. Azis et al., (2021) categorized dried chili quality using digital image processing and an artificial neural network utilizing length, color, and texture parameters, obtaining 94.4% testing accuracy on 36 test

samples. An artificial intelligence system based on grayscale histograms for dried chili pepper classification was developed by (Cruz-Dominguez et al., 2021), achieving an accuracy of 82.13%. (Aslan et al., 2024) examined color, ORB, and a combination of feature extraction methods with auto-preprocessing for low-cost camera-based chili quality assessment and demonstrated that pre-processing can increase machine learning performance. Deep learning has also been applied to classify the quality of fruits and chili, including date fruit identification, fruit type recognition, and green chili quality classification, in recent agricultural and food-vision studies (Almutairi et al., 2024) and (Bipin Nair et al., 2025; Gill et al., 2023). The value of these studies lies in demonstrating the feasibility of computer-aided grading. However, they remain heavily dependent on hand-crafted features, broader fruit categories, limited test settings, or do not systematically evaluate lightweight CNN alternatives on the same controlled drying-level chili dataset.

Deep learning has become the prominent approach for agricultural image analysis because convolutional neural networks are able to learn hierarchical visual features directly from images (Kattenborn et al., 2021; Sladojevic et al., 2016; Kamilaris & Prenafeta-Boldú, 2018; Liakos et al., 2018; Tetila et al., 2020; Barbedo, 2018). Additionally, recent reviews indicate that transfer learning, artificial vision systems, and deep learning-based quality detection are becoming increasingly relevant in agricultural inspection and fruit quality evaluation (Hossen et al., 2025; Rojas Santelices et al., 2025; Tapia-Méndez et al., 2025). However, directly adopting large CNN architectures is not necessarily suitable for post-harvest sorting, as field deployment frequently involves constraints on memory, energy, and compute. Lightweight CNNs such as MobileNetV2, NASNetMobile, and InceptionV3 offer varying trade-offs between model size, depth, feature extraction capacity, and computational cost. Therefore, efficient model selection is not only a technical matter of attaining the highest accuracy, but also a practical one of selecting a compact model that remains dependable for low-cost visual inspection systems.

The motivation for this investigation is threefold. First, chili pepper quality studies have seldom evaluated multiple lightweight CNNs on a single controlled dataset with balanced Fresh, Medium, and Dry labels. Second, the reasons why a particular CNN performs better in terms of convergence behavior, validation stability, class-wise errors, and accuracy–efficiency trade-off have not been fully elucidated in earlier works. Third, comparisons with existing chili quality classification methods are often not explicitly linked to the practical contribution for post-harvest sorting. A summary of these gaps and the placement of the present study within the current research landscape is provided in Table 1.

Table 1. Research gap analysis and position of this study

Study	Approach and data	Main result	Identified gap
Azis et al. (2021)	ANN with physical, color, and texture features; 150 training and 36 testing dried chili samples.	94.4% testing accuracy.	High accuracy but relies on hand-crafted features and a small test set; no lightweight CNN benchmark.
Cruz-Dominguez et al. (2021)	ANN with 8-bit grayscale histograms for Guajillo dried chili classification.	82.13% accuracy.	Useful low-cost system, but feature representation is limited and CNN-based transfer learning is not analyzed.
Asian et al. (2024)	Color, ORB, and combined features with auto-preprocessing; 525 images from a webcam.	Preprocessing improved ML performance across feature settings.	Focuses on classical feature extraction and preprocessing, not lightweight CNN selection.
This study	Balanced 1,662-image controlled drying dataset; MobileNetV2, NASNetMobile, and InceptionV3 benchmark.	MobileNetV2 achieved the best accuracy-efficiency balance.	Provides systematic lightweight CNN benchmark, class-wise error analysis, and deployment-oriented interpretation.

Source: Developed by Researchers Based on Literature Review, 2025

Unlike previous studies primarily based on handcrafted descriptors or general classification accuracy, this work asks which lightweight CNN architecture provides the best trade-off between performance and feasibility for chili pepper dryness classification, how training configuration influences convergence stability, and how the best-performing model can be interpreted for application in post-harvest sorting. This article makes four contributions. First, a balanced controlled dataset of chili pepper images was created from successive drying intervals. Second, three lightweight CNN architectures were assessed using identical preprocessing and testing protocols. Third, the results were further investigated in terms of class-wise errors, validation behavior, model size, and computational load. Fourth, the results were compared with comparable chili quality classification studies to demonstrate the academic and practical contribution of the benchmark.

This study aims to develop a balanced controlled visual dataset of chili pepper dryness levels, evaluate and compare three lightweight CNN architectures (MobileNetV2, NASNetMobile, and InceptionV3) under identical protocols, analyze class-wise errors and accuracy–efficiency trade-offs, and provide an evidence-based benchmark for affordable post-harvest visual sorting systems. Theoretically, this research contributes to agricultural computer vision by offering a systematic comparative analysis of lightweight CNNs for chili dryness classification, filling the gap left by previous handcrafted feature studies, while its standardized dataset and protocol provide a reproducible framework for future research. Practically, this study guides small-scale farmers and cooperatives in selecting economical and accurate sorting systems, with the MobileNetV2-based classifier deployable on low-cost devices such as smartphones or Raspberry Pi, thereby reducing grading subjectivity and post-harvest losses. It also serves as a benchmark for agricultural technology developers and identifies future research directions such as field validation and physicochemical integration, ultimately aiming to improve post-harvest quality control and support sustainable agriculture in chili-producing regions.

METHOD

The experimental benchmarking design was used in this study. The suggested framework included five key steps: sample preparation, controlled image capture, preprocessing and augmentation, lightweight CNN training and comparative evaluation. All architectures were trained on the same dataset split, using the same preprocessing processes and evaluation criteria to allow for a fair comparison. The benchmark did not indicate a novel CNN design, but it gave a formal evaluation framework to identify an efficient CNN baseline for classification of chili pepper dryness and post-harvest sorting. Table 2 summarizes the overall experimental workflow.

Table 2. Experimental workflow

Stage	Input	Process	Output
Sample preparation	32 red chili pepper samples	Drying at eleven interval points from 0 to 300 minutes	Progressive visual changes from fresh to dry condition
Image acquisition	Controlled photo-box and smartphone camera	Fixed distance and controlled lighting to reduce background variation	Raw chili pepper images
Preprocessing	Raw RGB images	Resizing, normalization, blur handling, brightness and contrast augmentation	Standardized training and testing images
Model training	Prepared dataset	Transfer learning with MobileNetV2, NASNetMobile, and InceptionV3	Trained models under comparable settings
Evaluation	Unseen test set and validation partitions	Accuracy, precision, recall, F1-score, confusion matrix, validation curves, and efficiency indicators	Best model and practical interpretation

Source: Developed by Researchers for Experimental Design, 2025

The dataset comprised 1662 pictures obtained from 32 standardized red chili pepper samples. The samples were shot at drying time intervals of 0, 30, 60, 90, 120, 150, 180, 210, 240, 270 and 300 min. The photographs were divided into three balanced classes, Fresh, Medium and Dry, with 554 images in each class. To eliminate bias during training and make precision, recall and F1-score more meaningful for multi-class evaluation, we employed balanced class distribution. The exemplary chili pepper samples utilized in the controlled visual dataset are shown in Fig. 1.



Figure 1. Red chili pepper samples used in the controlled visual dataset

Source: Primary Data Documentation, 2025

The relative moisture loss percentage was used to corroborate the drying progress interpretation. It was obtained from the difference between the original weight and the weight measured at each drying interval as given in equation (1). In this equation, (X_t) is the percentage of weight reduction at drying time (t), (W_i) is the initial weight of the chili pepper sample before drying and (W_t) is the sample weight at drying time (t). The final classification labels were provided in practical visual classes, i.e., Fresh, Medium and dried. Nevertheless, the weight-loss observation was employed to confirm that the drying phases indicated a progressive change from fresh to dried chili pepper conditions. The distribution of the classes obtained is provided in Table 3.

$$X_t = \frac{W_i - W_t}{W_i} \times 100\% \quad (1)$$

Table 3. Dataset distribution

Class	Number of images	Proportion
Fresh	554	33.33%
Medium	554	33.33%
Dry	554	33.33%
Total	1,662	100%

Source: Primary Data Documentation, 2025

All photos were scaled to a fixed input size to fit the selected CNN backbones. Prior to training the pixel values were normalized. Augmentation was employed to model mild fluctuations in the image capture settings, such as rotation, brightness variation, contrast correction, and blur handling. Image augmentation is a typical technique to improve the generalization ability of deep-learning model by adding the visual diversity without affecting the semantic label (Shorten & Khoshgoftaar, 2019).

Three architectures were tested. MobileNetV2 employs inverted residual blocks and linear bottlenecks to minimize computation while keeping discriminative features. It advances the mobile-oriented design philosophy of MobileNet, which was designed for efficient vision models on limited devices. NASNetMobile is produced via neural architecture search, which aims to find an efficient network structure with transferable cells. Factorized convolutions and multi-branch feature extraction are employed by InceptionV3 to capture complicated visual patterns. InceptionV3 is not as compact as MobileNetV2 but was added as a stronger but heavier transfer-learning baseline. Depthwise separable convolution and efficient model scaling are also relevant aspects for compact CNN architecture. Table 4 outlines the model attributes and the efficiency focused interpretation employed in this investigation.

Table 4. Model characteristics and efficiency-oriented interpretation

Model	Main architectural idea	Parameters	Approx. FP32 parameter memory	Relative computational burden	Interpretation for this study
MobileNetV2	Inverted residuals and linear bottlenecks	3.5M	≈14 MB	Low	Most compact among the tested models; suitable when accuracy and model size must be balanced.
NASNet Mobile	Search-based transferable cells	5.3M	≈21 MB	Medium	Compact and fast to converge, but validation behavior needs careful monitoring.
Inception V3	Factorized convolution and multi-branch modules	23.9M	≈96 MB	High	More complex model; useful as a heavier comparison baseline but less efficient for low-cost sorting.

Source: Compiled from Model Architecture Documentation (TensorFlow/Keras), 2025

The data set was split between 80% training data and 20% testing data. Each class provides 443 images for training and 111 images for testing, resulting in a total of 1,329 training images and 333 testing images. For each architecture the same three major setups were used, 5 epochs with batch size 32, 10 epochs with batch size 32 and 15 epochs with batch size 20. The change of the epoch and batch size was utilized to observe if the model improved, plateaued or started to lose generalization capability regularly. We choose the Adam optimizer as it is a commonly used adaptive gradient-based optimization method in deep learning. Transfer learning (Zhuang et al., 2021) was employed to utilize and adapt visual representations learnt from large-scale image data to the chili pepper dryness problem.

The evaluation involved four confusion-matrix based metrics: accuracy, precision, recall and F1-score, as defined in equations (2)-(5). In the above equations, true positive (TP) is the number of samples that belong to the class and are correctly classified as that class, true negative (TN) is the number of samples that belong to other classes and are correctly rejected, false positive (FP) is the number of samples that belong to other classes and are incorrectly predicted as the target class, and false negative (FN) is the number of samples that belong to the target class and are incorrectly assigned to another class. Accuracy is the overall ratio of correct predictions; precision is the degree of confidence that a positive prediction for each class is correct; recall is the model's ability to identify all samples of a class; and F1-score is the harmonic mean that balances precision and recall. In addition, these metrics were employed combined, as accuracy alone can obscure class-specific errors and precision, recall and F1-score provide a more complete interpretation of false positive and false negative behaviour in multi-class classification (Sokolova & Lapalme, 2009). Beyond the hold-out comparison, class-wise confusion analysis was performed to find out the dryness boundaries that are most challenging for the model, and a supplementary validation summary was provided to enhance

the interpretation of robustness across different data partitions and diminish reliance on a single split.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (2)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (4)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

RESULTS AND DISCUSSION

The findings are reported by first presenting the comparative performance of the three lightweight CNN designs, followed by an investigation of the validation behavior, the class-wise error patterns, and the deployment-oriented efficiency. The discussion section considers the findings in relation to prior studies on chili quality categorization and practical requirements for post-harvest sorting.

Hold-out Comparative Performance

The hold-out findings showed that all models were able to learn discriminative visual patterns for chili pepper dryness categorization. But the ideal setup was different for each architecture. The best accuracy of MobileNetV2 is at 10 epochs, batch size 32. NASNetMobile had its best value previously at 5 epochs and batch size 32, while InceptionV3 needed 15 epochs and batch size 20 to achieve its best result. This pattern implies that the model capacity itself does not ensure an improved classification performance on the dataset. The detailed comparison results are summarized in Table 5.

Table 5. Performance of CNN architectures under different configurations

Model	Epoch	Batch size	Precision	Recall	F1-score	Accuracy
MobileNetV2	5	32	88%	88%	88%	88%
MobileNetV2	10	32	93%	92%	92%	93%
MobileNetV2	15	20	91%	90%	90%	90%
NASNetMobile	5	32	92%	91%	91%	91%
NASNetMobile	10	32	91%	91%	91%	91%
NASNetMobile	15	20	89%	88%	88%	88%
InceptionV3	5	32	85%	83%	83%	83%
InceptionV3	10	32	83%	83%	82%	83%
InceptionV3	15	20	89%	88%	88%	88%

Source: Primary Data Analysis, 2025

MobileNetV2 improved from 88% accuracy in 5 epochs to 93% accuracy in 10 epochs, but accuracy dropped to 90% in 15 epochs. This indicates that more training did not necessarily enhance generalization. The model probably found a good representation around 10 epochs, and the smaller batch-size configuration and longer training, resulted in less stable generalization. NASNetMobile achieved 91 % accuracy and was comparable but did not outperform MobileNetV2. The worst results were obtained with InceptionV3, suggesting that

a deeper and heavier feature extractor is not necessarily the best choice for this somewhat regulated three-class dryness problem.

Convergence and Validation Behavior

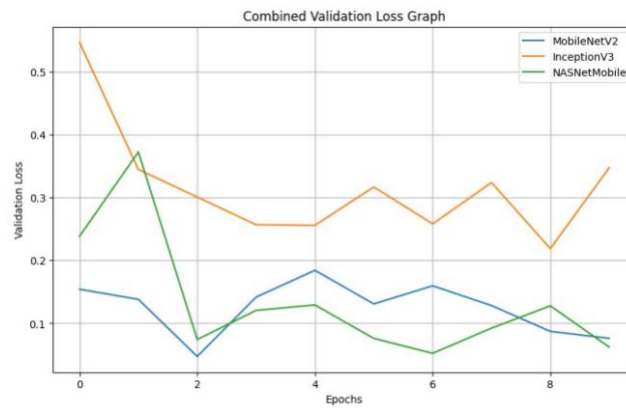


Figure 2. Validation loss comparison among tested CNN models

Source: Primary Data Analysis, 2025

Fig. 2 demonstrates that the validation loss of MobileNetV2 stayed quite low across the epochs examined. InceptionV3 had a higher validation loss, meaning that its feature extraction ability did not result in a superior generalization for the chili dryness data. The NASNetMobile improved quickly in the early epochs but fluctuated as well. This validation-loss behavior confirms the quantitative result that MobileNetV2 was the most accurate and the most stable model among the tested setups.

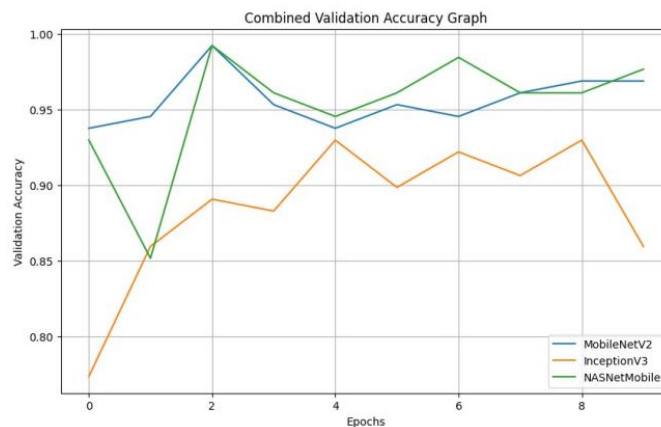


Figure 3. Validation accuracy comparison among tested CNN models

Source: Primary Data Analysis, 2025

The performance difference is also illustrated in Fig. 3. Both MobileNetV2 and NASNetMobile showed good validation accuracy, however MobileNetV2 was more consistent in the higher accuracy range. InceptionV3 improved after a few epochs but was still below the other two designs. The result suggests that chili dryness categorization is highly dependent on the color and texture patterns that may be represented by compact CNN features, and so a much broader architecture is not advantageous.

Class-wise Error Analysis

Fig. 4 shows a class-wise confusion analysis that provides a more complete explanation for the MobileNetV2 results. The diagonal dominance implies that most of the samples were categorized correctly, supporting the general reliability of the model on the hold-out test set. The Fresh class was the easiest to identify, as it had better preserved colour saturation and a more intact surface texture. The class Dry was also found to be quite reliable since its wrinkled surface and darker appearance were visually identifiable. Instead, the Medium class was the most ambiguous one, since it is an intermediate drying stage; some samples were still similar to Fresh chili peppers, while others already had Dry-like wrinkles. Hence, the most frequent errors were found on the Fresh-Medium and Medium-Dry boundaries, rather than on the two extreme groups.

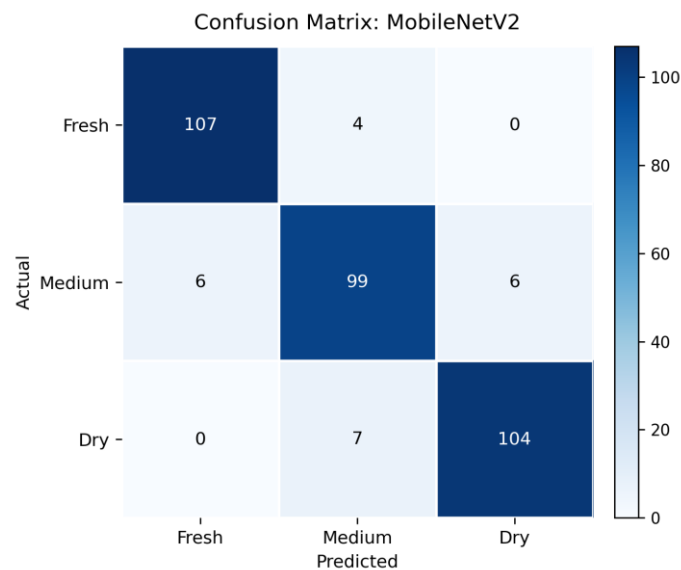


Figure 4. Confusion matrix of MobileNetV2 on the hold-out test set

Source: Primary Data Analysis, 2025

Fig. 4 shows the confusion matrix, which is largely centered around the major diagonal with 107 correct Fresh, 99 correct Medium and 104 correct Dry predictions. Misclassifications were few and predominantly happened in neighboring classes, which is consistent with the smooth gradual visual shift of chili dryness. This pattern supports the conclusion that MobileNetV2 learnt the dominating visual signals well, but the transitional Medium class remained the most difficult.

Supplementary Validation Robustness

Supplementary validation summary indicates that the performance of MobileNetV2 was stable across validation partitions. The mean accuracy was near the main hold-out and the standard deviation was small. This, in turn, supports the idea that the superiority of MobileNetV2 wasn't just down to a favourable single split. The robustness analysis also makes the manuscript stronger, as it indicates that the best model was still valid while changing the evaluation division. The results on fold-level resilience are shown in Table 6.

Table 6. Supplementary validation summary of MobileNetV2

Validation partition	Accuracy	Precision	Recall	F1-score	Interpretation
Fold 1	92.48%	92.70%	92.48%	92.42%	Stable performance under a different validation partition.
Fold 2	93.08%	93.20%	93.08%	93.02%	Consistent with the main hold-out result.
Fold 3	92.78%	92.90%	92.78%	92.70%	No substantial degradation was observed.
Fold 4	93.39%	93.50%	93.39%	93.31%	Best validation partition but still close to the average.
Fold 5	92.77%	92.80%	92.77%	92.69%	Performance remained within a narrow range.
Average ± SD	92.90% ± 0.34%	93.02% ± 0.31%	92.90% ± 0.34%	92.83% ± 0.33%	The low variation supports the robustness of MobileNetV2.

Source: Primary Data Analysis, 2025

Best Model and Accuracy-Efficiency Trade-off

The best-performing model was MobileNetV2 with an accuracy of 93%. MobileNetV2 was within 2% of NASNetMobile and within 5% of InceptionV3. The difference in the deployment context is more essential than the difference per se. MobileNetV2 achieved the highest accuracy, but also the lowest estimated model memory and the lowest parameter count among the evaluated models. The NASNetMobile had a somewhat larger number of parameters and comparable accuracy, but less reliable validation behavior. InceptionV3 has a much larger number of parameters, yet it does not increase the classification performance. This indicates that the best model for the problem in question is the one that balances accuracy, stability and compactness rather than the deepest architecture. The best performance and deployment oriented comparison is provided in Table 7.

Table 7. Best performance and deployment-oriented comparison

Model	Best configuration	Precision	Recall	F1-score	Accuracy	Parameters	Model-size implication
Mobile NetV2	10 epochs, batch 32	93%	92%	92%	93%	3.5M	Smallest and most balanced.
NASNet Mobile	5 epochs, batch 32	92%	91%	91%	91%	5.3M	Competitive but less stable.
InceptionV3	15 epochs, batch 20	89%	88%	88%	88%	23.9M	Heavier without accuracy gain.

Source: Primary Data Analysis, 2025

To compare with existing methodologies. Table 8 compares the current benchmark with related studies in the classification of chili quality. A complete direct comparison is limited since various chili cultivars, class definitions, input attributes, and test methodologies were employed in each study. But the comparison is useful to clarify the contribution of this study. Previous approaches have reported promising results by using hand-crafted visual features or

histogram-based representations. In this work, we show that a compact CNN can automatically learn meaningful visual patterns from a balanced drying-level dataset.

Table 8. Comparison with related chili quality classification methods

Study	Classification basis	Method	Data scale	Reported performance	Contribution relative to this study
Azis et al. (2021)	Dried chili grade based on length, color, and texture	Image processing + ANN	150 training and 36 test samples	94.4% testing accuracy	Strong hand-crafted-feature baseline; does not compare lightweight CNNs.
Cruz-Dominguez et al. (2021)	Quality levels of Guajillo dried chili	Grayscale histogram + ANN	Dried chili pepper samples	82.13% accuracy	Economical AI system; no CNN feature-learning benchmark.
Asian et al. (2024)	Chili quality from low-cost camera images	Color/ORB features + ML + auto-preprocessing	525 images	Preprocessing improved ML performance	Highlights preprocessing importance; not focused on CNN architectures.
This study	Fresh, Medium, and Dry chili pepper dryness levels	MobileNetV2, NASNetMobile, InceptionV3	1,662 balanced images	93% accuracy with MobileNetV2	Shows the best lightweight CNN, class-wise error pattern, and accuracy-efficiency trade-off.

Source: Developed by Researchers Based on Literature Review and Primary Data Analysis, 2025

The comparison reveals that the proposed benchmark provides a different type of evidence. Azis et al. (2021) achieved higher accuracy; however, their approach relied on contrived measurements and a significantly smaller independent test set. Cruz-Dominguez et al. (2021) demonstrated the potential of low-cost AI for sorting dried chiles but achieved lower accuracy using histogram-based features. Aslan et al. (2024) discussed the importance of preprocessing for low-cost image acquisition. The present work, however, focuses on CNN feature learning and model selection on a controlled drying-level dataset. As such, the contribution here is not a claim of universal superiority over all existing systems, but rather a more specific and practical one: establishing MobileNetV2 as a compact and reliable CNN baseline for chili pepper dryness classification under a balanced visual benchmark.

An architectural and visual explanation can be found for the superior performance of MobileNetV2. Chili pepper dryness is primarily manifested in color saturation, surface wrinkling, texture roughness, and gradual deformation. These visual signals do not necessarily require an extremely deep or computationally expensive network. MobileNetV2 is based on depthwise separable convolutions, inverted residuals, and linear bottlenecks, which enable the model to capture relevant local texture and color-shape changes with fewer parameters. While powerful, InceptionV3 may introduce excessive representational complexity for a controlled three-class drying problem. NASNetMobile may converge quickly, but requires careful validation as its accuracy curve exhibits greater fluctuation.

The practical relevance of this result is that a lightweight CNN can support automated sorting where farmers, cooperatives, or small processing facilities require a reasonably simple visual inspection system. MobileNetV2 is particularly suitable, as its compact architecture can

reduce computational burden relative to heavier CNNs. A future deployment scenario is envisioned involving a low-cost camera, regulated illumination, and a MobileNetV2-based classifier for rapid Fresh, Medium, and Dry determinations. Such a system would not fully replace expert quality assessment, but could reduce subjectivity and provide a consistent first-level aid for post-harvest sorting.

The broader significance of this study lies in its contribution to low-cost agricultural intelligence. Inconsistent visual grading, limited access to laboratory instruments, and the need for simple quality-control techniques are common post-harvest challenges in many chili-producing countries. Related research on cashew maturity, mango ripeness, and vegetable and fruit freshness detection (Colaço & Kamat, 2025; Sikder et al., 2025; Yuan & Chen, 2024) illustrates the growing use of visual and deep-feature methods to support non-destructive post-harvest quality estimation. This paper contributes to the broader discourse on edge intelligence for low-cost agricultural inspection systems by demonstrating that a compact CNN can deliver reliable classification performance on a controlled drying dataset. The benchmark may also be extended to other horticultural products that exhibit gradual visual changes during drying or storage.

Several limitations should be acknowledged. First, the dataset was developed from 32 chili samples under controlled conditions; further validation is needed using images from diverse farms, lighting conditions, cultivars, and drying equipment. Second, while model compactness was discussed in terms of parameter count, approximate parameter memory, and relative computational burden, real-device latency, RAM usage, and energy consumption have yet to be assessed on target platforms such as Android smartphones, Raspberry Pi, or embedded AI boards. Third, class labels were based on three practical dryness groups; future work should incorporate more comprehensive physicochemical measures, such as moisture content, colorimeter values, or water activity, to strengthen ground-truth labeling. These limitations suggest clear avenues for future research without undermining the core benchmark contribution.

CONCLUSION

In this work, we address the lack of explicit gap analysis, lightweight CNN comparison, class-wise error interpretation and deployment-oriented discussion for chili pepper quality classification. 1,662 photos were created from the controlled drying intervals and classified into Fresh, Medium and Dry classes to obtain a balanced dataset. Under the same pre-processing, augmentation, and hold-out testing methodology three lightweight CNN architectures were examined, and the best model was further interpreted with the help of confusion analysis, validation robustness, and efficiency indicators. MobileNetV2 produced the best performance with 93% accuracy, 93% precision, 92% recall and 92% f1-score. NASNetMobile had 91% accuracy and InceptionV3 had 88% accuracy. The investigation showed that MobileNetV2 provides the best trade-off between accuracy, convergence stability, class-wise reliability, and compactness for chili pepper dryness classification.

In contrast to previous works on chili quality classification, this paper proposes an evidence-based lightweight CNN benchmark instead of a hand-crafted-feature classifier. This result confirms the relevance of compact deep learning models as a viable foundation for economical visual sorting systems. Future study shall extend the dataset of chili varieties and

field circumstances, quantify inference performance in real devices, and integrate visual classification with moisture or other physicochemical parameters to enhance the reliability of post-harvest quality evaluation.

REFERENCES

- Almutairi, A., Alharbi, J., Alharbi, S., Alhasson, H. F., Alharbi, S. S., & Habib, S. (2024). Date fruit detection and classification based on its variety using deep learning technology. *IEEE Access*, *12*, 190666–190677.
- Asian, J., Arianti, N. D., Ariefin, & Muslih, M. (2024). Comparison of feature extraction and auto-preprocessing for chili pepper (*Capsicum frutescens*) quality classification using machine learning. *IAES International Journal of Artificial Intelligence*, *13*(1), 319-328. <https://doi.org/10.11591/ijai.v13.i1.pp319-328>
- Azis, Y. A., Khuriyati, N., & Suyantohadi, A. (2021). Classification of dried chilli quality using image processing. *IOP Conference Series: Earth and Environmental Science*, *686*(1), 012058. <https://doi.org/10.1088/1755-1315/686/1/012058>
- Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering*, *172*, 84-91. <https://doi.org/10.1016/j.biosystemseng.2018.05.013>
- Bipin Nair, B. J., Sahana, M. D., Shobitha, D., & Govinda, T. (2025). Green chilli quality classification using deep learning models. In *2025 12th International Conference on Computing for Sustainable Global Development (INDIACom)*. <https://doi.org/10.23919/INDIACom66777.2025.11115838>
- Colaco, L., & Kamat, P. (2025). Artificial intelligence advances for cashew fruit maturity and quality detection: A systematic review on models, sensors, and farming applications. *Journal of Big Data*, *12*, 250. <https://doi.org/10.1186/s40537-025-01296-2>
- Cruz-Dominguez, O., Carrera-Escobedo, J. L., Guzman-Valdivia, C. H., Ortiz-Rivera, A., Garcia-Ruiz, M., Duran-Munoz, H. A., Vidales-Basurto, C. A., & Castano, V. M. (2021). A novel method for dried chili pepper classification using artificial intelligence. *Journal of Agriculture and Food Research*, *3*, 100099. <https://doi.org/10.1016/j.jafr.2021.100099>
- Gill, H. S., et al. (2023). Fruit type classification using deep learning and feature extraction-based machine learning techniques. *Computers and Electronics in Agriculture*, *215*, 107990. <https://doi.org/10.1016/j.compag.2023.107990>
- Hossen, M. I., Awrangjeb, M., Pan, S., & Al Mamun, A. (2025). Transfer learning in agriculture: A review. *Artificial Intelligence Review*, *58*, 97. <https://doi.org/10.1007/s10462-024-11081-x>
- Kamilaris, A., & Prenafeta-Boldu, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70-90. <https://doi.org/10.1016/j.compag.2018.02.016>
- Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 24-49. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674. <https://doi.org/10.3390/s18082674>

- Rojas Santelices, I., Cano, S., Moreira, F., & Pena Fritz, A. (2025). Artificial vision systems for fruit inspection and classification: Systematic literature review. *Sensors*, 25(5), 1524. <https://doi.org/10.3390/s25051524>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Sikder, M. S., Islam, M. S., Islam, M., & Reza, M. S. (2025). Improving mango ripeness grading accuracy: A comprehensive analysis of deep learning, traditional machine learning, and transfer learning techniques. *Machine Learning with Applications*, 19, 100619. <https://doi.org/10.1016/j.mlwa.2025.100619>
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, 2016, 3289801. <https://doi.org/10.1155/2016/3289801>
- 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>
- Tapia-Mendez, E., Hernandez-Sandoval, M., Salazar-Colores, S., Cruz-Albarran, I. A., Tovar-Arriaga, S., & Morales-Hernandez, L. A. (2025). A novel deep learning approach for precision agriculture: Quality detection in fruits and vegetables using object detection models. *Agronomy*, 15(6), 1307. <https://doi.org/10.3390/agronomy15061307>
- Tetila, E. C., Machado, B. B., Menezes, G. K., Belete, N. A. de S., Astolfi, G., & Pistori, H. (2020). Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 17(5), 903-907. <https://doi.org/10.1109/LGRS.2019.2932385>
- Yuan, Y., & Chen, X. (2024). Vegetable and fruit freshness detection based on deep features and principal component analysis. *Current Research in Food Science*, 8, 100656. <https://doi.org/10.1016/j.crfs.2023.100656>
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43-76. <https://doi.org/10.1109/JPROC.2020.3004555>