Article

Mitigating Spatial Scale Loss in CNN-Based Fine-Grained Image Classification: Application to Date Fruit Grading

Ziaul Haque [1], Murat Koklu [2], Mohammed Aquil Mirza [3], Marwan Omar [4] and Saidova Mukhayyokhon [5, *]

[1] Department of Plant Protection, Faculty of Agricultural Sciences, Aligarh Muslim University, Aligarh 202002, India
[2] Technology Faculty, Department of Computer Engineering, Selcuk University, Konya 42031, Türkiye
[3] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon PQ806, Hong Kong
[4] ITM Department, Illinois Institute of Technology, Chicago, IL60616, USA
[5] Ground transport systems, Tashkent State Technical University, Tashkent 100095, Uzbekistan

* Correspondence: m.saidova87@mail.ru (S.M.)

Abstract

Accurate classification of date fruit varieties and size grades is critical for automated grading and post-harvest quality assessment. However, conventional image preprocessing techniques based on uniform resizing often distort size-dependent visual cues, leading to misclassification among size levels within the same variety. To address this limitation, this study proposes a size-preserving rescaling strategy for deep learning–based date fruit classification. Experiments are conducted on a curated dataset comprising 5,836 images distributed across 12 classes, representing four date varieties (Aseel, Dandhi, Karblain, and Kupro), each categorized into three size levels: large, medium, and small. Five convolutional neural network architectures—MobileNetV3, DenseNet121, InceptionV3, ResNet101, and VGG16—are evaluated using identical training, validation, and test splits under a supervised learning framework. When standard resized inputs are used, the highest classification accuracy achieved is 82.18%, with macro F1-scores close to 0.82. In contrast, incorporating the proposed size-preserving rescaling approach leads to substantial performance improvements across all models. The best results are obtained with ResNet101, achieving an accuracy of 94.44%, a macro precision of 0.9476, and a macro F1-score of 0.9446, followed closely by DenseNet121 with 94.32% accuracy. These findings demonstrate that preserving size information during preprocessing significantly enhances class separability and reduces size-level confusion, making the proposed approach well suited for practical date fruit grading systems.

Keywords: date fruit; size-preserving rescaling; deep learning; image classification; automated fruit grading

Citation: Haque Z., Koklu M., Mirza M.A., Omar M., Mukhayyokhon S. Mitigating Spatial Scale Loss in CNN-Based Fine-Grained Image Classification: Application to Date Fruit Grading. *Impact in Agriculture*. **2025**, *1*, 5. https://doi.org/10.65500/agriculture-2025-005

Received: 17 November 2025 | Revised: 02 December 2025 | Accepted: 13 December 2025 | Published: 30 December 2025

1. Introduction

Date palm (Phoenix dactylifera L.) is one of the most important fruit crops in arid and semi-arid regions. According to FAOSTAT, global date production exceeded 9.7 million metric tons in recent years. Egypt leads worldwide production with approximately 1.9 million tons, followed by Saudi Arabia at about 1.6 million tons, Algeria at 1.3 million tons, Iran at nearly 1.0 million tons, and Iraq at more than 0.6 million tons annually [1].

Together, these countries account for a dominant share of global supply and export volume. As production scales, post-harvest operations such as sorting, grading, and packaging increasingly determine economic value. Commercial grading of date fruits depends on several visual attributes. Variety, maturity stage, surface quality, and fruit size directly affect market price and consumer acceptance. Among these, size plays a central role in packaging standards and pricing tiers. In many production facilities, grading remains manual. This process is slow, inconsistent, and sensitive to operator experience. With increasing production volumes, manual grading introduces delays and variability that reduce overall efficiency.

Automated visual grading systems have been studied as a solution to this problem. Early approaches relied on classical image processing and machine learning. Features such as color histograms, texture descriptors, shape parameters, and geometric measurements were extracted manually and fed into classifiers such as support vector machines, k-nearest neighbors, or neural networks [2]. These systems performed adequately under controlled lighting and background conditions, but their accuracy dropped when applied to images captured under real production environments.

The introduction of deep learning shifted research toward end-to-end image classification. Convolutional neural networks learn hierarchical representations directly from raw images and reduce dependence on handcrafted features. Transfer learning using pre-trained architectures such as VGG, ResNet, Inception, DenseNet, and MobileNet became common practice in agricultural vision tasks [3–5]. One of the early deep learning frameworks for date fruit classification explored fine-tuned pre-trained models within a smart city context, supported by edge computing for low-latency processing [6]. The study demonstrated system feasibility, but performance analysis focused on overall classification rather than class-specific behavior. Subsequent work expanded the scope toward defect detection and ripening stage recognition using VGG-based architectures, reporting overall accuracies above 95 percent under controlled conditions [7].

Robotic harvesting and orchard-level applications introduced additional complexity. A large-scale study proposed deep learning models for real-time classification of date fruit type and maturity under natural environmental conditions, using more than 8,000 images captured in orchards [8]. Reported accuracies exceeded 97 percent for type and maturity classification, with inference times below 40 ms. However, size grading was not treated as an independent classification objective. Other studies combined classification and regression tasks to estimate fruit type, maturity level, and weight. Deep learning models were used for type and maturity estimation, while regression models such as support vector machines were applied for weight prediction [9,10]. Although these systems achieved high maturity classification accuracy, weight estimation accuracy remained lower, often below 85 percent. Size classification at the image level was not explicitly addressed.

Dataset construction received increasing attention. Public datasets were introduced to support localization, recognition, and classification of date fruits, including annotations for size, maturity, and variety[11,12]. These datasets reduced barriers for benchmarking, yet many studies continued to report only overall accuracy or average precision. Class-wise analysis, particularly for size categories, was often omitted. A consistent trend across the literature is the widespread use of fixed-size input representations. Most deep learning pipelines resize images to standard dimensions such as 224 by 224 or 299 by 299 pixels to match network input requirements. This step is applied universally, regardless of task objectives. It is rarely questioned or analyzed.

For variety recognition, resizing often works well because discriminative cues are dominated by color, texture, and shape. For grading tasks where size is a primary factor, resizing alters the image geometry. Large and small fruits are normalized to identical spatial extents. Absolute and relative scale information is removed before the image reaches the network. This limitation appears indirectly in reported results. Several studies note confusion between visually similar classes, particularly intermediate categories, despite high overall accuracy [12,13]. Medium-sized classes often show lower recall and precision compared to extreme categories. These patterns persist across architectures, suggesting that the issue lies in data representation rather than model depth or parameter count.

Attempts to mitigate this problem often focus on model complexity. Feature fusion, ensemble learning, attention mechanisms, and hybrid architectures have been proposed to boost accuracy [14–18]. While these methods improve performance for variety recognition, they increase computational cost and system complexity. They also operate on resized inputs and therefore do not restore size information lost during preprocessing.

Data augmentation has also been widely adopted. Techniques such as random scaling, cropping, rotation, and synthetic image generation using GANs increase dataset size and balance class distributions [19,20]. These methods improve robustness but further distort spatial scale, which is critical for size-based grading. Detection-based approaches use object detectors such as YOLO and Faster R-CNN to localize fruits before classification [21–23]. These systems are effective for counting and harvesting tasks, but they introduce bounding box regression errors and increase inference latency. They also require additional annotation effort and computational resources.

Recent reviews highlight the importance of dataset quality, annotation strategy, and representation choices in agricultural AI systems [24,25]. A large comparative study on dataset complexity shows that classifier performance is strongly influenced by feature representation, class structure, and dimensionality [26,27]. These findings support the view that preprocessing decisions directly affect classification outcomes.

Despite extensive research, the effect of resizing on size-based date fruit classification has not been systematically studied. Most works assume that convolutional networks learn scale-invariant features sufficient for grading tasks. Experimental evidence suggests otherwise.

In this work, we investigate size classification within the same date fruit species, a problem that remains insufficiently addressed in existing deep learning based grading systems. The dataset comprises images from multiple date varieties and size categories. An analysis of the original images shows that fruit size is strongly reflected in spatial occupancy and orientation, which are removed by standard resizing to fixed input dimensions. This preprocessing practice suppresses scale information that is critical for size discrimination. To address this limitation, a size-preserving rescaling strategy is introduced, where images are embedded into a fixed canvas while maintaining original scale and aspect ratio. This approach preserves size-related cues at the input level without altering network architecture, providing a more suitable representation for size-based grading tasks.

The main contributions of this work are as follows.

- An analysis of how image resolution and orientation encode size information in date fruit images.
- Evidence that standard resizing degrades size discrimination across multiple CNN architectures.
- A size-preserving rescaling strategy that retains spatial scale without modifying network design.
- A controlled comparison across five deep learning models using identical training parameters.
- Consistent performance gains, with accuracy improving from approximately 0.82 to above 0.94 and reduced size-level confusion.

2. Materials and Methods

This section describes the dataset, preprocessing strategies, model architectures, and experimental protocol employed in this study. The proposed framework is designed to evaluate the impact of input representation on size-based classification of date fruits. In particular, the methodology contrasts a conventional resizing-based preprocessing pipeline with a size-preserving rescaling strategy that retains spatial scale information. The complete workflow, including data preparation, preprocessing variants, model training, and evaluation, is summarized in Figure 1. Each component of the framework is described in detail in the following subsections.

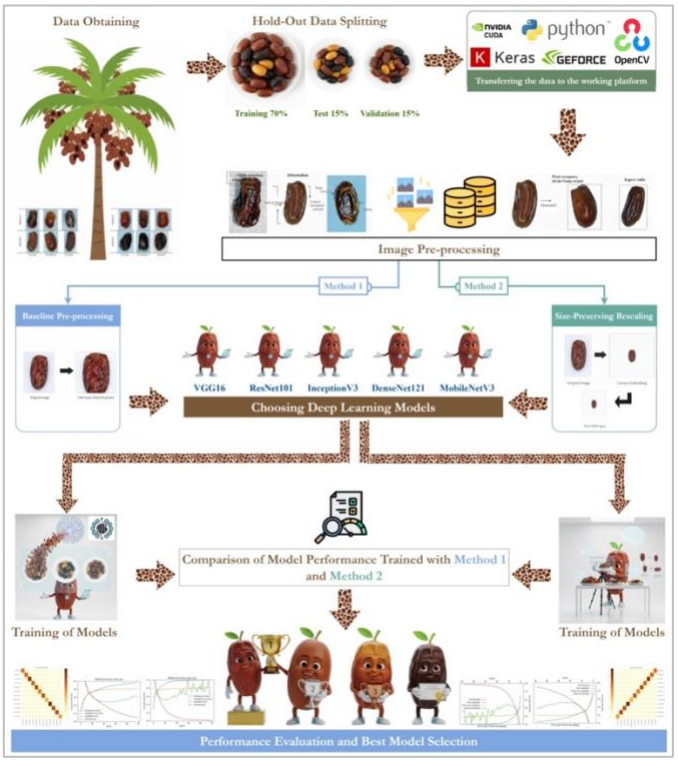


Figure 1. Overall methodological pipeline for date fruit variety and size classification.

2.1 Dataset Description

The experiments in this study are conducted using a publicly available size-based date fruit image dataset released in 2025 [28]. The dataset was prepared specifically for classification tasks involving both variety and size differentiation of date fruits. It contains images from four

commonly cultivated date varieties, namely Aseel, Dandhi, Karblain, and Kupro. Each variety is divided into three size categories: large, medium, and small, resulting in a total of twelve distinct classes.

The complete dataset comprises 5,836 images. The data are partitioned into training, validation, and test sets using a fixed split to ensure consistency across experiments. The training set contains 4,079 images, the validation set contains 876 images, and the test set contains 881 images. This corresponds to approximately 70% of the data used for training, 15% for validation, and 15% for testing. Each class is represented across all three subsets. The detailed class-wise distribution is summarized in Table 1.

Table 1. Details of the original dataset

Class	Class Name	Training	Validation	Test	Total
1	Aseel Large	287	62	62	411
2	Aseel Medium	350	75	76	501
3	Aseel Small	441	94	95	630
4	Dandhi Large	330	71	71	472
5	Dandhi Medium	410	88	88	586
6	Dandhil Small	329	71	71	471
7	Karblain Large	349	75	75	499
8	Karblain Medium	214	46	47	307
9	Karblain Small	445	95	96	636
10	Kupro Large	358	77	77	512
11	Kupro Medium	354	76	77	507
12	Kupro Small	212	46	46	304
Total		**4079**	**876**	**881**	**5836**

Images were captured under natural acquisition conditions. Variations are present in background, illumination, orientation, and distance from the camera. Fruits appear both horizontally and vertically oriented, depending on placement during image capture. No artificial constraints were imposed during acquisition, reflecting realistic grading scenarios encountered in agricultural and post-harvest environments.

2.2 Visual Characteristics of the Dataset

A preliminary inspection of the dataset reveals systematic differences in image resolution and spatial occupancy across size categories. Larger fruits generally occupy a greater portion of the image frame, while smaller fruits appear with reduced pixel coverage. This effect is amplified by orientation. Horizontally placed fruits exhibit higher width-to-height ratios, while vertically placed fruits exhibit higher height-to-width ratios.

Measured image resolutions vary across size categories. For horizontally oriented fruits, typical resolutions are approximately 618 × 448 pixels for large, 594 × 394 pixels for medium, and 527 × 372 pixels for small categories. For vertically oriented fruits, typical resolutions are approximately 398 × 625 pixels for large, 390 × 584 pixels for medium, and 381 × 571 pixels for small categories. These variations encode implicit size information through spatial scale rather than explicit geometric features.

These characteristics are visually illustrated in sample images from the dataset. Figure 2 presents representative examples dataset before any resizing or preprocessing is applied. The figure highlights the differences in pixel occupancy and aspect ratio across size categories.

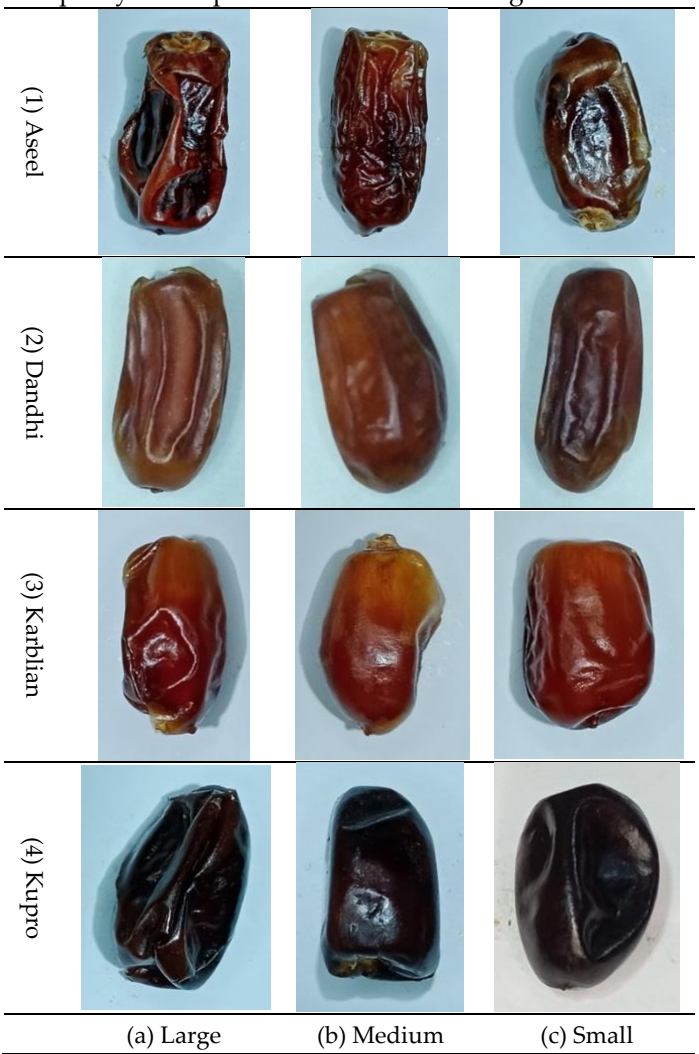


Figure 2. Representative samples of date fruit varieties and size categories used in the study: (1) Aseel, (2) Dandhi, (3) Karblain, and (4) Kupro, shown across three size classes—(a) Large, (b) Medium, and (c) Small.

2.3 Baseline Preprocessing Pipeline

For baseline experiments, a standard preprocessing pipeline is applied. All images are loaded in RGB format and normalized using pixel-wise scaling. No color correction or background segmentation is performed. Images are resized directly to meet the fixed input size requirements of each convolutional neural network architecture.

Architectures such as MobileNetV3, DenseNet121, ResNet101, and VGG16 require input dimensions of 224 × 224 pixels, while InceptionV3 requires 299 × 299 pixels. During resizing, the original aspect ratio is not preserved. Images are uniformly scaled along both axes to fit the square input dimensions.

This resizing strategy is widely adopted in deep learning pipelines due to its simplicity and compatibility with pre-trained models. However, it forces all fruits to occupy similar pixel areas regardless of their original size or orientation. As a result, spatial scale differences across size categories are suppressed at the input level.

2.4 Motivation for Size-Preserving Rescaling

Size-based grading relies on relative spatial information rather than fine-grained texture or color differences. In the given dataset, fruits from the same variety exhibit similar surface appearance, color distribution, and texture patterns across size categories. When spatial scale is removed, remaining visual cues become insufficient for reliable size discrimination.

The baseline preprocessing pipeline treats all images as scale-invariant objects. While this assumption is acceptable for variety classification, it is not suitable for size-based classification within the same species. This observation motivates the introduction of a preprocessing strategy that preserves spatial scale while maintaining compatibility with convolutional neural network inputs.

2.5 Size-Preserving Rescaling Strategy

To retain size-related spatial information, a size-preserving rescaling strategy is adopted. Instead of directly resizing images to square inputs, each image is first embedded into a fixed-size canvas while maintaining its original resolution and aspect ratio.

An analysis of the dataset shows that the maximum observed image resolution does not exceed 650 × 650 pixels. Based on this observation, a canvas of size 660 × 650 pixels is defined to accommodate all images without cropping or scaling. Each image is placed at the center of the canvas. No resizing is applied at this stage.

The background area surrounding the image is filled with a constant value to ensure uniformity across samples. This background does not introduce additional texture or patterns. The placement preserves the original spatial occupancy of the fruit within the canvas, allowing relative size differences to remain encoded.

Figure 3 illustrates this process. Figures 3(a) and 3(c) show original horizontal and vertical images with their native resolutions. Figures 3(b) and 3(d) show the corresponding images embedded into the fixed-size canvas. The preserved spatial scale and aspect ratio are visually evident, while the canvas ensures consistent outer dimensions.

After embedding, the canvas image is resized once to the required network input size. Since the relative scale between fruit and background is preserved, spatial size information remains encoded in the final input representation.

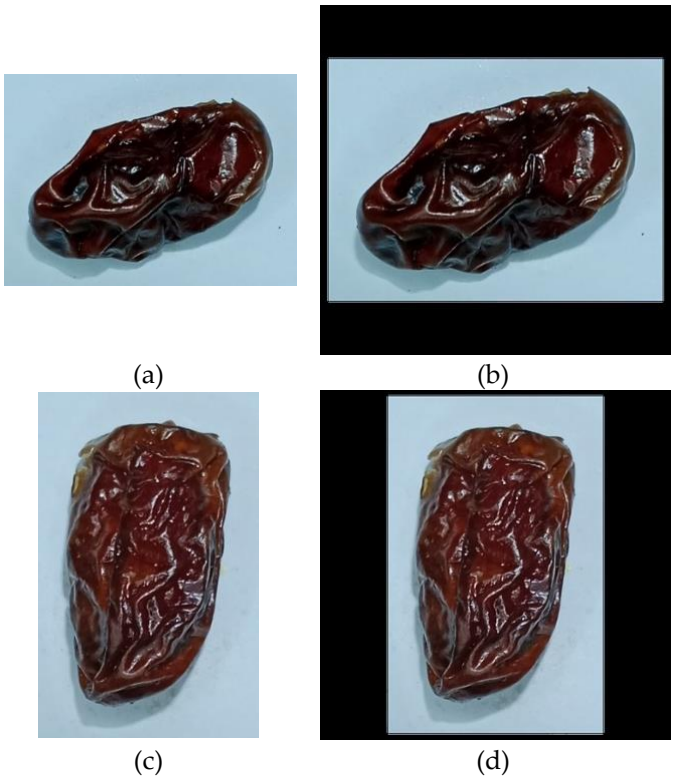


Figure 3. Size-preserving rescaling examples. (a) Original horizontal image. (b) Horizontal image after rescaling. (c) Original vertical image. (d) Vertical image after rescaling.

2.6 Deep Learning Architectures

Five convolutional neural network architectures are employed to evaluate the effect of preprocessing on size-based classification. The models are selected based on three criteria: architectural diversity, representational capacity,

and established use in agricultural image analysis [29–32]. Specifically, MobileNetV3 represents a lightweight architecture optimized for efficiency, DenseNet121 and InceptionV3 represent moderate-depth models with enhanced feature reuse and multi-scale representation, while ResNet101 and VGG16 represent deeper architectures with higher representational capacity. This selection enables systematic analysis of whether the proposed preprocessing strategy benefits models across different depths and complexities.

All models are initialized using ImageNet pre-trained weights to ensure stable convergence and fair comparison. The final classification layer of each network is replaced with a fully connected layer matching the twelve target classes. No additional layers or architectural modifications are introduced. This design isolates the effect of input representation, ensuring that any observed performance differences arise from the preprocessing strategy rather than changes in network structure or optimization capacity.

2.7 Training Configuration

To ensure controlled and fair comparison, all models are trained using identical hyperparameters. The Adam optimizer is employed with a fixed learning rate of 0.00001. Batch size is set to 16, and training is conducted for 50 epochs.

Preliminary trials are conducted using learning rates ranging from 0.1 to 0.000001 and batch sizes of 8, 16, and 32. The selected configuration provides stable training behavior across all architectures. Early stopping is not applied, allowing full observation of convergence trends.

Training is performed using the same configuration for both baseline preprocessing and size-preserving rescaling setups. All other factors, including initialization, optimizer settings, and data splits, remain unchanged.

2.8 Experimental Design and Evaluation Protocol

Two experimental settings are defined to isolate the effect of input representation on size-based classification. In the first setting, models are trained using the baseline preprocessing pipeline, where images are resized directly to fixed square input dimensions. In the second setting, models are trained using the proposed size-preserving rescaling strategy. All other factors, including data splits, model initialization, and training hyperparameters, remain unchanged across settings.

Each convolutional neural network is trained independently under both configurations. No weight sharing or cross-training is performed between experiments. This design ensures that any observed differences arise solely from the preprocessing strategy rather than architectural or optimization changes.

All experiments are conducted on a workstation running Windows 10 Pro. The system is equipped with an Intel i5 processor operating at 2.9 GHz, 16 GB of RAM, and an NVIDIA GeForce GTX 1660 GPU. The software environment consists of Python version 3.8, OpenCV version 4.7 for image processing, and Keras version 2.8 for model implementation and training. This setup provides stable execution time and reproducible training behavior across runs.

Model evaluation is performed using the held-out test set. Performance is assessed using accuracy, precision, recall, and F1-score. These metrics are computed at the class level for all size and variety categories. Macro-averaged metrics are reported to assign equal importance to each class, while weighted averages reflect the underlying class distribution. Confusion matrices are generated to visualize correct and incorrect predictions at the class level. These outputs provide a detailed view of classification behavior under both preprocessing strategies.

3. Results and Discussion

3.1. Overview of Evaluation Protocol

The experimental results are reported under two controlled settings. Set 1 corresponds to models trained using standard network resizing, where input images are directly resized to fixed dimensions required by each architecture. Set 2 corresponds to models trained using the proposed size-preserving rescaling strategy. In both settings, five convolutional neural networks are evaluated under identical training conditions. Training and validation accuracy and loss trends are illustrated in Figures 4 and 6. Class-level prediction behavior is examined through confusion matrices in Figures 5 and 7. Quantitative performance metrics, including accuracy, precision, recall, and f1 score, are summarized in Tables 2 and 3. The discussion focuses on convergence behavior, stability of validation performance, and the extent of size-level confusion within the same fruit variety.

3.2. Training and Validation Behavior on the Original Dataset

Figure 4 presents the training and validation accuracy and loss curves for all five models trained on the original dataset using standard resizing. Across all architectures, training accuracy increases rapidly during the first 15 to 23 epochs and exceeds 0.90 well before the midpoint of

training. By contrast, validation accuracy stabilizes at noticeably lower levels and shows limited improvement beyond approximately 20 epochs.

For MobileNetV3 in Figure 4(a), training accuracy approaches 0.97 by epoch 50, while validation accuracy plateaus near 0.80. Training loss decreases steadily from above 1.0 to below 0.15, whereas validation loss begins to rise after epoch 20, indicating divergence between training and validation behavior. DenseNet121 in Figure 4(b) shows a similar pattern, with training accuracy approaching 0.99 and validation accuracy remaining around 0.82. Validation loss exhibits visible oscillations after mid-training, despite continuous reduction in training loss.

InceptionV3, shown in Figure 4(c), converges more slowly and exhibits pronounced fluctuations in validation loss throughout training. Validation accuracy remains below 0.80 for most of the training period, consistent with the instability observed in the loss curve. ResNet101 in Figure 4(d) reaches high training accuracy rapidly, exceeding 0.98 before epoch 15, but validation accuracy plateaus early near 0.81 and validation loss shows irregular spikes. VGG16 in Figure 4(e) converges more gradually and exhibits the lowest validation accuracy, stabilizing around 0.76, with validation loss remaining consistently higher than that of the other models.

Overall, Figure 4 shows a consistent pattern across architectures. All models fit the training data effectively, but validation performance saturates early and remains substantially lower than training performance. This behavior suggests limited generalization under standard resizing, despite prolonged training.

3.3. Class-Level Performance on the Original Dataset

The confusion matrices for the original dataset are shown in Figure 5. Across all models, correct predictions concentrate along the diagonal at the variety level, while most errors occur between size categories within the same variety.

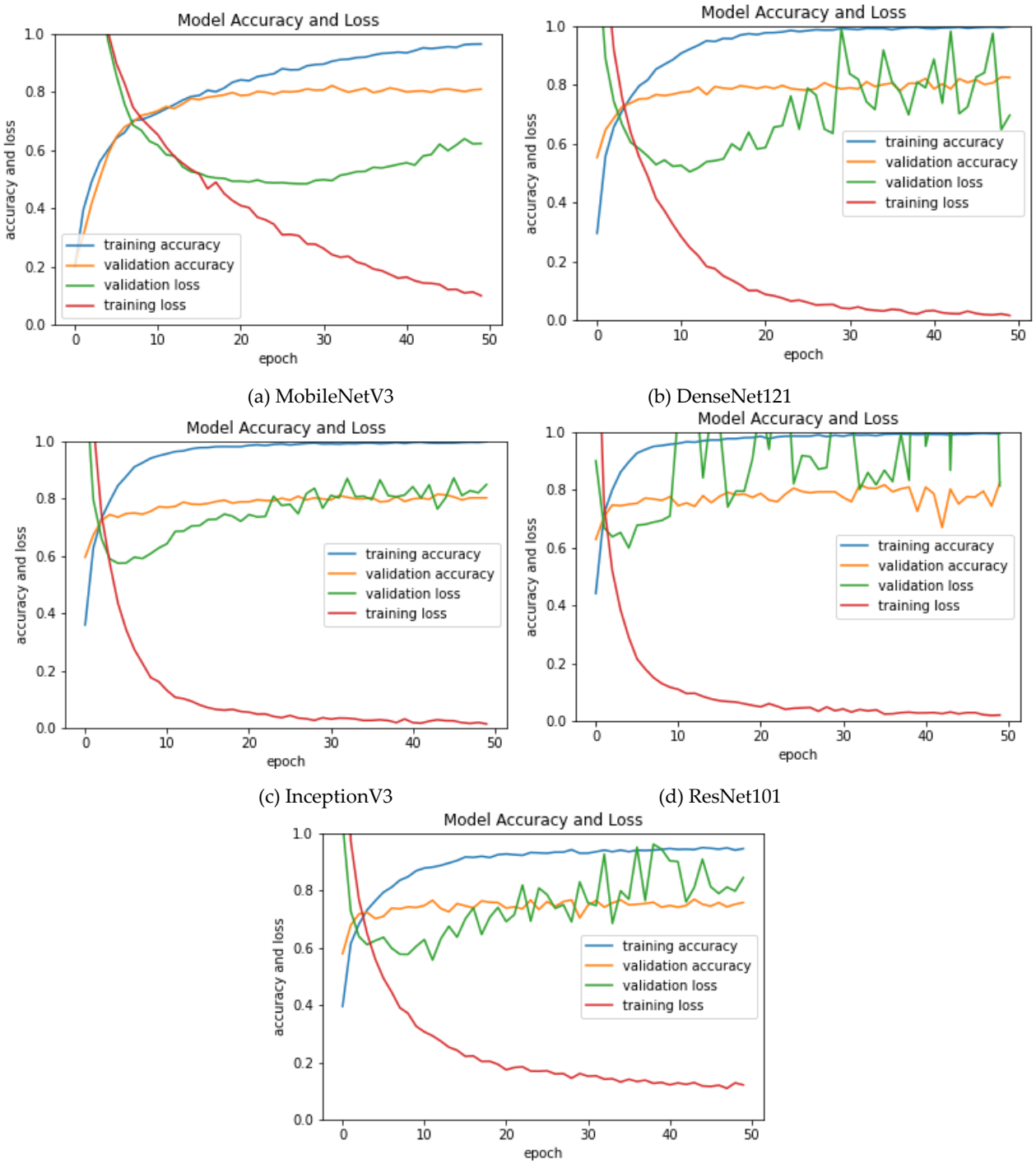
For MobileNetV3 in Figure 5(a), 43 out of 62 Aseel Large samples are correctly classified, while 15 are misclassified as Aseel Medium. Similar patterns appear for Dandhi, where only 41 Dandhi Large samples are correctly classified, with 19 misclassified as Dandhi Medium and 11 as Dandhi Small. Karblain and Kupro varieties exhibit stronger diagonal dominance, with limited size confusion.

DenseNet121 in Figure 5(b) improves recognition for Karblain and Kupro sizes, with 72 out of 75 Karblain Large samples correctly classified and near-perfect classification for Kupro Large and Medium. However, size-level confusion persists for Aseel and Dandhi. For example, 23 Dandhi Medium samples are misclassified as Dandhi Large, and 6 as Dandhi Small.

InceptionV3 in Figure 5(c) shows fragmented predictions for Dandhi sizes. Only 41 Dandhi Large samples are correctly classified, while 19 are confused with Dandhi Medium and 11 with Dandhi Small. This pattern aligns with the unstable validation loss observed in Figure 4(c). ResNet101 in Figure 5(d) achieves stronger diagonal dominance for larger sizes but continues to misclassify medium-sized fruits. VGG16 in Figure 5(e) shows the highest degree of confusion, particularly for Dandhi Large and Medium, where correct classifications drop below 50 samples in several cases.

These confusion matrices confirm that standard resizing preserves sufficient information for variety recognition but suppresses size-related cues. Misclassification is dominated by confusion between large, medium, and small categories of the same variety, rather than cross-variety errors.

Table 2 quantifies this behavior. Overall accuracy ranges from 0.7650 for VGG16 to 0.8218 for MobileNetV3 and DenseNet121. Macro recall values remain below 0.82 for all models, reflecting reduced sensitivity to medium-sized classes. Weighted metrics follow similar trends, confirming that size discrimination is the primary limiting factor under standard resizing.



(a) MobileNetV3

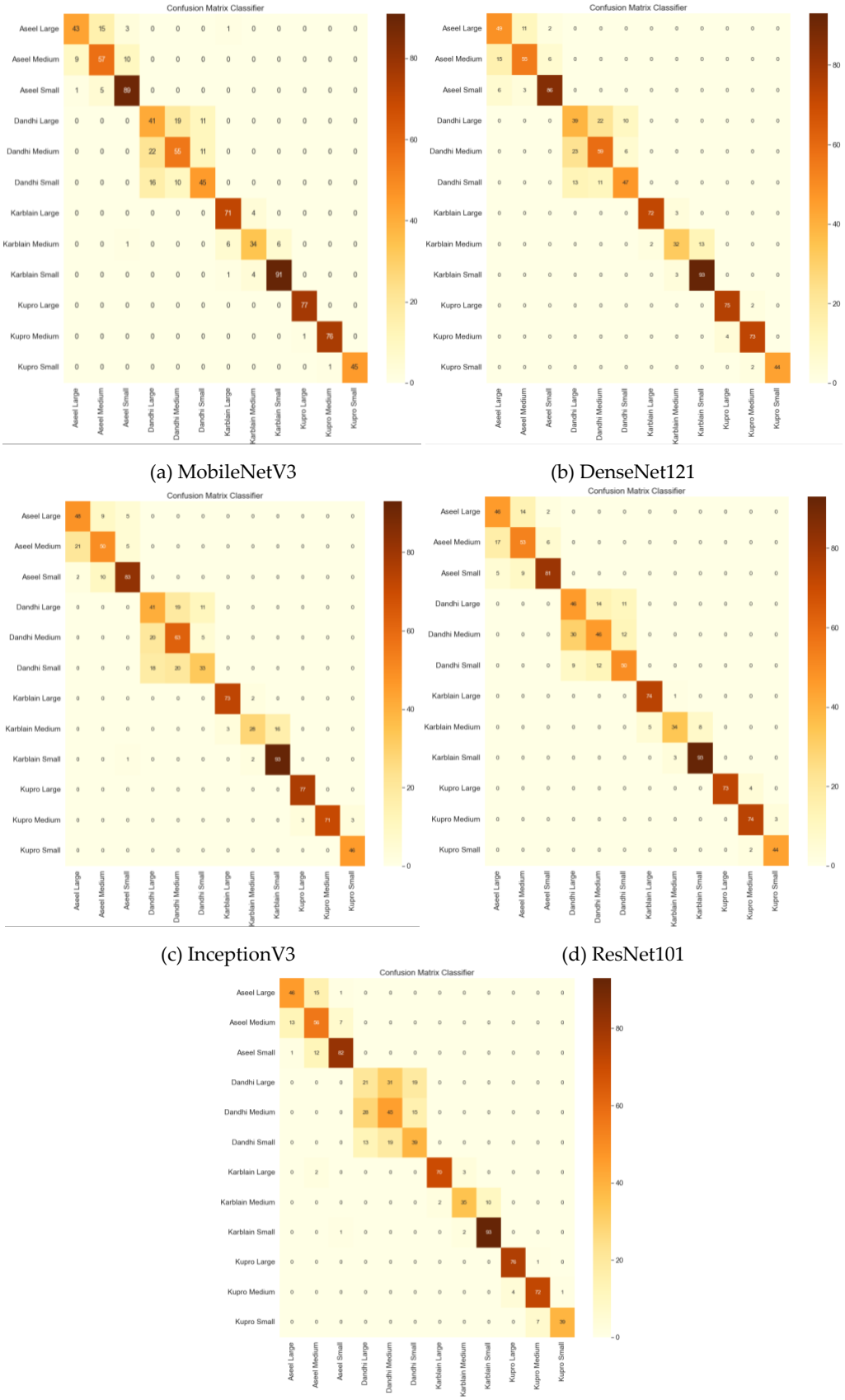
(b) DenseNet121

(c) InceptionV3

(d) ResNet101

(e)VGG16

Figure 4. Training and validation accuracy and loss curves for all models using standard resizing.



(a) MobileNetV3

(b) DenseNet121

(c) InceptionV3

(d) ResNet101

(e)VGG16

Figure 5. Confusion matrices for all models using standard resizing.

Table 2. Performance Metrics on Original Dataset (Set 1)

Model	Accuracy	Macro Precision	Macro Recall	Macro F1	Weighted Precision	Weighted Recall	Weighted F1
MobileNetV3	0.8218	0.8235	0.8167	0.8191	0.8222	0.8218	0.8211
DenseNet121	0.8218	0.8258	0.8158	0.8192	0.8242	0.8218	0.8216
InceptionV3	0.8014	0.8070	0.7937	0.7951	0.8051	0.8014	0.7988
ResNet101	0.8104	0.8135	0.8093	0.8093	0.8135	0.8104	0.8101
VGG16	0.7650	0.7707	0.7596	0.7637	0.7661	0.7650	0.7644

3.4. Training and Validation Behavior with Size-Preserving Rescaling

Figure 6 shows the training and validation curves obtained after applying the size-preserving rescaling strategy. Compared to Figure 4, all models demonstrate faster convergence, improved alignment between training and validation accuracy, and more stable validation loss behavior.

For MobileNetV3 in Figure 6(a), validation accuracy rises sharply within the first 10 epochs and stabilizes above 0.93, closely tracking training accuracy throughout the remainder of training. Validation loss decreases steadily and remains below 0.25, without the upward trend observed in Figure 4(a). DenseNet121 in Figure 6(b) exhibits similarly smooth convergence, with validation accuracy exceeding 0.94 and validation loss remaining stable after early epochs.

InceptionV3 in Figure 6(c) shows a marked reduction in validation loss oscillations compared to its original setting. Validation accuracy stabilizes near 0.93, while validation loss remains consistently below 0.35. ResNet101 in Figure 6(d) converges rapidly and maintains close alignment between training and validation curves, with minimal fluctuation. VGG16 in Figure 6(e) demonstrates substantial improvement, with validation accuracy increasing from approximately 0.76 in Figure 4(e) to above 0.94, and validation loss reduced to below 0.30.

These trends indicate that preserving spatial scale at the preprocessing stage improves convergence reliability and stabilizes validation behavior across architectures of varying depth and capacity.

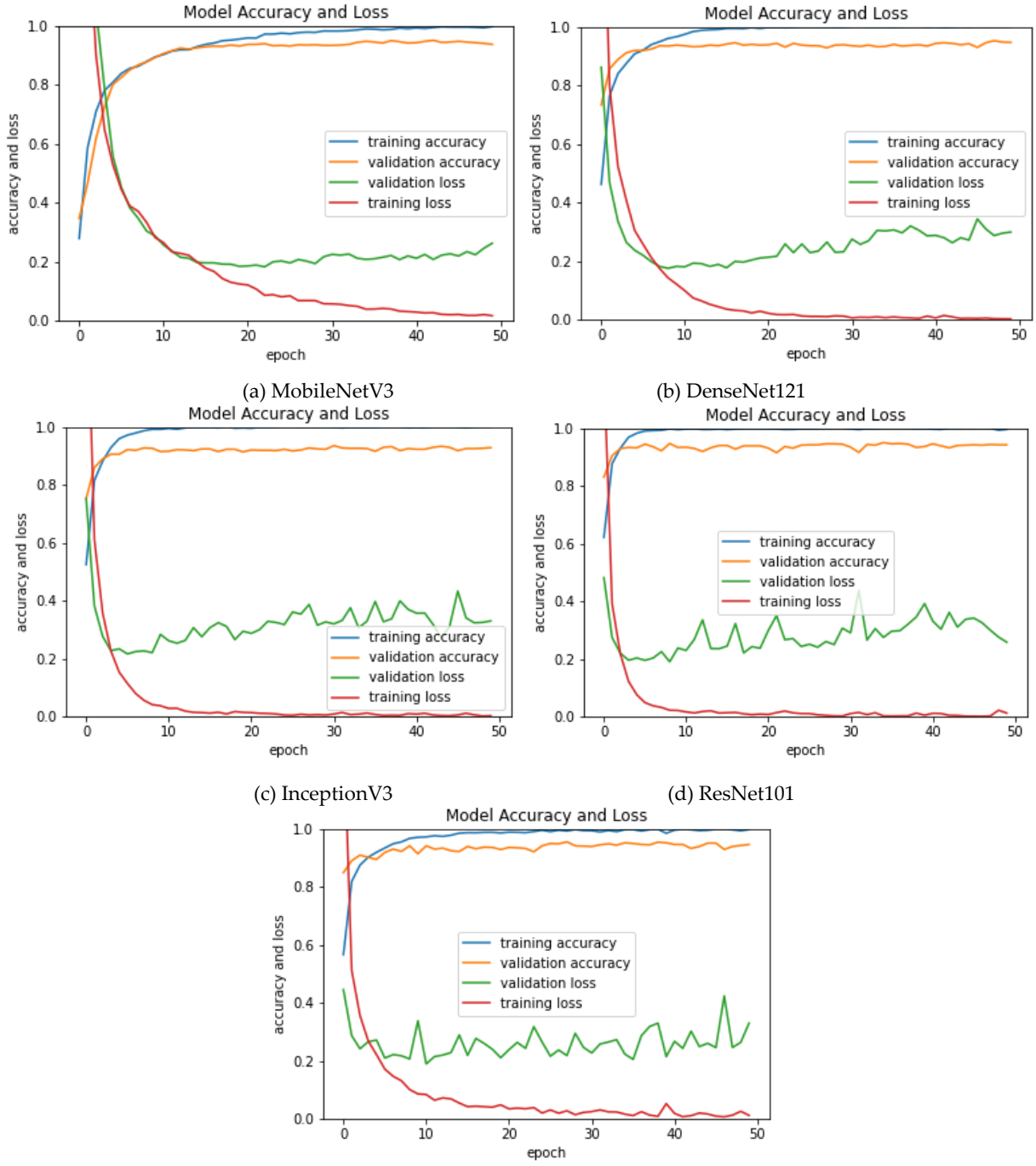
3.5. Class-Level Performance with Size-Preserving Rescaling

Figure 7 presents the confusion matrices for models trained on the rescaled dataset. In contrast to Figure 5, predictions concentrate strongly along the diagonal for all varieties and sizes, with substantially fewer off-diagonal entries.

For MobileNetV3 in Figure 7(a), 61 out of 62 Aseel Large samples are correctly classified, with only one misclassification. Dandhi Medium and Small also show improved separation, with 78 and 58 correct predictions, respectively. DenseNet121 in Figure 7(b) achieves near-perfect classification for Karblain Small, with all 96 samples correctly classified, and shows minimal confusion across other sizes.

InceptionV3 in Figure 7(c) demonstrates balanced performance across all classes, with strong diagonal dominance for Aseel, Dandhi, Karblain, and Kupro sizes. ResNet101 in Figure 7(d) maintains consistent classification across all varieties, with sparse and isolated off-diagonal errors. VGG16 in Figure 7(e) shows a pronounced reduction in size-level confusion compared to Figure 5(e), particularly for Dandhi and Karblain categories.

Table 3 summarizes these improvements. Accuracy values increase to the range of 0.9330 to 0.9444 across models. Macro f1 scores rise above 0.93 for all architectures, indicating balanced performance across size classes. Weighted metrics confirm consistent predictions across the full dataset. These numerical gains correspond closely with the stabilized training behavior observed in Figure 6 and the diagonal dominance evident in Figure 7.

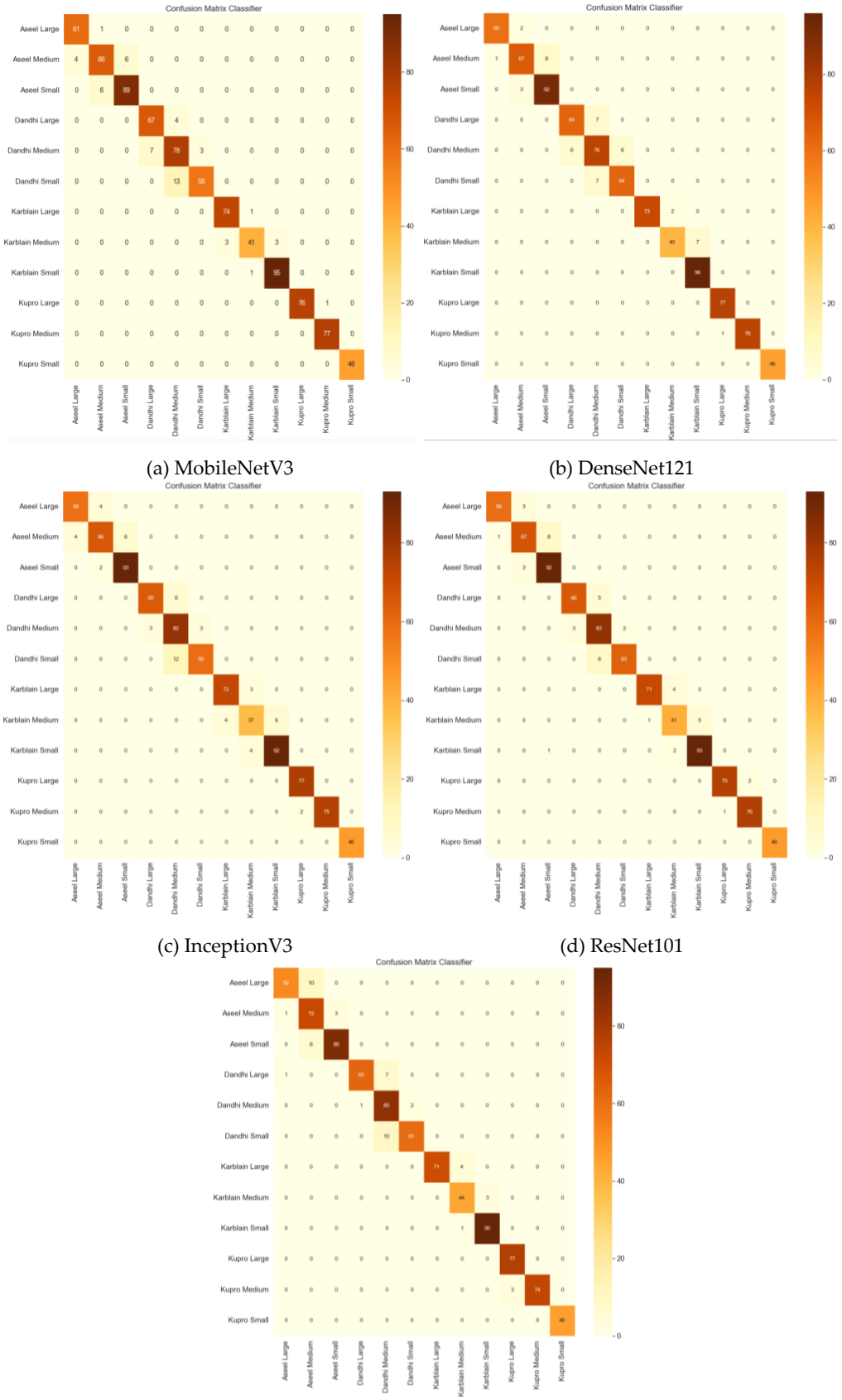


(a) MobileNetV3 (b) DenseNet121

(c) InceptionV3 (d) ResNet101

(e)VGG16

Figure 6. Training and validation accuracy and loss curves for all models with size-preserving rescaling.



(a) MobileNetV3

(b) DenseNet121

(c) InceptionV3

(d) ResNet101

(e)VGG16

Figure 7. Confusion matrices for all models with size-preserving rescaling.

Table 3. Performance Metrics with Size-Preserving Rescaling (Set 2)

Model	Accuracy	Macro Precision	Macro Recall	Macro F1	Weighted Precision	Weighted Recall	Weighted F1
MobileNetV3	0.9398	0.9440	0.9393	0.9408	0.9409	0.9398	0.9396
DenseNet121	0.9432	0.9482	0.9413	0.9443	0.9438	0.9432	0.9431
InceptionV3	0.9330	0.9351	0.9284	0.9309	0.9344	0.9330	0.9328
ResNet101	0.9444	0.9476	0.9425	0.9446	0.9457	0.9444	0.9445
VGG16	0.9410	0.9470	0.9391	0.9412	0.9457	0.9410	0.9415

3.6. Comparative Summary of Results

A direct comparison between Figures 4 and 6 highlights the role of input representation in training dynamics. Under standard resizing, validation accuracy saturates early while validation loss diverges from training loss, despite extended training. With size-preserving rescaling, validation accuracy improves rapidly and remains stable, accompanied by reduced loss variability across all models.

Comparison of Figures 5 and 7 further shows a clear reduction in size-level misclassification without any change to network architecture or training parameters. Overall accuracy improves from approximately 0.80–0.82 in Table 2 to above 0.93 in Table 3 across all architectures. These gains appear consistently for lightweight networks such as MobileNetV3 as well as deeper models such as ResNet101, indicating that performance improvement does not depend on model complexity.

These results indicate that the primary limitation in size-based date fruit classification lies in input representation rather than network capacity. Under standard resizing, models converge on the training set but fail to retain spatial cues required for reliable size separation. This limitation manifests as early validation saturation, unstable loss behavior, and systematic confusion among size classes within the same variety. Preserving relative object scale through size-aware rescaling restores this missing information and leads to stable convergence and improved class balance. The consistency of this effect across five distinct architectures confirms that the observed gains result from correcting a structural preprocessing limitation rather than model-specific optimization.

4. Conclusions and Future Directions

This study examined size-based classification within the same date fruit variety under a controlled deep learning setting. The analysis focused on a common but often overlooked preprocessing step, direct image resizing. Experiments across five convolutional architectures showed that standard resizing suppresses spatial scale information required for reliable size discrimination. This effect appeared consistently in training behavior, validation stability, and class-level predictions. Models achieved strong variety recognition but showed persistent confusion among size categories of the same variety.

A size-preserving rescaling strategy was introduced to address this limitation. By embedding each image into a fixed canvas while preserving its original scale and aspect ratio, relative size information was retained without modifying network architecture or training parameters. Across all evaluated models, this adjustment led to stable convergence, reduced validation loss divergence, and improved balance across size classes. The consistency of the improvement across lightweight and deep architectures indicates that the observed gains arise from correcting an input-level constraint rather than increasing model capacity.

The findings highlight an important implication for agricultural image analysis. When the target attribute relates directly to physical scale, preprocessing decisions directly determine what information remains learnable. Model-centric optimization alone is insufficient if discriminative cues are removed before training begins. Simple adjustments at the data preparation stage can therefore have a larger impact than architectural changes, while maintaining computational efficiency and deployment feasibility.

Despite these contributions, several limitations should be acknowledged. The dataset was collected under controlled imaging conditions with consistent background and lighting. While this setup allows clear isolation of scale effects, it does not capture the full variability encountered in field environments. In addition, the study focused on four date fruit varieties and three discrete size categories. Continuous size estimation and grading thresholds were not addressed. The evaluation was limited to image-based cues and did not incorporate depth, weight, or multi-view information that may further improve grading reliability.

Future work should extend the proposed preprocessing strategy to datasets collected under varying illumination, background clutter, and camera distance. Cross-orchard and cross-season validation would help assess robustness under real production conditions. Integration with object detection pipelines may allow size-aware preprocessing in multi-fruit scenes. Further studies may also explore regression-based size estimation and hybrid grading systems that combine visual scale with complementary physical measurements. These directions would support the development of reliable and scalable grading systems for practical agricultural deployment.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable.

Writing Assistance Disclosure: This manuscript was drafted by the authors. Grammarly and ChatGPT were used occasionally for grammar checking and language refinement only. No content generation or analytical assistance was sought from AI tools.

Data Availability Statement: The dataset used in this study is publicly available. The *Size-Based Date Fruit Dataset for Classification* can be accessed via Mendeley Data at https://doi.org/10.17632/D986HYR2KB.1.

Conflicts of Interest: The authors declare no conflicts of interest.

References:

1. Date Production by Country 2025 Available online: https://worldpopulationreview.com/country-rankings/date-production-by-country (accessed on 14 December 2025).
2. Koklu, M.; Kursun, R.; Taspinar, Y.S.; Cinar, I. Classification of Date Fruits into Genetic Varieties Using Image Analysis. *Math Probl Eng* 2021, *2021*, doi:10.1155/2021/4793293.
3. Nath, B.; Gulzar, Y.; Tamang, S.; Alkanan, M. Comparative Evaluation of Deep Learning Architectures for Brinjal Fruit Disease Classification. *Emir J Food Agric* 2025, *37*, 1–14, doi:10.3897/EJFA.2025.172982.
4. Gulzar, Y.; Ünal, Z.; Kadir¸ Şahbaz; Alkanan, M. PTL-Inception: Integrating Deep Learning and Taxonomy for Desert Plant Classification. *Diversity 2025, Vol. 17, Page 806* 2025, *17*, 806, doi:10.3390/D17110806.
5. Gulzar, Y. Applications of Transfer Learning in Sunflower Disease Detection: Advances, Challenges, and Future Directions. *Turkish Journal of Biology* 2025, *49*, 534–549, doi:10.55730/1300-0152.2763.
6. Nasiri, A.; Taheri-Garavand, A.; Zhang, Y. Image-Based Deep Learning Automated Sorting of Date Fruit. *Postharvest Biol Technol* 2019, *153*, 133–141, doi:10.1016/j.postharvbio.2019.04.003.
7. Albarrak, K.M.; Gulzar, Y.; Hamid, Y.; Mehmood, A.; Soomro, A.B. A Deep Learning-Based Model for Date Fruit Classification. *Sustainability (Switzerland)* 2022, *14*, doi:10.3390/su14106339.
8. Altaheri, H.; Alsulaiman, M.M.; Muhammad, G. Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning. *IEEE Access* 2019, *7*, 117115–117133, doi:10.1109/ACCESS.2019.2936536.
9. Faisal, M.; Albogamy, F.R.H.; ElGibreen, H.A.; Algabri, M.; Al-Qershi, F. Deep Learning and Computer Vision for Estimating Date Fruits Type, Maturity Level, and Weight. *IEEE Access* 2020, *8*, 206770–206782, doi:10.1109/ACCESS.2020.3037948.
10. Faisal, M.; Alsulaiman, M.M.; Arafah, M.A.; Mekhtiche, M.A. IHDS: Intelligent Harvesting Decision System for Date Fruit Based on Maturity Stage Using Deep Learning and Computer Vision. *IEEE Access* 2020, *8*, 167985–167997, doi:10.1109/ACCESS.2020.3023894.
11. Pérez-Pérez, D.B.; Salomón-Torres, R.; García-Vázquez, J.P. Dataset for Localization and Classification of Medjool Dates in Digital Images. *Data Brief* 2021, *36*, doi:10.1016/j.dib.2021.107116.
12. Maitlo, A.K.; Shaikh, R.A.; Arain, R.H. A Novel Dataset of Date Fruit for Inspection and Classification. *Data Brief* 2024, *52*, doi:10.1016/j.dib.2023.110026.
13. Maitlo, A.K.; Shaikh, R.A.; Arain, R.H.; Mujtaba, G. CNN-Based Intelligent System for Date Fruit Classification Using Novel Dataset. *VFAST Transactions on Software Engineering* 2024, *12*, 134–144, doi:10.21015/vtse.v12i4.1987.
14. Aiadi, O.; Khaldi, B.; Kherfi, M.L.; Mekhalfi, M.L.; Alharbi, A.R. Date Fruit Sorting Based on Deep Learning and Discriminant Correlation Analysis. *IEEE Access* 2022, *10*, 79655–79668, doi:10.1109/ACCESS.2022.3194550.
15. Alresheedi, K.M.; Aladhadh, S.; Khan, R.U.; Qamar, A.M. Dates Fruit Recognition: From Classical Fusion to Deep Learning. *Computer Systems Science and Engineering* 2022, *40*, 151–166, doi:10.32604/CSSE.2022.017931.
16. Hassan, E.; Abu-Ghazalah, S.; El-Rashidy, N.M.; Abd El-Hafeez, T.; Shams, M.Y. DenseNet Model with Attention Mechanisms for Robust Date Fruit Image Classification. *International Journal of Computational Intelligence Systems* 2025, *18*, doi:10.1007/s44196-025-00809-4.
17. Eser, M.; Bilgin, M.; Yasin, E.T.; Koklu, M. Using Pretrained Models in Ensemble Learning for Date Fruits Multiclass Classification. *J Food Sci* 2025, *90*, doi:10.1111/1750-3841.70136.
18. Hassan, E.; Abu-Ghazalah, S.; El-Rashidy, N.M.; Abd El-Hafeez, T.; Shams, M.Y. Sustainable Deep Vision Systems for Date Fruit Quality Assessment Using Attention-Enhanced Deep Learning Models. *Front Plant Sci* 2025, *16*, doi:10.3389/fpls.2025.1521508.
19. Ufuah, D.; Thomas, G.M.; Balocco, S.; Manickavasagan, A. A DATA AUGMENTATION APPROACH BASED ON GENERATIVE ADVERSARIAL NETWORKS FOR DATE FRUIT CLASSIFICATION. *Appl Eng Agric* 2022, *38*, 975–982, doi:10.13031/aea.15107.
20. Ibrahim, D.M.; Elshennawy, N.M. Improving Date Fruit Classification Using CycleGAN-Generated Dataset. *CMES - Computer Modeling in Engineering and Sciences* 2022, *130*, doi:10.32604/cmes.2022.016419.
21. Almutairi, A.; Alharbi, J.; Alharbi, S.; Alhasson, H.F.; Alharbi, S.S.; Habib, S. Date Fruit Detection and Classification Based on Its Variety Using Deep Learning Technology. *IEEE Access* 2024, *12*, 190666–190677, doi:10.1109/ACCESS.2024.3433485.
22. Lipiński, S.; Sadkowski, S.; Chwietczuk, P. Application of AI in Date Fruit Detection—Performance Analysis of YOLO and Faster R-CNN Models. *Computation* 2025, *13*, doi:10.3390/computation13060149.
23. Yousaf, J.; Abuowda, Z.; Ramadan, S.; Salam, N.; Almajali, E.R.F.; Hassan, T.; Gad, A.; Alkhedher, M.A.; Ghazal, M.A.

Autonomous Smart Palm Tree Harvesting with Deep Learning-Enabled Date Fruit Type and Maturity Stage Classification. *Eng Appl Artif Intell* 2025, *139*, doi:10.1016/j.engappai.2024.109506.

24. Noutfia, Y.; Ropelewska, E. What Can Artificial Intelligence Approaches Bring to an Improved and Efficient Harvesting and Postharvest Handling of Date Fruit (Phoenix Dactylifera L.)? A Review. *Postharvest Biol Technol* 2024, *213*, doi:10.1016/j.postharvbio.2024.112926.
25. Zangana, H.M.; Li, S.; Wani, S. Diffusion Models for Agricultural Imaging: A Systematic Review of Methods, Applications and Future Prospects. *Impact in Agriculture* 2025, *1*, 3, doi:10.65500/agriculture-2025-003.
26. Acevedo-Sanchez, G.; Alarcón-Paredes, A.; Yáñez-Márquez, C. Effect of Agriculture-Related Dataset Complexity on Classical Machine Learning and Deep Learning Classifiers Performance. *Comput Electron Agric* 2025, *239*, doi:10.1016/j.compag.2025.110941.
27. Ayoub, S.; Baig, I.; Ashraf, M.; Okasha, M. Impact of Dataset Quality on Deep Learning Models for Dragon Fruit and Leaf Health Classification. *Impact in Agriculture* 2025, *1*, 1, doi:10.65500/agriculture-2025-001.
28. Shaikh, R.A.; Mujtaba, G.; Maitlo, A.K.; Ali, D.; Shaikh, H. Size-Based Date Fruit Dataset for Classification. 2025, *1*, doi:10.17632/D986HYR2KB.1.
29. Eser, M.; Bilgin, M.; Yasin, E.T.; Koklu, M. Using Pretrained Models in Ensemble Learning for Date Fruits Multiclass Classification. *J Food Sci* 2025, *90*, e70136, doi:10.1111/1750-3841.70136;WGROUP:STRING:PUBLICATION.
30. Gulzar, Y. PapNet: An AI-Driven Approach for Early Detection and Classification of Papaya Leaf Diseases. *Applied Fruit Science 2025 67:4* 2025, *67*, 1–11, doi:10.1007/S10341-025-01466-9.
31. Gulzar, Y. Papaya Leaf Disease Classification Using Pre-Trained Deep Learning Models: A Comparative Study. *Applied Fruit Science* 2025, *67*, 1–10, doi:10.1007/S10341-025-01533-1/METRICS.
32. Gulzar, Y.; Ünal, Z. Time-Sensitive Bruise Detection in Plums Using PlmNet with Transfer Learning. *Procedia Comput Sci* 2025, *257*, 127–132, doi:10.1016/J.PROCS.2025.03.019.