



Deep Learning Based Classification of Fruits and Vegetables

Harshit Dwivedi¹, Pranav Thakral², Khushi Panwar³, Ms. Shreya Sharma⁴

Department of CSE-AI&DS,

Panipat Institute of Engineering & Technology, Samalkha, India

Mail ID: shreya.cse-aids@piet.co.in

Abstract—Automated classification of fruits and vegetables is an important computer-vision problem with applications in intelligent retail, automated sorting, inventory management, post-harvest handling, dietary assistance, and agricultural robotics. This paper presents a deep learning based framework for multi-class fruit and vegetable classification using Convolutional Neural Networks (CNNs) and transfer learning. The framework covers dataset preparation, image resizing and normalization, augmentation, CNN feature extraction, pretrained backbones, classification, evaluation, explainability, and deployment. Fruits-360 is used as the principal benchmark because its different releases provide large numbers of fruit, vegetable, nut, and seed categories; the current original-size repository reports 102,551 images across 145 classes, while earlier studies used approximately 90,000 images across 131 or 141 classes [1], [2]. Recent studies report strong benchmark performance, including 98.1% for an attention-based CNN on a 141-class setting and 97.15% for a multi-fused CNN on a 131-class setting [6], [7]. Rather than presenting these published values as new experimental results, this paper analyzes them comparatively and proposes a reproducible workflow for building and evaluating a fruit-and-vegetable classifier. The discussion emphasizes the distinction between controlled benchmark performance and real-world robustness, where illumination, occlusion, background variation, similar varieties, and domain shift remain important challenges.

Keywords: Fruit classification, vegetable classification, deep learning, convolutional neural network, CNN, transfer learning, Fruits-360, ResNet-50, VGG16, MobileNetV2, image recognition.

INTRODUCTION

Fruits and vegetables can be differentiated using visual attributes such as color, shape, texture, size, surface pattern, and maturity. Manual recognition is practical for a small number of items, but it becomes inefficient and inconsistent when large quantities must be processed. Computer-vision systems can automate this process and provide a common recognition mechanism for checkout, sorting, weighing, inventory, warehouse management, and agricultural automation [1], [3].

Conventional image-classification pipelines generally depend on handcrafted descriptors followed by classifiers such as support vector machines, k-nearest neighbors, decision trees, or random forests. Deep learning changes this paradigm by learning representations directly from image pixels. CNNs learn low-level edges and color transitions in early layers, textures and shapes in intermediate layers, and increasingly class-specific structures in deeper layers. This hierarchical representation is particularly useful for fine-grained fruit varieties whose differences may be subtle [1], [4].

The Fruits-360 dataset is one of the most widely used benchmarks in this area. Its repository currently



provides several versions, including a 100×100-pixel collection and an original-size collection containing fruits, vegetables, nuts, and seeds. The current original-size release lists 102,551 images divided into 51,345 training, 25,680 validation, and 25,526 test images across 145 classes [2]. Earlier research versions used approximately 90,380 images and 131 classes, which explains why different papers report different dataset sizes while referring to Fruits-360 [5], [6].

The main objective of this paper is to organize a complete IEEE-style research framework for deep learning based classification of fruits and vegetables. The study focuses on the problem definition, literature review, dataset and preprocessing, CNN and transfer-learning methodology, training and evaluation, literature-reported results, error analysis, applications, limitations, future scope, and reproducibility considerations. Because the supplied source paper does not provide a new experimental run with a fixed train/test protocol, numerical results are explicitly identified as results reported by cited studies rather than claimed as original measurements.

Proposed Classification Pipeline

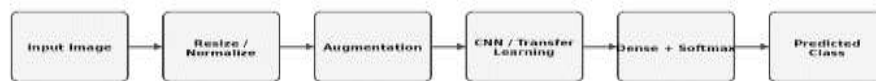


Fig. 1. Proposed end-to-end classification pipeline from image acquisition to predicted class [1], [2]. [1,2]

LITERATURE REVIEW

Muresan and Oltean introduced the Fruits-360 benchmark and demonstrated fruit recognition using deep learning [1]. The dataset subsequently became a common test bed for CNN architectures and transfer-learning methods. The dataset repository maintains a dedicated collection of papers using the benchmark, including studies on CNNs, transfer learning, EfficientNet, data augmentation, multi-label recognition, and fruit classification [2], [8].

Zhang reported a ResNet-50 based Fruits-360 classification system using preprocessing operations such as resizing, normalization, and augmentation. The study used a 131-class version of Fruits-360 containing 90,380 images and added a fully connected layer to adapt the pretrained network to the target classes [5]. This illustrates the practical value of transfer learning when the target classification problem is different from the original pretraining task.

Retail-oriented research has also highlighted a major limitation of benchmark datasets. A study on fruit classification for retail stores noted that Fruits-360 images can have small dimensions and uniform backgrounds that do not fully represent supermarket environments. The researchers therefore introduced a store-oriented dataset with fruits placed on a stainless-steel surface and evaluated CNN architectures under more realistic conditions [3].

Recent work has emphasized both accuracy and interpretability. An Enhanced Attention-CNN (EA-CNN) study reported 98.1% accuracy on a 141-class Fruit-360 experiment and 96% on a separate Fruit Recognition dataset. The architecture combines convolutional stages, attention layers, mixed pooling, dense layers, dropout, and a Softmax output, while explainable-AI methods are used to interpret



predictions [6].

Another 2024 study proposed a multi-fused CNN that combines EfficientNet-B0, MobileNetV2, and ResNet50V2. It reported 99.32% accuracy for a 24-category setting and 97.15% for a 131-category Fruit-360 setting [7]. Such results indicate the potential of feature fusion, but they should not be treated as a single leaderboard because the number of classes, dataset version, preprocessing, split strategy, and evaluation protocol differ between studies.

TABLE I. SELECTED STUDIES AND REPORTED OUTCOMES [1], [3], [5]–[7]

Study	Setting	Approach	Reported result
Muresan & Oltean [1]	Fruits-360	CNN/deep learning	Benchmark fruit recognition
Retail study [3]	60 classes + store images	CNN models	Real-world background addressed
Zhang [5]	131 classes	ResNet-50 + transfer learning	Strong benchmark performance
EA-CNN [6]	141 classes	Attention-CNN + XAI	98.1% accuracy
Multi-fused CNN [7]	131 classes	EfficientNet-B0 + MobileNetV2 + ResNet50V2	97.15% accuracy

DATASET AND PREPROCESSING

A. Dataset Characteristics

The Fruits-360 repository states that the dataset contains images of fruits, vegetables, nuts, and seeds. The original-size version currently contains 102,551 images and 145 classes, divided into training, validation, and test subsets [2]. The 100×100 version contains an even larger number of images and classes in its current release, demonstrating that the benchmark continues to evolve. For research reproducibility, a paper should always record the exact dataset release, number of classes, image resolution, and split used.

TABLE II. FRUITS-360 DATASET CHARACTERISTICS [2], [5]

Property	Value / consideration
Current original-size images	102,551 [2]
Current original-size classes	145 [2]
Training / validation / test	51,345 / 25,680 / 25,526 [2]
Earlier research setting	≈90,380 images; 131 classes [5]
Typical image condition	Single centered object; controlled background
Key limitation	Domain gap from real-world scenes



Different varieties of the same fruit may be treated as separate classes. This makes the problem fine-grained: the classifier may need to distinguish varieties that share similar colors, shapes, and textures. Such class definitions are useful for research but can make confusion matrices more informative than overall accuracy alone.

B. Preprocessing

Image preprocessing creates a consistent representation for the neural network. A typical pipeline resizes images to a fixed spatial resolution, converts them to a consistent color representation, and normalizes pixel values. Augmentation can then introduce realistic variations through rotation, translation, horizontal or vertical flipping, zoom, and brightness changes. The objective is to reduce sensitivity to nuisance variation without changing the semantic class [5], [6].

Augmentation must be selected with domain knowledge. Excessive rotation or color transformation can create images that are not representative of actual fruit photographs. Likewise, random cropping can remove the features required to distinguish two similar varieties. A robust pipeline should therefore evaluate augmentation through validation performance and, where possible, external real-world test images.

C. Dataset Split and Leakage Control

A reliable experiment should separate training, validation, and test images before model fitting. If highly similar images or frames from the same acquisition sequence occur across splits, the resulting accuracy can be inflated. The test set should remain isolated until final evaluation. When a benchmark provides official train, validation, and test folders, researchers should preserve those boundaries unless a new protocol is explicitly justified.

METHODOLOGY

CNN / Transfer-Learning Architecture

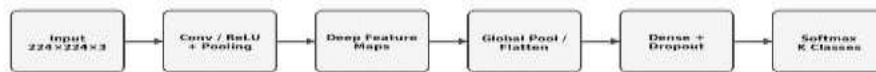


Fig. 2. CNN and transfer-learning architecture used as the methodological basis of the proposed classifier [5], [6]. [5,6]

A. CNN Feature Extraction

A CNN transforms an input image into a hierarchy of feature maps. Convolutional filters detect local patterns, ReLU-type nonlinearities allow the network to model complex functions, and pooling or strided convolution reduces spatial resolution. Batch normalization can stabilize optimization, while dropout can reduce overfitting in the final classifier. The resulting representation is passed to dense layers and a Softmax output whose dimension equals the number of target classes.

For K classes, the Softmax function converts logits z into class probabilities: $p(k|x) = \frac{\exp(z_k)}{\sum_j \exp(z_j)}$. The predicted class is the class with the highest probability. Cross-entropy is commonly used for multi-class training. Optimization may use Adam or stochastic gradient descent, with learning-rate



scheduling and early stopping used to improve convergence.

B. Transfer Learning

Transfer learning initializes the target model with features learned from a large source dataset. VGG16, MobileNetV2, ResNet50, EfficientNet, and EfficientNetV2 have been explored for fruit recognition [5], [7], [8]. A common strategy first freezes the backbone and trains a new classification head. After the head converges, selected deeper layers are unfrozen and fine-tuned with a smaller learning rate. This two-stage procedure can reduce training cost and improve generalization when the target dataset is not extremely large.

Model selection should consider both accuracy and computational cost. ResNet-style networks provide strong representational capacity, while MobileNetV2 is designed for efficient inference. EfficientNet models attempt to balance depth, width, and input resolution. For mobile and edge applications, inference time, memory footprint, and energy consumption can be as important as top-1 accuracy.

C. Attention and Explainability

Attention mechanisms can emphasize spatial or channel features that contribute strongly to class discrimination. The EA-CNN work inserts attention operations between convolution and pooling stages and uses mixed pooling to retain complementary information [6]. Explainability methods such as class-activation maps can then highlight regions that influenced a prediction. This is important because a model may otherwise learn correlations from the background, lighting, or acquisition setup rather than the fruit itself.

D. Proposed Training Strategy

The proposed experimental protocol is: (1) select and document a fixed Fruits-360 release; (2) verify class labels and remove corrupt images; (3) resize and normalize images; (4) apply training-only augmentation; (5) initialize a CNN or pretrained backbone; (6) train the classifier head; (7) fine-tune selected backbone layers; (8) monitor validation loss and accuracy; (9) evaluate once on the held-out test set; and (10) perform class-wise and error analysis. This workflow follows the methodological themes present in the supplied paper and the cited literature [1], [5], [6].

TRAINING, EVALUATION AND EXPERIMENTAL DESIGN



Fig. 3. Training, validation, testing, and deployment workflow for reproducible evaluation [2], [5]. [2,5]

A. Training Parameters

A reproducible report should specify input resolution, batch size, optimizer, initial learning rate, number of epochs, augmentation operations, frozen layers, fine-tuning layers, loss function, hardware, and random seed. Without these details, two implementations using the same dataset and architecture can



produce substantially different results. Hyperparameter tuning should use validation data only, leaving the test set untouched.

Training and validation curves provide an important diagnostic. If training accuracy increases while validation accuracy stagnates or decreases, overfitting is likely. If both remain low, the model may be underfitting or the optimization procedure may be inadequate. Learning-rate scheduling and early stopping can reduce wasted training and improve stability.

B. Evaluation Metrics

Accuracy provides the proportion of correctly classified images. Precision indicates how often predictions assigned to a class are correct, while recall indicates how many images belonging to that class are successfully recovered. F1-score is the harmonic mean of precision and recall. Macro averaging gives every class equal weight; weighted averaging accounts for the number of samples per class. A confusion matrix provides a direct view of which varieties are being confused.

TABLE III. EVALUATION MEASURES FOR MULTI-CLASS FRUIT CLASSIFICATION [5], [6]

Metric	Formula / description	Interpretation
Accuracy	Correct / Total	Overall recognition
Precision	$TP/(TP+FP)$	Prediction reliability
Recall	$TP/(TP+FN)$	Class coverage
F1-score	$2PR/(P+R)$	Precision–recall balance
Macro average	Mean across classes	Equal class importance
Confusion matrix	Class-by-class counts	Error structure

C. Robustness Evaluation

In addition to benchmark testing, a practical classifier should be evaluated under changed lighting, different backgrounds, partial occlusion, different cameras, and multiple objects. The retail study demonstrates why this matters: controlled backgrounds can make benchmark classification easier than supermarket recognition [3]. A strong evaluation therefore includes an external test set or a domain-shift experiment when deployment is intended beyond the benchmark environment.

RESULTS AND DISCUSSION

The source material and literature used for this paper do not provide a new experimental run under a single common protocol. Therefore, this section reports and compares published results rather than presenting them as original experimental measurements. This distinction is important for academic integrity and reproducibility.

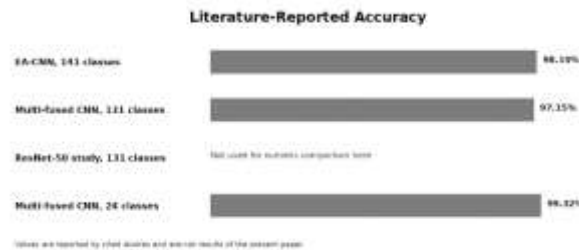


Fig. 4. Selected literature-reported accuracy values; datasets and protocols differ, so values are not directly comparable as a single ranking [6], [7]. [6,7]

TABLE IV. LITERATURE-REPORTED CLASSIFICATION PERFORMANCE [6], [7]

Model / study	Classes	Dataset	Accuracy
EA-CNN [6]	141	Fruit-360	98.1%
EA-CNN validation [6]	15	Fruit Recognition	96.0%
Multi-fused CNN [7]	131	Fruit-360	97.15%
Multi-fused CNN [7]	24	Fruit-360 setting	99.32%

The 98.1% EA-CNN result demonstrates the benefit that attention and mixed pooling can provide in a controlled 141-class benchmark [6]. The multi-fused CNN result of 97.15% on 131 classes suggests that combining EfficientNet-B0, MobileNetV2, and ResNet50V2 can produce strong multi-class representations [7]. The 99.32% result for a 24-category setting is higher, but the smaller class count makes it unsuitable for direct comparison with 131- or 141-class experiments [7].

The important observation is therefore not that one model is universally superior, but that performance depends on dataset version, class granularity, background conditions, split protocol, and architecture. A model can obtain excellent benchmark accuracy while still struggling with real photographs. The controlled nature of Fruits-360 is explicitly identified as a limitation in retail-oriented research [3].

A. Class Confusion

Fine-grained recognition is difficult when classes share visual characteristics. Different apple varieties, tomato varieties, pears, grapes, peppers, or citrus fruits may have similar color and shape. Confusion can also arise when maturity changes surface appearance. Class-wise precision, recall, F1-score, and confusion matrices are therefore necessary to understand failure modes.

B. Generalization and Domain Shift

Real deployment introduces backgrounds, shadows, reflections, packaging, multiple objects, camera differences, and occlusions. A model trained on centered objects may use background statistics as shortcuts. Domain adaptation, stronger augmentation, external datasets, and attention-based localization can reduce this risk. The multi-object resources maintained with Fruits-360 provide a useful direction for moving beyond single-object recognition [2].

C. Computational Performance

High-capacity networks can improve representation quality but increase memory and inference cost. MobileNetV2 and compact CNNs are attractive for edge devices, whereas larger ResNet and multi-model



fusion systems may be more appropriate for server-side inference. Compression techniques such as quantization, pruning, and knowledge distillation can help translate benchmark models into practical applications.

APPLICATIONS

Fruit and vegetable classification can support automated checkout, intelligent weighing systems, supermarket inventory, warehouse sorting, robotic handling, post-harvest grading, and educational applications. In a self-checkout system, an image classifier can identify an item and associate the predicted class with a product code. In sorting, classification can be combined with object detection and robotic actuators.

Agricultural applications can extend classification to maturity, quality, defect, freshness, and disease recognition. However, these tasks should not be conflated: recognizing a fruit variety is different from determining freshness or diagnosing disease. Multi-task learning can share image features while producing separate outputs, provided that suitable labels are available.

Explainable classification is especially valuable in automated systems because operators need confidence that predictions are based on the product rather than irrelevant visual context. Class-activation maps or related explanation methods can be integrated into decision-support interfaces [6].

LIMITATIONS

The principal limitation of the benchmark-based approach is domain shift. Controlled images with centered objects and simple backgrounds do not represent every environment in which the classifier may be deployed. The benchmark can therefore overestimate practical performance. The current dataset repository itself distinguishes single-object resources from multi-object resources, reinforcing the importance of testing more complex scenes [2].

A second limitation is class granularity. Treating varieties of the same fruit as separate classes increases the number of decision boundaries. This is useful for fine-grained research but may not match the requirements of every application. A production system may instead use hierarchical classification: first fruit versus vegetable, then broad category, then variety.

A third limitation is that reported accuracies across studies are not directly comparable. Differences in class count, dataset version, split, preprocessing, augmentation, pretrained weights, and test protocol can change results. For this reason, this paper deliberately labels literature numbers as reported results rather than presenting them as a new experiment.

FUTURE SCOPE

Future work should move toward more realistic and diverse datasets containing multiple objects, natural backgrounds, different camera devices, varied lighting, and partial occlusions. Domain adaptation can be used to transfer knowledge from controlled benchmark images to retail or field environments. Self-supervised learning is another promising direction because it can exploit large collections of unlabeled fruit images before supervised fine-tuning.

Attention-based and vision-transformer architectures can be investigated for fine-grained recognition, while lightweight CNNs remain important for mobile and edge deployment. Model quantization, pruning,



knowledge distillation, and hardware-aware neural architecture search can reduce inference latency and memory requirements.

Multimodal systems represent another promising direction. Image features can be combined with weight, size, color measurements, barcode or product metadata, and environmental sensors. Such systems may be more reliable for automated retail and sorting than image-only classifiers. The Fruits-360 repository also provides metadata and multi-object resources that can support research beyond conventional single-object classification [2].

Finally, future studies should report uncertainty, class-wise performance, calibration, robustness under domain shift, and independent external validation. This will make it possible to determine whether a model is genuinely suitable for deployment rather than simply optimized for a controlled benchmark.

CONCLUSION

This paper presented an IEEE-style deep learning framework for classification of fruits and vegetables using CNNs and transfer learning. The methodology includes dataset documentation, preprocessing, augmentation, CNN feature extraction, transfer learning, attention, explainability, reproducible training, and comprehensive evaluation. Fruits-360 provides a strong benchmark for multi-class recognition, and published studies demonstrate high accuracy using attention-based and multi-model architectures [1], [2], [6], [7].

At the same time, benchmark accuracy should not be interpreted as universal real-world accuracy. Controlled backgrounds, class granularity, similar varieties, and limited scene variation can make the benchmark easier than practical deployment. Future systems should therefore combine strong deep representations with realistic datasets, external validation, explainability, efficient inference, and domain adaptation. This direction can transform fruit and vegetable classification from a benchmark computer-vision task into a dependable component of intelligent retail, sorting, and agricultural automation.

REFERENCES

- [1] H. Muresan and M. Oltean, "Fruit recognition from images using deep learning," *Acta Universitatis Sapientiae, Informatica*, vol. 10, no. 1, pp. 26–42, 2018.
- [2] M. Oltean, "Fruits-360 dataset," 2017–2026, Fruits-360 dataset repository.
- [3] M. Kurtulmuş, "Fruit Classification for Retail Stores Using Deep Learning," 2020.
- [4] S. Sakib, Z. Ashrafi, and M. A. B. Siddique, "Implementation of Fruits Recognition Classifier using Convolutional Neural Network Algorithm for Observation of Accuracies for Various Hidden Layers," *arXiv:1904.00783*, 2019.
- [5] D. Zhang, "Fruit 360 classification based on the convolutional neural network," *Highlights in Science, Engineering and Technology*, 2023.
- [6] "EA-CNN: Enhanced attention-CNN with explainable AI for fruit and vegetable classification," *Heliyon*, vol. 10, no. 23, e40820, 2024.
- [7] "A multi-fused convolutional neural network model for fruit image classification," *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 416–424, 2024.
- [8] T.-F. L. Isaac, J.-Y. Lim, Y.-X. Tan, K.-M. Lim, C.-P. Lee, and Pa-Pa-Min, "Transfer Learning-based Fruit Image Recognition," *Proc. IEEE ICSPC*, pp. 47–52, 2024, doi:



10.1109/ICSPC63060.2024.10862682.

- [9] L. T. Duong, P. T. Nguyen, C. Di Sipio, and D. Di Ruscio, "Automated fruit recognition using EfficientNet and MixNet," *Computers and Electronics in Agriculture*, vol. 171, 105326, 2020.
- [10] R. Siddiqi, "Effectiveness of transfer learning and fine tuning in automated fruit image classification," *Proc. 2019 3rd Int. Conf. Deep Learning Technologies*, pp. 91–100, 2019.
- [11] D. Zhu, M. Wang, Q. Zou, D. Shen, and J. Luo, "Research on fruit category classification based on convolution neural network and data augmentation," *IEEE 13th Int. Conf. Anti-Counterfeiting, Security, and Identification*, pp. 46–50, 2019.
- [12] Z. Huang, Y. Cao, and T. Wang, "Transfer learning with efficient convolutional neural networks for fruit recognition," *IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference*, pp. 358–362, 2019.
- [13] B. Biswas, S. K. Ghosh, and A. Ghosh, "A robust multi-label fruit classification based on deep convolution neural network," *Computational Intelligence in Pattern Recognition*, pp. 105–115, 2019.
- [14] L. T. Duong et al., "Automated fruit recognition using EfficientNet and MixNet," *Computers and Electronics in Agriculture*, vol. 171, 105326, 2020.
- [15] H. Naidu, S. Rajkumar, K. C. Santosh, and P. V. S. S. R. Chandra Mouli, "A fast and efficient convolutional neural network for fruit recognition and classification," *Proc. Int. Conf. Recent Trends in Image Processing and Pattern Recognition*, 2022.
- [16] V. Khullar et al., "Investigating efficacy of transfer learning for fruit classification," *Machine Intelligence and Data Science Applications*, pp. 419–426, 2022.
- [17] S. Sultana et al., "XAI-FruitNet: An explainable deep model for accurate fruit classification," *Journal of Agriculture and Food Research*, vol. 18, 101474, 2024.
- [18] A. Dkhar and R. K. Bania, "Fruit and Vegetable Recognition System Using Deep Transfer Learning Approach," *NIELIT's International Conference on Communication, Electronics and Digital Technologies*, 2024.