

A Robust Deep Learning Framework for Multi-Class Fruit and Vegetable Classification Using Optimized AlexNet


 Batool Abdulsatar Abdulghani^{1,2*}, Adnan Ismail Al-Sulaifanie², Ahmed Khorsheed Mohammed²
¹ Department of Cybersecurity Engineering, Technical College of Engineering, Duhok Polytechnic University, Duhok 42001, Iraq

² Department of Electrical and Computer Engineering, College of Engineering, University of Duhok, Duhok 42001, Iraq

Corresponding Author Email: batool.abdulghani@dpu.edu.krd

 Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420623>

ABSTRACT

Received: 28 August 2025

Revised: 19 November 2025

Accepted: 1 December 2025

Available online: 31 December 2025

Keywords:
ANN, deep learning, alexnet, fruit classification, LeakyReLU, class imbalance

Automated sorting of fruits and vegetables is critical in modern agriculture and industry. However, this task faces a number of challenges, including the need for high classification accuracy, real-time processing speed, interclass similarities like the resemblance of chili peppers and bell peppers, and intraclass variability like the size and color differences among class apples. To tackle these challenges, this study implemented changes to the pre-trained AlexNet deep learning model, in which the first seven layers were frozen for feature extraction, replacing ReLU activations with LeakyReLU to improve discrimination of visually similar species, and class-weighted loss concerning imbalance among underrepresented classes like ginger (68 samples) and orange (69 samples). The model achieved 98.04% accuracy on the 36-class dataset (3,818 images), demonstrating a 2.47% improvement over the baseline AlexNet (95% confidence interval [1.12%, 4.19%]) and a 56.2% reduction in classification errors. As a side effect, computational efficiency improved, achieving 127.13 images per second for training and 55.45 images per second for testing on GPU hardware, demonstrating an optimal balance of performance and efficiency for practical deployment. This study revealed a solution for automated sorting of produce, where accuracy, morphological ambiguities, and operational speed posed critical constraints.

1. INTRODUCTION

The classification of fruits and vegetables is an innovation in agricultural technology that improves food security, supply chain management, and waste minimization [1]. Human errors in manual sorting in the fresh produce industry account for approximately 30-40% of post-harvest losses [2]. Applications for recognition of fruits can remove the human component in fruit harvesting processes [3]. The multitude of fruit types makes their categorization difficult [4]. The efficiency of an image recognition system has been enhanced by recent developments in computer vision [5]. There is tremendous potential to solve these two persistent problems using deep learning-based computer vision systems. Firstly, interclass similarities, in which distinct species, such as bell and chili peppers, share visual look-alikes. Secondly, intraclass variability—differences in the same cultivar's appearance owing to environmental conditions—can be substantial [6]. The need for accurate classification systems dealing with fruits and vegetables has been spurred by the growing automation in food quality assessment, smart farming, and precision nutrition [7].

Current approaches using transfer learning with pretrained convolutional neural networks (CNNs), a category of Artificial Neural Network (ANN), have received scant

attention to accuracy on controlled datasets, while often overlooking critical operational requirements [8]. The impact of advanced CNN-based image processing frameworks has significantly advanced the field of image categorization [9]. In 2012, Krizhevsky [10] won the ImageNet competition with his model, AlexNet, which demonstrated outstanding performance in image categorization. This architecture marks a significant breakthrough in deep learning for imaging and has served as the basis for subsequent complex CNN architectures [11, 12].

Although the classification of fruits and vegetables has advanced with image processing and machine learning techniques, several areas remain that can be addressed by advanced deep learning frameworks such as AlexNet. The model implemented in reference [13] used color and texture feature extraction, PCA for feature selection, and classification with an artificial neural network (ANN). Even though it achieved a remarkable 98.3% accuracy, this approach relied heavily on handcrafted features, which cannot match the performance of deep learning algorithms that automatically extract features from images. With the success of CNNs across various domains, many researchers shifted their focus to fruit recognition using CNNs. Despite CNN being effective at solving the image recognition problem, it still has certain limitations: it does not require an additional feature extraction

process during classification; instead, the entire feature extraction and classification are performed in a single link. As a result, the model structure is straightforward and the implementation process is comparatively simple.

With a range of species and strong generalization, Zeng [14] classified 26 distinct fruits and vegetables using the VGG architecture. The classification accuracy was 95.6%, which satisfies the requirements for fruit classification but falls short in terms of speed and morphological ambiguities.

In a comparison of the AlexNet deep learning approach with traditional methods such as BP neural networks and SVM classifiers, Zhu et al. [15] demonstrated that deep learning outperformed conventional methods by a large margin, achieving 92.1% accuracy on the test set. They presented a deep learning-based vegetable image classification system using the AlexNet model, demonstrating high performance.

Divya Shree et al. [16] developed a two-phase model for fruit recognition and estimation of their nutritional value using a pre-trained AlexNet model. A CNN was used to classify fruits in the first phase, after which pertinent dietary information for the detected fruit was displayed. Its performance on 15 fruit categories was evaluated and about 91% accuracy was reported.

In reference [17], the results were compared with algorithms such as AlexNet, GoogLeNet, and ResNet-50, which were fine-tuned, along with metrics such as accuracy, precision, sensitivity, and specificity. The study showed that the new deep convolutional neural network, Fruit114Net, enabled efficient classification while being low-cost in terms of computational, training time, and resource requirements. As noted, however, the accuracy, precision, and sensitivity metrics reported were much lower than those achieved with fine-tuned models.

Fu et al. [18] reported that an optimized version of GoogLeNet achieved 96.88% accuracy during training and 98% during testing for fruit and vegetable classification. The optimization focused on parameter reduction and modifications to the Inception-based design, improving training speed from 11.38 images per second to 33.68 images per second. An efficient model for the classification and identification of eight types of date fruit was presented by Albarak et al. [19]. The proposed model is based on the MobileNetV2 architecture and achieved an accuracy of 99%, outperforming other well-known models such as AlexNet, VGG16, InceptionV3, and ResNet, but no details on testing and training speeds are provided.

Amin et al. [20] proposed a method based on Convolutional Neural Networks (CNNs) combined with transfer learning for automatic classification of fruit freshness. In this instance, AlexNet was applied for transfer learning and later fine-tuned for better results. Its application to three publicly available datasets demonstrated that the method achieved an average accuracy of over 99%. However, the study did not address speed of processing, intra-category ambiguity, visual variability, and the range of feature recognition for class imbalance, all of which are crucial for practical implementation.

Vegetable classification was performed by the study [21] using the AlexNet CNN model and block-based compressive sensing (CS) to enhance computational and storage efficiency. A standalone AlexNet model achieved 98% maximum test accuracy. With block-based CS, the performance declined to 96.66%. There was no mention of dealing with interclass resemblance or intraclass variation in focus on fruit

classification, particularly within the context of fruit classification.

Thirumalraj et al. [22] developed an automated fruit recognition system that employs a modified AlexNet model for feature extraction and an FSSATM (Fruit Shift Self-Attention Transform Mechanism) classifier. This system sought to improve classification performance on the comprehensive Fruit-360 dataset, which contains 90,483 images across 131 fruit classes. The model achieving 98% accuracy is impressive; however, it did not report training or testing times, nor did it appear to address concerns about category or contour overlap, or variation in the features of samples from a single class.

A recent study [23] compared traditional machine learning and deep learning approaches for fruit recognition and classification using the Fruit-360 dataset. The authors achieved 99.85% accuracy with the AlexNet model. However, while the approach demonstrated strong classification performance, it did not examine critical deployment challenges, such as processing speed and visual resemblance across different fruit types or appearance inconsistencies within the same category—factors that significantly affect real-world classification systems.

Kurniawan et al. [24] used the deep Convolutional Neural Network AlexNet to improve the efficiency and accuracy of palm oil fruit maturity classification. Though the model achieved an accuracy of 0.9962, which met classification benchmarks, it fell short in processing speed and in handling morphological ambiguities.

Recent studies have demonstrated the performance gains possible with increasingly complex architectures. Notably, Wang et al. [25] proposed an advanced framework for eggplant disease detection based on multimodal data fusion and an embedding attention mechanism. Their method integrates image data with environmental sensor inputs using a dual-stream network (ResNet-50 and BERT) and dynamically weights features via a specialized attention module. Although this approach achieves notable accuracy, it also exemplifies a trend toward high computational complexity.

This work solves all these problems, creating a landmark achievement in the automatic classification of fruits and vegetables by addressing a triad of the most important and closely related classification problems: achieving real-time classification accuracy, processing speed, and tolerance to similarity in classification features within a class and between classes. Unlike prior approaches that addressed accuracy, speed, or class imbalance in isolation. As an extension to the previous study and finding a better classification system, an innovative optimization of the AlexNet architecture has been investigated to achieve superior performance on a complex and imbalanced dataset containing 36 classes with substantial variations in sample sizes, background conditions, and object density, like multiple peppers appearing in a single image. Given the modifications made to the baseline AlexNet, the main key contributions of this work can be summed up as follows:

- A modified AlexNet architecture is designed and validated for an end-to-end classification system, which is particularly customized for surmounting the core challenges in fruit and vegetable imagery: high inter-class similarity, significant intra-class variability, and the practical necessity for real-time processing,

- It is shown that with a strategic and holistic modification,

a classic CNN can achieve state-of-the-art performance on a complex 36-class dataset while retaining superior computational efficiency compared to much larger modern architectures. This was achieved through the integrated use of layer freezing, LeakyReLU activations, and a class-weighted loss function,

- Extensive empirical analysis is provided that encompasses per-class metrics, computational complexity, and thorough statistical validation.

This paper is organized as follows: the proposed system is described in Section 2; Section 3 presents the experimental results and their analysis; and Section 4 offers the conclusion and an outlook on limitations and future research directions.

2. THE PROPOSED SYSTEM DESIGN

The evolution of an actual fruit and vegetable classification

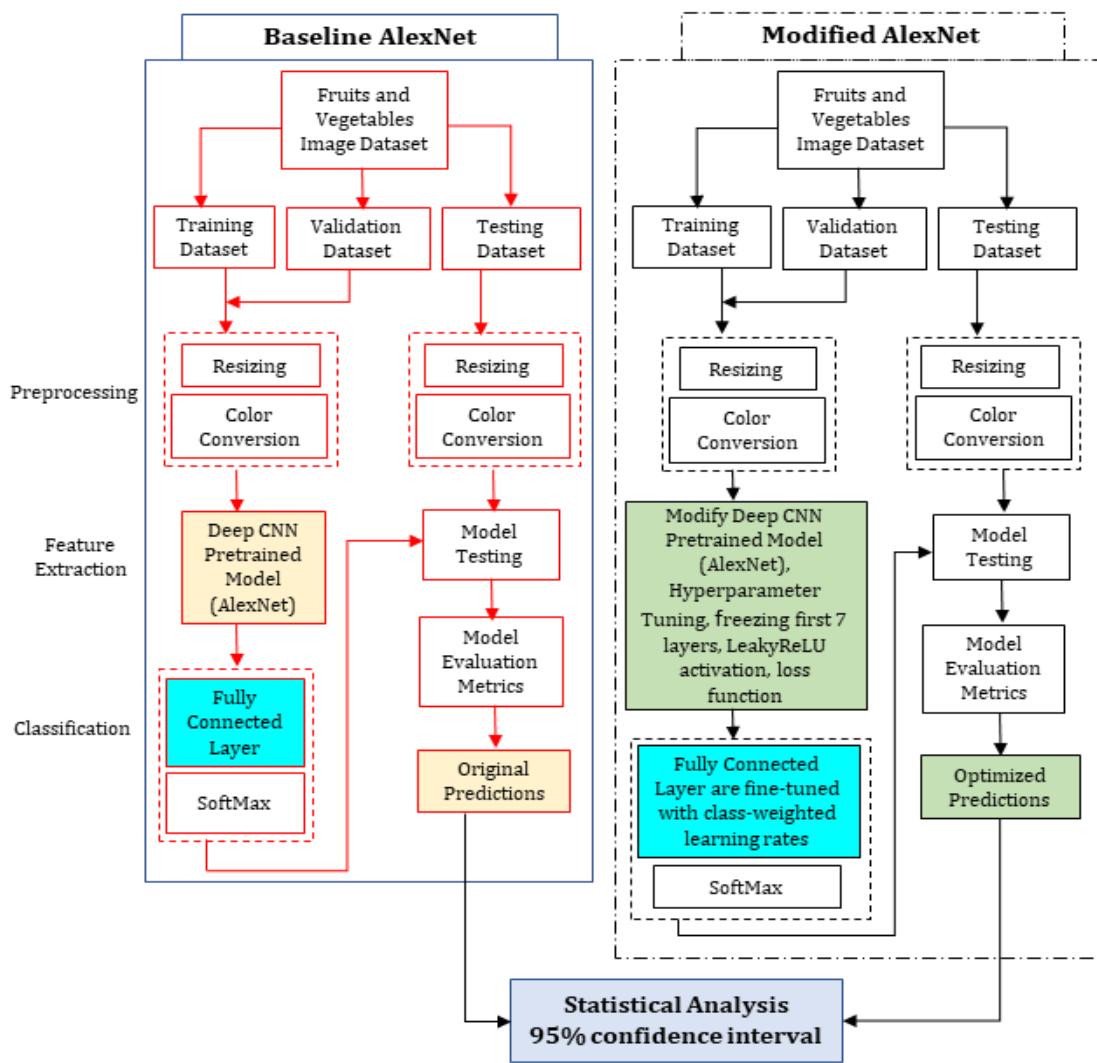


Figure 1. Diagram for the proposed study

2.1 Dataset

The dataset used in this study was obtained from the publicly available Kaggle dataset titled "Fruits and Vegetables Image Recognition Dataset" [26]. It comprised 3,818 images, categorized into 36 distinct classes of fruits and vegetables. These categories included the following fruit classes: banana,

system required a holistic approach following a conventional deep learning pipeline, including three essential stages: data preprocessing, network modification, and tailored training protocols. This section outlines the systematic design of the proposed solution, which incorporates key enhancements to optimize performance on a 36-class agricultural dataset. Selected changes were made to the baseline AlexNet architecture to improve feature discrimination among closely related species while enhancing computational efficiency. These changes benefited to some degree both feature discrimination and robustness to overfitting for visually similar produce. Diverse categories of produce were subjected to a thorough training regimen to obtain stable model performance with strong convergence. These elements, together, designed a complete system that could solve practical agricultural automation problems as accurately and quickly as possible. The complete system design is illustrated in Figure 1.

apple, pear, grapes, orange, kiwi, watermelon, pomegranate, pineapple, and mango; and the following vegetable classes: cucumber, carrot, capsicum, onion, potato, lemon, tomato, radish, beetroot, cabbage, lettuce, spinach, soybean, cauliflower, bell pepper, chili pepper, turnip, corn, sweetcorn, sweet potato, paprika, jalapeño, ginger, garlic, peas, and eggplant. Sample images from the dataset are shown in Figure

2, illustrating the visual diversity across fruit and vegetable categories, including varying object sizes, background complexities, and lighting conditions. The dataset demonstrated evident intraclass variability in the spatial arrangement of objects within images, with certain classes, such as eggplant and onions, containing both single and grouped items. While this non-uniformity could, in theory, bias feature learning toward object scale or background context, it ultimately enhanced the model's robustness to real-world agricultural conditions.

As illustrated in Figure 3, the dataset was split into three

subsets: 3,110 images for training, 350 for validation, and 358 for testing. Some issues were resolved before the actual training began. Each image was transformed to 227 by 227 so that it could be used by the AlexNet architecture. Additionally, grayscale images were converted to RGB to provide a consistent three-channel input. A class distribution analysis was performed to assess and address class imbalance. Moreover, the training dataset was randomly shuffled during training to reduce potential sequence bias and improve generalization. The detailed distribution of images across all categories and dataset splits is provided in Table 1.

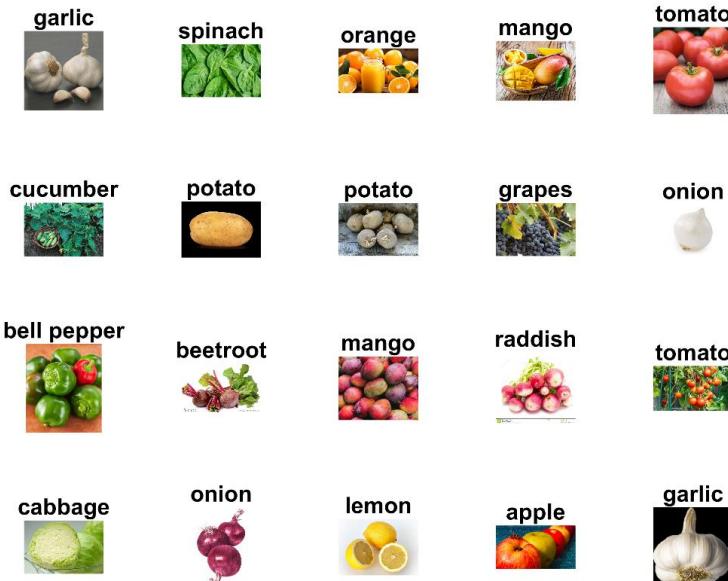


Figure 2. Sample images from selected fruit and vegetable categories in the dataset

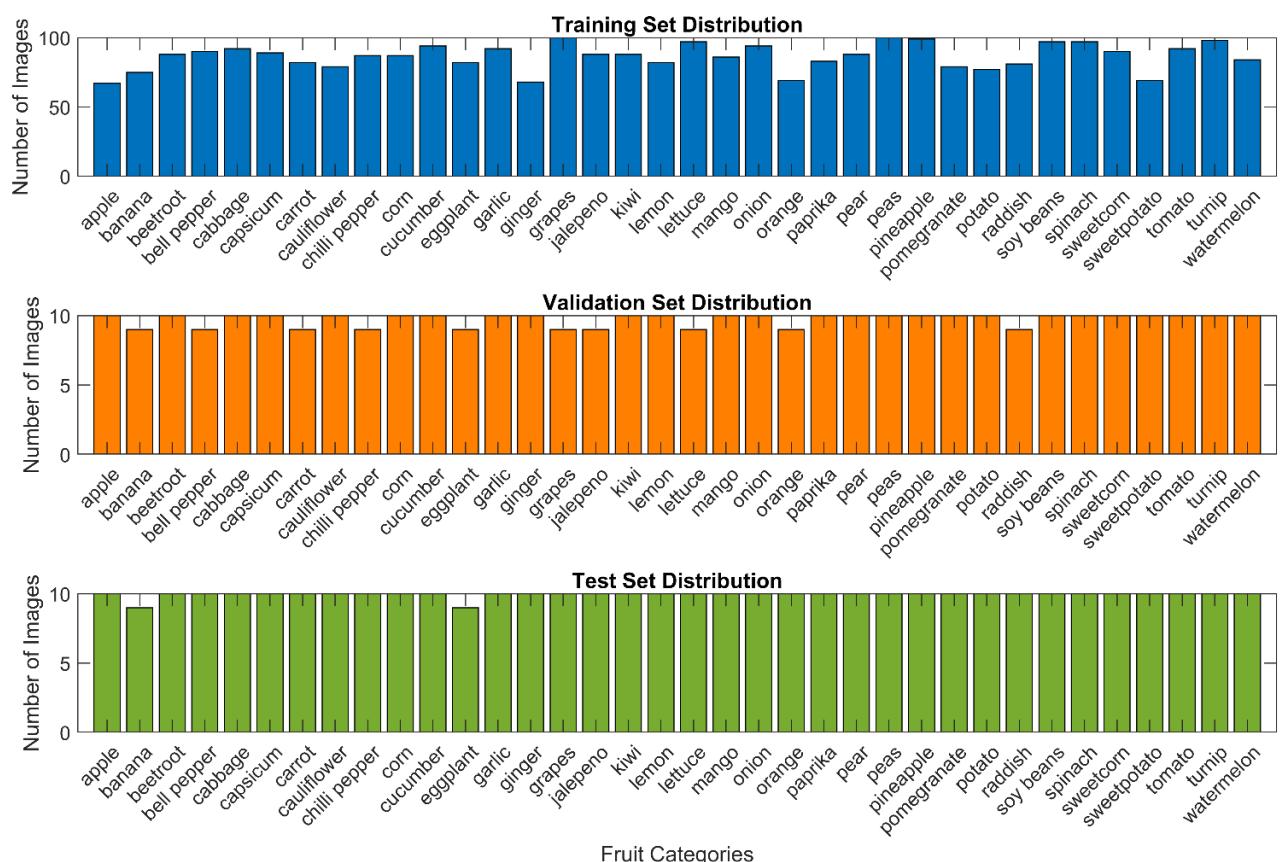


Figure 3. The number of images per class in the training, validation and testing sets

Table 1. Class distribution of training, validation, and testing sets

Category	Training	Validation	Testing	Total	Category	Training	Validation	Testing	Total
peas	100	10	10	120	corn	87	10	10	107
pineapple	99	10	10	119	jalapeño	88	9	10	107
grapes	100	9	10	119	chili pepper	87	9	10	106
turnip	98	10	10	118	mango	86	10	10	106
soy beans	97	10	10	117	watermelon	84	10	10	104
spinach	97	10	10	117	paprika	83	10	10	103
lettuce	97	9	10	116	lemon	82	10	10	102
cucumber	94	10	10	114	carrot	82	9	10	101
onion	94	10	10	114	radish	81	9	10	100
tomato	92	10	10	112	eggplant	82	9	9	100
garlic	92	10	10	112	pomegranate	79	10	10	99
cabbage	92	10	10	112	cauliflower	79	10	10	99
sweetcorn	90	10	10	110	potato	77	10	10	97
capsicum	89	10	10	109	banana	75	9	9	93
bell pepper	90	9	10	109	sweet potato	69	10	10	89
pear	88	10	10	108	orange	69	9	10	88
kiwi	88	10	10	108	ginger	68	10	10	88
beetroot	88	10	10	108	apple	67	10	10	87
Total Dataset Images				3818					

2.2 AlexNet architecture modification

The AlexNet architecture is considered the first deep convolutional neural network that was used for image classification in large databases. It consisted of five convolutional and max-pooling layers, followed by three dense layers, including fully connected layers with ReLU

activation. It is worth noting that the neural network was proposed by Krizhevsky et al. [10] in 2012. Although baseline models are helpful for many applications, in the case of fruit and vegetable classification, they were not valuable, as they were unable to discriminate fine-grained features due to the lack of class imbalance.

Table 2. Layer configuration of the proposed modified AlexNet-based classifier

Layer	Name	Type	Activations
1	data 227×227×3 images with ‘zerocenter’ normalization	Image Input	227(S) × 227(S) × 3(C) × 1(B)
2	conv1 96 11×11×3 convolutions with stride [4 4] and padding [0 0 0]	2-D Convolution	55(S) × 55(S) × 96(C) × 1(B)
3	relu1_leaky LeakyReLULayer	LeakyReLULayer	55(S) × 55(S) × 96(C) × 1(B)
4	norm1 cross channel normalization with 5 channels per element	Cross Channel Normalization	55(S) × 55(S) × 96(C) × 1(B)
5	pool1 3×3 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	27(S) × 27(S) × 96(C) × 1(B)
6	conv2 2 groups of 128 5×5×48 convolutions with stride [1 1] and padding [2 2 2]	2-D Grouped Convolution	27(S) × 27(S) × 256(c) × 1(B)
7	relu2_leaky LeakyReLULayer	LeakyReLULayer	27(S) × 27(S) × 256(C) × 1(B)
8	norm2 cross channel normalization with 5 channels per element	Cross Channel Normalization	27(S) × 27(S) × 256(C) × 1(B)
9	pool2 3×3 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	13(S) × 13(S) × 256(C) × 1(B)
10	conv3 384 3×3×256 convolutions with stride [1 1] and padding [1 1 1 1]	2-D Convolution	13(S) × 13(S) × 384(C) × 1(B)
11	relu3_leaky LeakyReLULayer	LeakyReLULayer	13(S) × 13(S) × 384(C) × 1(B)
12	conv4 2 groups of 192 3×3×192 convolutions with stride [1 1] and padding [1 1 1 1]	2-D Grouped Convolution	13(S) × 13(S) × 384(C) × 1(B)
13	relu4_leaky LeakyReLULayer	LeakyReLULayer	13(S) × 13(S) × 384(C) × 1(B)
14	conv5 2 groups of 128 3×3×192 convolutions with stride [1 1] and padding [1 1 1 1]	2-D Grouped Convolution	13(S) × 13(S) × 256(C) × 1(B)
15	relu5_leaky LeakyReLULayer	LeakyReLULayer	13(S) × 13(S) × 256(C) × 1(B)
16	pool5 3×3 max pooling with stride [2 2] and padding [0 0 0 0]	2-D Max Pooling	6(S) × 6(S) × 256(C) × 1(B)
17	fc6 4096 fully connected layer	Fully Connected	1(S) × 1(S) × 4096(C) × 1(B)
18	relu6_leaky LeakyReLULayer	LeakyReLULayer	1(S) × 1(S) × 4096(C) × 1(B)
19	drop6 50% dropout	Dropout	1(S) × 1(S) × 4096(C) × 1(B)
20	fc7 4096 fully connected layer	Fully Connected	1(S) × 1(S) × 4096(C) × 1(B)
21	relu7_leaky LeakyReLULayer	LeakyReLULayer	1(S) × 1(S) × 4096(C) × 1(B)
22	drop7 50% dropout	Dropout	1(S) × 1(S) × 4096(C) × 1(B)
23	Fruit Feature Learner 36 fully connected layer	Fully Connected	1(S) × 1(S) × 36(C) × 1(B)
24	Prob softmax	Softmax	1(S) × 1(S) × 36(C) × 1(B)
25	Fruit Classifier Class weighted crossentropy with ‘apple’ and 35 other classes	Classification Output	1(S) × 1(S) × 36(C) × 1(B)

To address these limitations, the AlexNet architecture was modified (Table 2) with three key adaptations. First, all standard ReLU activations were replaced with LeakyReLU units ($\alpha=0.01$) to improve gradient propagation and mitigate

neuron saturation. Secondly, the last fully connected layer was modified to support 36 output classes, and class-weighted loss was used to address dataset imbalance. Thirdly, a transfer learning approach was used, in which low-level feature

extraction in the first seven convolutional layers was preserved by freezing those layers. In contrast, the remaining layers were trained to domain-specific features. These modifications enhanced feature discrimination for agricultural imagery while maintaining computational efficiency. The use of LeakyReLU ensured stable gradient flow during backpropagation, particularly for inputs that were negatively activated. Additionally, the class-weighted loss function effectively balanced learning across underrepresented categories in the training dataset.

2.3 Training protocol

The modified architecture was trained using an Adam optimizer with an initial learning rate of 3×10^{-4} , which was reduced by a factor of 0.1 every five epochs. This approach balanced rapid convergence during early training stages with precise weight updates in later phases. Validation metrics were reviewed after every epoch, and an accuracy-improving validation window of 10 iterations triggered early stopping to avoid overfitting. A mini-batch size of 128 was used, combined with L2 regularisation ($\lambda=0.0005$) to combat overfitting. A combination of early-layer freezing and deeper-layer fine-tuning achieved effective transfer learning, reducing trainable parameters by 58% compared to full-network training. Convergence occurred between 12 and 15 epochs, at which point the training loss was below 0.01, and the class-weighted loss function was stabilised and optimised.

2.4 Rationale for architectural selection and integration

The specific changes to the AlexNet architecture were guided by a coherent design philosophy tailored to the application at hand. Equally, basing the system on AlexNet is motivated by the requirement for a lean, efficient architecture suitable for edge deployment—one that strikes an outstanding balance between representational power and computational footprint compared to more modern, more computationally expensive networks.

In this context, the changes had specific purposes. Freezing the first seven convolutional layers served a dual purpose. First, it greatly reduces the number of parameters to train, thereby speeding up training and reducing the risk of overfitting when dealing with a relatively small dataset. Second, it keeps intact the generic low-level feature detectors, which are edge and texture filters, for instance, which had been learned from ImageNet and maintain their universal value on agricultural scenes with no need to relearn them.

Replacing the default ReLU activation with LeakyReLU was essential to prevent dying neurons, which would have been highly undesirable in a network fine-tuned as deeply as this one. Such a change ensures that the gradient flow during backpropagation remains consistent when learning minor differences in features between visually quite similar classes, such as bell peppers and chilli peppers.

Simultaneously, the natural imbalance in this dataset, such as 68 samples of ginger and 120 peas, would naturally bias a standard model towards the majority class. Implementing a class-weighted loss function is a direct counterbalance to such biases, since it imposes a higher penalty for misclassifying instances from underrepresented classes, thereby forcing the model to be more sensitive to their discriminative features. Isolated, each technique addresses a specific weakness, while together they form a synergistic system that is robust, efficient,

and accurately tuned for the complexities of fruit and vegetable classification. This integrated approach ensures the model is not only accurate but also practical for real-world agricultural applications.

3. EXPERIMENTAL AND RESULT ANALYSIS

The experimental evaluation was conducted on a dataset of 3,818 high-resolution fruit and vegetable images spanning 36 categories to validate performance improvements achieved through architectural modifications to the AlexNet model. All implementations were developed in MATLAB R2023b using the Deep Learning Toolbox and executed on a high-performance workstation equipped with Windows 11 Professional 64-bit, a 13th Gen Intel Core i7-13700H processor (2.40GHz), 16 GB RAM, 500GB SSD storage, and NVIDIA GeForce RTX 4060 GPU (8GB VRAM) acceleration. A modified and a baseline architecture were implemented alongside each other, along with the additional criteria of training stability and computational efficiency, to form a comprehensive assessment framework.

The selection of the specific hyperparameters for the modified AlexNet architecture, particularly the LeakyReLU activation and the seven-layer freezing strategy, was not arbitrary but was determined through a structured ablation study. This study evaluated variants using ReLU and LeakyReLU ($\alpha=0.01$, $\alpha=0.1$), along with different layer-freezing strategies (3-9 frozen layers). The final combination of LeakyReLU ($\alpha=0.01$) and seven frozen layers emerged as the Pareto-optimal choice, delivering superior accuracy for fruit and vegetable recognition without the computational burden of training the whole network.

Model performance has significantly improved with the changes made. The modified architecture achieved a test accuracy of 98.04% on the evaluation dataset, representing a 2.51 percentage point increase over the baseline model's 95.53% accuracy.

This enhancement was statistically validated through bootstrap analysis, with the improvement falling within a 95% confidence interval of 1.12% to 4.19%. The model has proven more reliable, with the error rate reduced from 4.47 per cent to 1.96 per cent. These metrics are visually contrasted in Figure 4, which presents a direct comparison of original and optimized accuracy and error rates using a double-bar plot.

Training dynamics exhibited notable improvements, as depicted in Figure 5. Compared to the baseline architecture, the modified model achieved peak validation accuracy about 3 epochs earlier, demonstrating improved convergence. This accelerated training occurred alongside enhanced stability, with reduced fluctuation in loss values across epochs. The experiment addressed two fundamental challenges in agricultural image classification: interclass similarities among visually analogous produce such as bell peppers vs. chili peppers and intraclass variability stemming from natural differences in color, size, and morphology such as red vs. green apples. The modified AlexNet architecture demonstrated notable success in discriminating between taxonomically distinct but visually similar categories, as evidenced by the confusion matrix presented in Figure 6. While most class pairs achieved $\geq 99\%$ accuracy, residual misclassifications primarily occurred between produce with overlapping morphological features, such as tomatoes and cherries, where shape and texture differences were subtle.

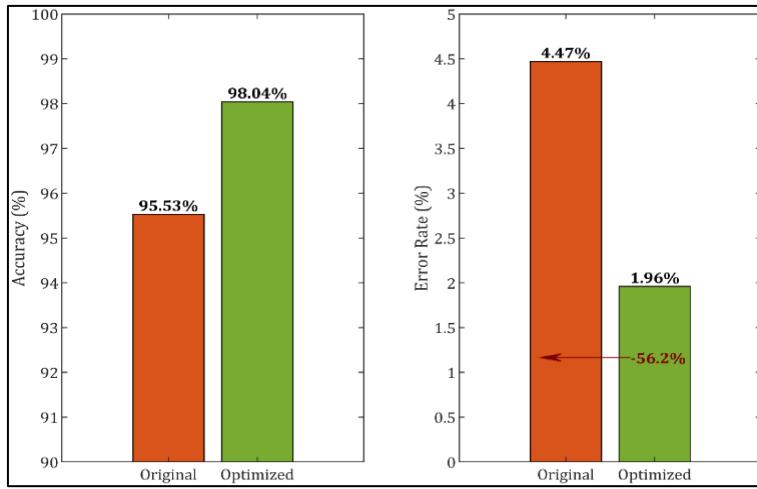


Figure 4. Comparison of original versus optimized accuracy and error rates

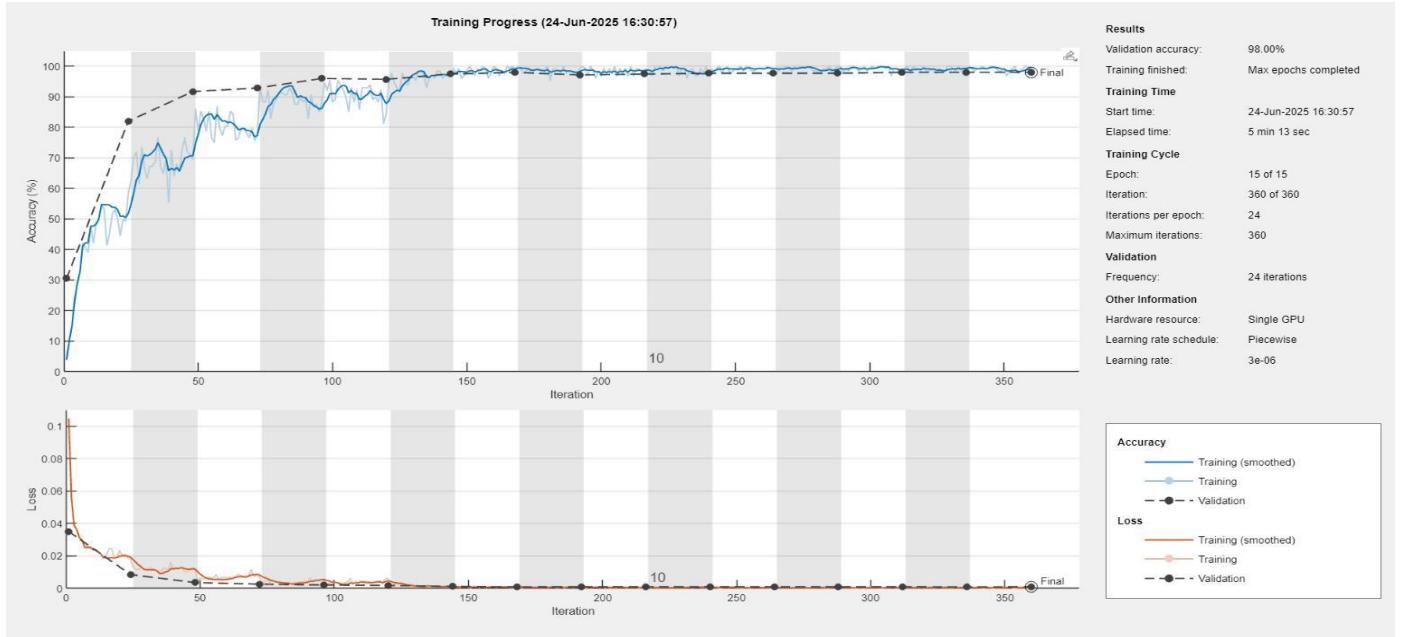


Figure 5. Training dynamics of the modified AlexNet for accuracy and loss curves across 15 epochs

Intraclass variability was effectively mitigated by combining frozen early layers (preserving generic feature extractors) with LeakyReLU activations, thereby maintaining gradient flow across diverse color and texture distributions within a single category. The class-weighted loss function further compensated for underrepresented classes, ensuring robust performance even for classes with limited training samples, such as ginger and orange. Considering the natural heterogeneity of lighting conditions, occlusion, and varying stages of produce maturation, these adaptations were crucial. As shown in Table 3, the modified architecture outperformed the baseline model across all evaluation metrics, demonstrating consistent improvement.

Computational efficiency metrics revealed substantial gains in processing speed. Training throughput increased by 123%, from 56.94 to 127.13 images per second, while testing speed improved by 4%, from 53.32 to 55.45 images per second. As the classification accuracy was preserved, these improvements highlighted the model's effectiveness for real-time deployment. The bootstrap distribution in Figure 7 shows the frequency of mean accuracy differences between the optimized and baseline models. The 95% confidence interval

for this difference (1.12% to 4.19%) excludes zero, indicating a statistically significant improvement at the $p < 0.05$ level and confirming that the observed gains are not due to random chance.

A broad class-wise performance evaluation was thus carried out to assess the effectiveness of the class-weighted loss function in mitigating dataset imbalance. As can be seen from Table 4, despite there being significant variation in the number of samples available to train each class (67-100 samples per class), the model ensured relatively strong performance in all 36 classes. For example, the smallest classes, such as ginger and orange, with 68 and 69 samples, respectively, attained 100% recall; whereas the largest classes, such as peas and grapes, each with 100 samples, also returned 100% recall. That is a very narrow margin of only 20.0% between the minimum and maximum recall values across the dataset. Furthermore, the proposed class-weighted loss function successfully mitigated imbalance; 25 out of 36 classes (69.4%) achieved perfect recall, while only 3 classes scored below 90%. This therefore proves that the approach presented is practical for addressing class imbalance while sustaining very high overall classification accuracy.

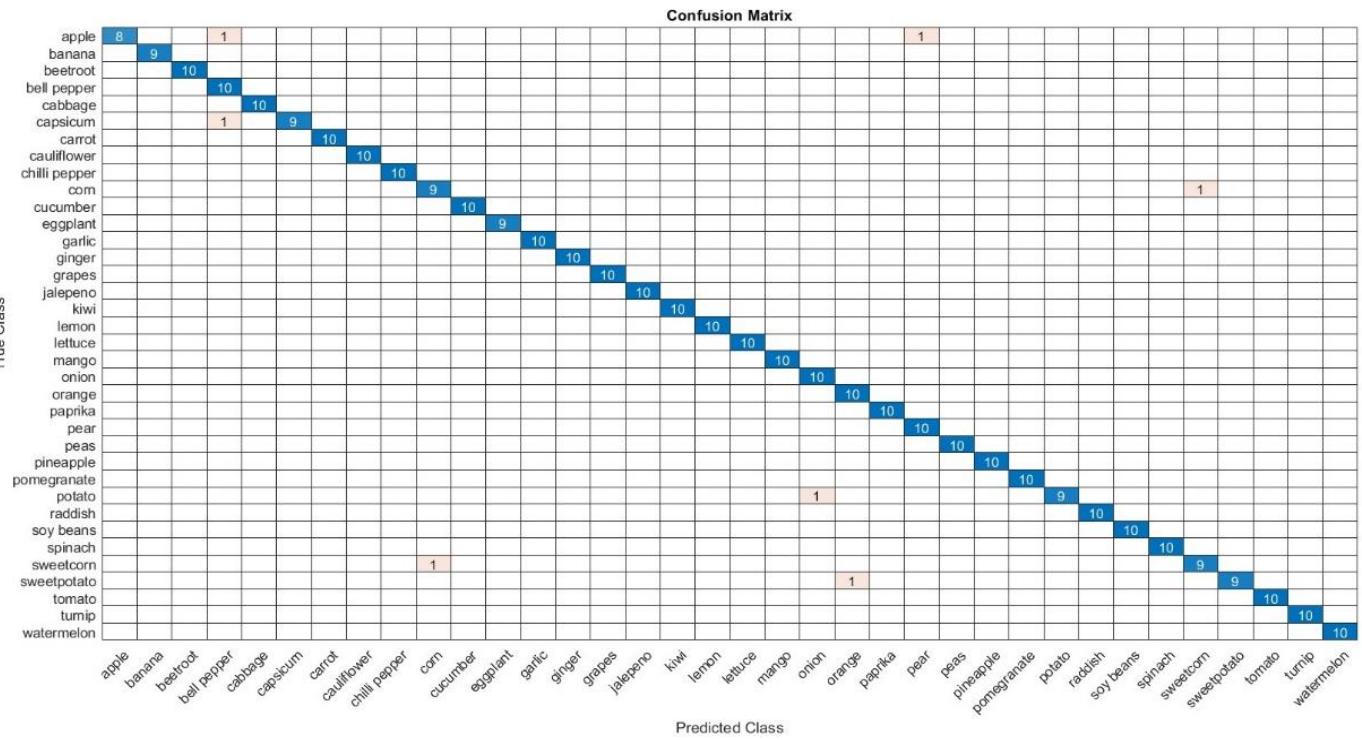


Figure 6. Confusion matrix for the modified AlexNet (Rows: True labels; Columns: Predicted labels)

Table 3. Comparative performance metrics of baseline versus modified AlexNet

Metric	Baseline Model	Modified Model	Improvement
Validation Accuracy (%)	95.43	98.00	+2.57
Training Accuracy (%)	87.50	98.44	+10.94
Training Speed (img/s)	56.94	127.13	+123%
Testing Accuracy (%)	95.53	98.04	+2.51
Testing Speed (img/s)	53.32	55.45	+4.0%
Error Rate (%)	4.47	1.96	-2.51
Macro Precision (%)	96.15	98.22	+0.0207
Macro Recall (%)	95.49	98.06	+0.0257
Macro F1-Score (%)	95.41	98.05	+0.0264
Weighted Precision (%)	96.13	98.21	+0.0208
Weighted Recall (%)	95.53	98.04	+0.0251
Weighted F1-Score (%)	95.42	98.04	+0.0262

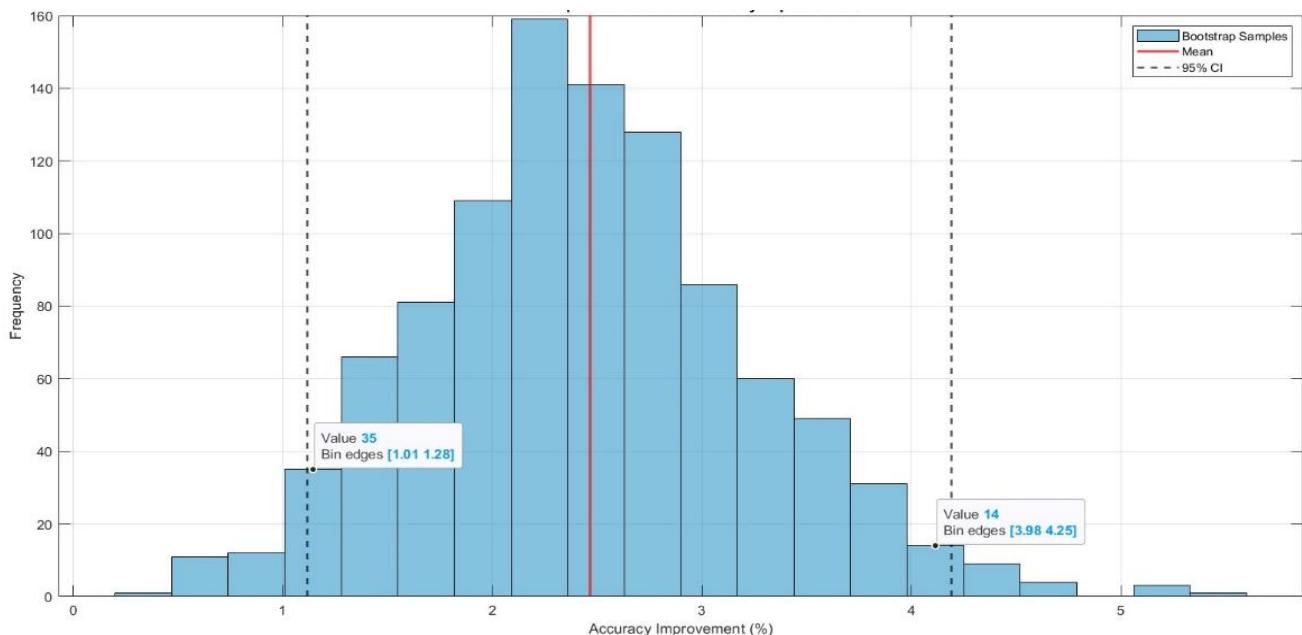


Figure 7. Bootstrap distribution of accuracy improvement

Table 4. Comprehensive class-wise performance analysis

Class	Training Samples	Precision	Recall	F1-Score	Class	Training Sample	Precision	Recall	F1-Score
apple	67	88.9	80	84.2	lettuce	97	100	100	100
banana	75	100	88.9	94.1	mango	86	100	100	100
beetroot	88	100	100	100	onion	94	100	100	100
bell pepper	90	81.8	90	85.7	orange	69	100	100	100
cabbage	92	100	100	100	paprika	83	100	100	100
capsicum	89	81.8	90	85.7	pear	88	100	100	100
carrot	82	100	100	100	peas	100	100	100	100
cauliflower	79	100	100	100	pineapple	99	100	100	100
chilli pepper	87	90.9	100	95.2	pomegranate	79	100	100	100
corn	87	88.9	80	84.2	potato	77	100	80	88.9
cucumber	94	100	100	100	raddish	81	100	100	100
eggplant	82	100	100	100	soy beans	97	100	100	100
garlic	92	100	100	100	spinach	97	100	100	100
ginger	68	100	100	100	sweetcorn	90	81.8	90	85.7
grapes	100	100	100	100	sweetpotato	69	100	90	94.7
jalepeno	88	100	100	100	tomato	92	100	100	100
kiwi	88	90.9	100	95.2	turnip	98	100	100	100
lemon	82	90.9	100	95.2	watermelon	84	100	100	100

Table 5. Cross-dataset performance on Fruits-360 test set

Class	Precision %	Recall %	F1 %	Support
banana	77	52.4	62.4	166
beetroot	88.8	52.7	66.1	150
cauliflower	-	0	-	234
corn	100	30	46.2	150
cucumber	-	0	-	130
eggplant	89.1	100	94.3	156
kiwi	0	0	-	156
lemon	48.4	45.1	46.7	164
mango	38.6	81.9	52.5	166
orange	100	48.1	65	160
pear	15.9	6.1	8.8	164
pineapple	0	0	-	166
Overall Accuracy %		33.8		

Cross-dataset validation using the publicly available Fruits-360 dataset [27], which features distinct imaging conditions with uniform white backgrounds compared to the primary

training dataset [26] used in this work, quantitatively characterized the model's generalization boundaries. The evaluation achieved 33.83% accuracy on 15 overlapping classes (Apple, Banana, Beetroot, Cauliflower, Corn, Cucumber, Eggplant, Kiwi, Lemon, Mango, Orange, Pear, Pineapple, Pomegranate, Watermelon) without fine-tuning. As detailed in Table 5, this analysis provides crucial benchmarking data that maps the domain-adaptation challenge in agricultural vision systems, offering transparent insights into the capabilities and limitations of deep learning approaches under significant domain shifts.

In addition to the baseline AlexNet, our modified model was rigorously evaluated against two other modern deep learning architectures: ResNet-18 and VGG-19. This comprehensive benchmarking was conducted to situate our model's performance within the current landscape properly. All models were trained and evaluated under identical conditions on the 36-class fruit-and-vegetable dataset.

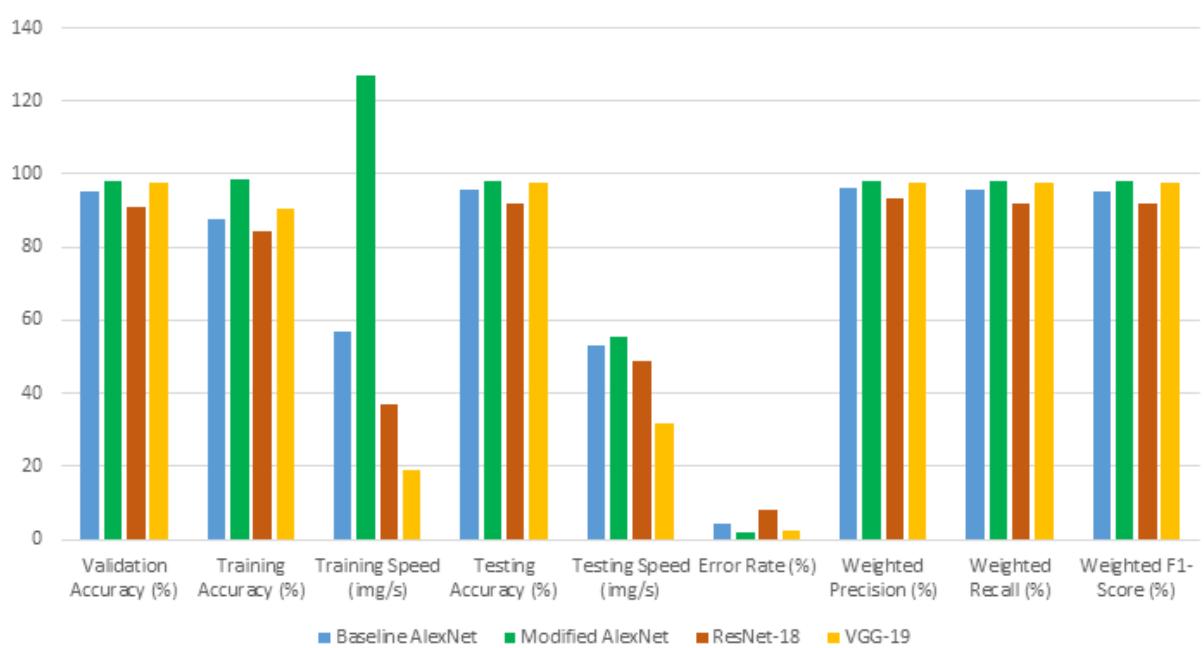


Figure 8. Comprehensive comparative analysis

The analysis reveals two key findings. First, our model demonstrates a superior trade-off between accuracy and efficiency. It achieves a top-tier testing accuracy of 98.04%, marginally outperforming VGG-19 (97.49%) and significantly surpassing ResNet-18 (91.9%) and the Baseline AlexNet (95.53%). Second, and more critically, this high accuracy is attained with exceptional computational efficiency. As Figure 8 illustrates, our Modified AlexNet occupies the most desirable position in the accuracy-speed space, delivering VGG-level accuracy but at a training speed more than six times faster. This efficiency stems directly from our design choices—the frozen layers and optimized architecture—which yield a model that is both powerful and practical, avoiding the computational overhead of the vastly larger VGG-19 network.

The overall results highlight that, with MATLAB's Deep Learning Toolbox and modern GPU hardware, the accuracy and efficiency of deep learning models for automated agricultural classification can be optimally improved through thoughtful changes to the model architecture. The improvements in training stability, classification performance, and computational speed position the modified AlexNet architecture as a viable solution for real-world fruit and vegetable recognition applications in precision agriculture and food quality control systems.

4. CONCLUSIONS

This study successfully addressed the problems of inter-class similarities and intraclass variability in the classification of fruits and vegetables by systematically altering the design. The model successfully distinguished visually similar items—such as bell and chili peppers—while accounting for natural variation within categories, achieving 98.04% test accuracy and better performance on minority-class samples. This success illustrates the significance of leveraging domain-specific adaptations, such as LeakyReLU for fine-grained feature exploitation, frozen layers for stability across diverse inputs, and weighted loss for countering imbalance, to agricultural computer vision systems. The practical viability of the solution was underscored by real-time processing capabilities and a testing speed of 55.45 images per second. Such features are essential for industries such as automated sorting where accuracy and efficiency are crucial.

This study demonstrates high-performance classification while honestly quantifying domain adaptation challenges through rigorous cross-dataset evaluation. The 33.83% accuracy on Fruits-360 establishes a critical baseline that advances the field by providing measurable generalization boundaries. Future research should build upon these empirical insights to develop domain-robust architectures capable of operating across diverse agricultural imaging conditions.

REFERENCES

[1] Calicioglu, O., Flammini, A., Bracco, S., Bellù, L., Sims, R. (2019). The future challenges of food and agriculture: An integrated analysis of trends and solutions. *Sustainability*, 11(1): 222. <https://doi.org/10.3390/su11010222>

[2] Lipinski, B., Hanson, C., Lomax, J., Kitinoja, L., Waite, R., Searchinger, T. (2013). Reducing food loss and waste (Working paper, installment 2 of creating a sustainable food future). Washington, DC: World Resources Institute. http://pdf.wri.org/reducing_food_loss_and_waste.pdf.

[3] Gracia, N.S., Gladston, A., Nehemiah, K.H. (2025). LwSANet: Light weight self-attention network model to recognize fruits from images. *Traitement du Signal*, 42(1): 183-200. <https://doi.org/10.18280/ts.420117>

[4] Alzubaidi, L., Al-Shamma, O., Fadhel, M.A., Arkah, Z.M., Awad, F.H. (2019). A deep convolutional neural network model for multi-class fruits classification. In *International Conference on Intelligent Systems Design and Applications*. Cham: Springer International Publishing, pp. 90-99. https://doi.org/10.1007/978-3-030-49342-4_9

[5] Zhai, H. (2016). Research on image recognition based on deep learning technology. In *2016 4th International Conference on Advanced Materials and Information Technology Processing (AMITP 2016)*. Atlantis Press, pp. 266-270. <https://doi.org/10.2991/amitp-16.2016.53>

[6] Jana, S., Parekh, R. (2016). Intra-class recognition of fruits using color and texture features with neural classifiers. *International Journal of Computer Applications*, 148(11): 1-6. <https://doi.org/10.5120/ijca2016911283>

[7] Hamid, N.N.A.A., Razali, R.A., Ibrahim, Z. (2019). Comparing bags of features, conventional convolutional neural network and AlexNet for fruit recognition. *Indonesian Journal of Electrical Engineering and Computer Science*, 14(1): 333-339. <https://doi.org/10.11591/ijeecs.v14.i1.pp333-339>

[8] Kamilaris, A., Prenafeta-Boldú, 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>

[9] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of The ACM*, 60(6): 84-90. <https://doi.org/10.1145/3065386>

[10] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.

[11] Omonigho, E.L., David, M., Adejo, A., Aliyu, S. (2020). Breast cancer: Tumor detection in mammogram images using modified alexnet deep convolution neural network. In *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Ayobo, Nigeria, pp. 1-6. <https://doi.org/10.1109/icmcecs47690.2020.240870>

[12] Shanthi, T., Sabeenian, R.S. (2019). Modified Alexnet architecture for classification of diabetic retinopathy images. *Computers & Electrical Engineering*, 76: 56-64. <https://doi.org/10.1016/j.compeleceng.2019.03.004>

[13] Septiarini, A., Sunyoto, A., Hamdani, H., Kasim, A.A., Utaminingrum, F., Hatta, H.R. (2021). Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features. *Scientia Horticulturae*, 286: 110245. <https://doi.org/10.1016/j.scienta.2021.110245>

[14] Zeng, G. (2017). Fruit and vegetables classification system using image saliency and convolutional neural network. In *2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)*, Chongqing, China, pp. 613-617. <https://doi.org/10.1109/itoec.2017.8122370>

[15] Zhu, L., Li, Z., Li, C., Wu, J., Yue, J. (2018). High performance vegetable classification from images based on AlexNet deep learning model. *International Journal of Agricultural and Biological Engineering*, 11(4): 217-223. <https://doi.org/10.25165/j.ijabe.20181104.2690>

[16] Divya Shree, B., Brunda, R., Shobha Rani, N. (2019). Fruit detection from images and displaying its nutrition value using deep Alex network. In *Soft Computing and Signal Processing: Proceedings of ICSCSP 2018*. Singapore: Springer Singapore, 2: 599-608. https://doi.org/10.1007/978-981-13-3393-4_61

[17] Orquia, J.J.D. (2020). Automated fruit classification using deep convolutional neural network. *Philippine Social Science Journal*, 3(2): 177-178. <https://doi.org/10.52006/main.v3i2.188>

[18] Fu, Y.S., Song, J., Xie, F.X., Bai, Y., Zheng, X., Gao, P. (2021). Circular fruit and vegetable classification based on optimized GoogLeNet. *IEEE Access*, 9: 113599-113611. <https://doi.org/10.1109/access.2021.3105112>

[19] Albarak, K., Gulzar, Y., Hamid, Y., Mehmood, A., Soomro, A.B. (2022). A deep learning-based model for date fruit classification. *Sustainability*, 14(10): 6339. <https://doi.org/10.3390/su14106339>

[20] Amin, U., Shahzad, M.I., Shahzad, A., Shahzad, M., Khan, U., Mahmood, Z. (2023). Automatic fruits freshness classification using CNN and transfer learning. *Applied Sciences*, 13(14): 8087. <https://doi.org/10.3390/app13148087>

[21] Irawati, I.D., Budiman, G., Saidah, S., Rahmadiani, S., Latip, R. (2023). Block-based compressive sensing in deep learning using AlexNet for vegetable classification. *PeerJ Computer Science*, 9: e1551. <https://doi.org/10.7717/peerj-cs.1551>

[22] Thirumalraj, M.A., Rajalakshmi, B., Kumar, B.S., Stephe, S. (2024). Automated fruit identification using modified AlexNet feature extraction based FSSATM classifier. *Research Square*. <https://doi.org/10.21203/rs.3.rs-4074664/v1>

[23] Salim, N.O., Mohammed, A.K. (2024). Comparative analysis of classical machine learning and deep learning methods for fruit image recognition and classification. *Traitement du Signal*, 41(3): 1331-1343. <https://doi.org/10.18280/ts.410322>

[24] Kurniawan, R., Samsuryadi, S., Mohamad, F.S., Wijaya, H.O.L., Santoso, B. (2024). Classification of palm oil fruit ripeness based on AlexNet deep convolutional neural network. *SINERGI Учредители: Universitas Mercu Buana*, 29(1): 207. <https://doi.org/10.22441/sinergi.2025.1.019>

[25] Wang, X., Yan, F., Li, B., Yu, B., Zhou, X., Tang, X., Jia, T., Lv, C. (2025). A multimodal data fusion and embedding attention mechanism-based method for eggplant disease detection. *Plants*, 14(5): 786. <https://doi.org/10.3390/plants14050786>

[26] Seth, K. (2020). Fruits and vegetables image recognition dataset. *Kaggle*. <https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition>, accessed on Aug. 9, 2024.

[27] Fruits-360 dataset: A dataset of images containing fruits, vegetables, nuts and seeds. <https://www.kaggle.com/datasets/moltean/fruits>, accessed on Aug. 11, 2025.