ISSN 2959-6157

Classification of Pistachio Species based on Deep Learning Models

Han Li

School of Computer Science and Technology, Jiangxi Normal University, Nanchang, China
*Corresponding author: lihanv@jxnu.edu.cn

Abstract:

Pistachios are highly nutritious and have a large market presence. Since different species vary in quality and can be sold at different prices, it is crucial to classify prominent species accurately. Additionally, given the large number of species, improving the speed of species identification is also important. However, traditional methods for identifying pistachio species can be challenging in both accuracy and efficiency. This study examines the utilization of Deep Learning technologies to automatically classify the two pistachio species from Turkey: Kirmizi and Siirt. The performance of Convolutional Neural Networks (CNNs) is compared to that of Vision Transformers. Evaluated CNN models include ResNet and EfficientNet, which were trained and tested using both actual and augmented images. The experimental results show that ResNet achieves the best accuracy and fastest inference time, the peak accuracy is 99.30%. Data augmentation improves performance but increases inference time. The research findings emphasize the potential of Deep Learning for improving pistachio species classification.

Keywords: Pistachio; Convolutional Neural Networks; Vision Transformer; Data Augmentation.

1. Introduction

Pistachios are highly nutritious, contain unsaturated fatty acids essential for humans [1], and are widely used in snack foods. Their dark green kernels make them especially popular in the ice cream and pastry industries. As a high-cost agricultural product, pistachio prices vary based on quality, with different varieties commanding different prices. Turkey is a major global producer of pistachios, offering many varieties. Kirmizi and Siirt pistachios are particularly popular for their abundant fruits and reduced tendency to periodicity. Thus, classifying these two varieties from others is important.

In recent years, deep learning models have made significant advances and have been commonly utilized in computer vision (CV) domains, such as image classification, autonomous driving, and medical imaging. More recent innovations, such as Transformers, have been used to enhance their capability and efficiency. This paper introduces three deep-learning models for pistachio species classification and discusses their performance in Section 3.

2. Literature Review

Many studies have explored various methods for classifying pistachio species, including both machine learning and deep learning approaches. While traditional CNNs are commonly used, newer models are still relatively underutilized. Below is a summary of some of these studies.

The study [2] aimed to distinguish between Kirmizi and Siirt species using deep learning algorithms, including AlexNet, VGG16, and VGG19. Classification success was assessed using five key metrics. The results indicated 94.41% accuracy with AlexNet, 98.84% accuracy with VGG16, and 98.14% accuracy with VGG19.

A computer vision system (CVS) was applied, incorporating image processing techniques and artificial intelligence methods [3]. Images were initially segmented, followed by the extraction of morphological and shape features. The method used to reduce dimensionality was Principal Component Analysis (PCA), after which k-NN and weighted k-NN were utilized for classification. With 10-fold cross-validation, the PCA-based weighted k-NN method achieved an accuracy of 94.18%.

CNN algorithms were utilized to classify defective and perfect pistachios [4]. A dataset comprising 958 images was used, and the results showed that the accuracies achieved were 95.8% with GoogleNet, 97.2% with ResNet, and 95.83% with VGG16.

Another study [5] aimed to classify wild pistachios into four different ripeness levels. Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Artificial Neural Networks (ANN) were used as classifiers. Following image preprocessing, feature extraction and selection, the optimal features were used to train the classifiers. Classification rates achieved were 93.75% with LDA, 97.5% with QDA, and 100% with ANN, showing

the efficacy of the imaging algorithm when paired with both linear and non-linear classification methods for assessing the ripeness stages of wild pistachios.

The study [6] developed a system to identify whether the mouths of various pistachios are opened or closed. The models based on CNNs, including ResNet50, ResNet152, and VGG16, were employed for feature extraction and classification of the images. The classification accuracies for the models were recorded at 85.28%, 85.19%, 83.32%.

3. Method

This section introduces three models used for classifying pistachio species, each with different structures, as categorized in Fig.1.

Fig.1 Architecture classification of the methods

3.1 ResNet

ResNet was proposed to solve the problem of network degradation [7]. The integration of shortcut connections and residual representations made these networks simpler to optimize relative to earlier models, and their increased depth contributed to improved accuracy. The architecture of the residual block is illustrated in Fig.2.

X represents the input image of a block, $F = W_2\sigma(W_1X)$ represents the operating function in the block, which includes two weight layers: W_1, W_2, and σ indicates the RELU activation function. The biases are omitted for simplicity. If the dimensions of X and F are equal, the block's output can be obtained by:

$$Y = F\left(X, \{W_i\}\right) + X \qquad (1)$$

Else,

$$Y = F\left(X, \{W_i\}\right) + W_s X \qquad (2)$$

In which, W_i represents the i-th weight layer, and W_s is a linear projection performed by the shortcut connections to match the same dimensions.

In image classification tasks [8, 9, 10], ResNet has achieved better performance. This architecture is also used to train the pistachio classifier in this study.

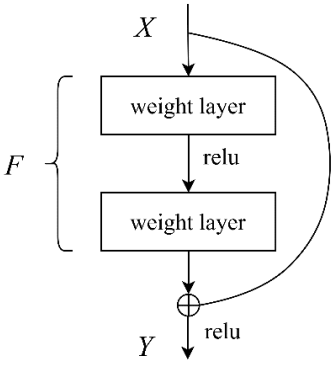

Fig.2 The architecture of the residual block

3.2 EfficientNet

Compared to merely increasing the depth in a neural network, fully balancing the depth, width, and resolution of a network can lead to better performance [11]. Therefore, EfficientNet was proposed. This approach employs a compound scaling method, which uses a compound coefficient φ to systematically scale the network's depth, width, and resolution coherently:

$$\begin{aligned} depth &: d = \alpha^{\varphi} \\ width &: w = \beta^{\varphi} \\ resolution &: r = \gamma^{\varphi} \\ s.t. &\ \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ &\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned} \qquad (3)$$

The constants α, β, γ can be obtained through a small grid search. The purpose of optimization is to ensure that the model achieves maximum accuracy within specific resource limitations, as shown:

$$\begin{aligned} &max_{d,\omega,r} Accuracy(N(d,w,r)) \\ &s.t. Memory(N) \leq target\_memory \\ &FLOPS(N) \leq target\_flops \end{aligned} \qquad (4)$$

Where N represents ConvNet. *Memory* refers to the amount of RAM required by the model during computation. *FLOPS* refers to Floating Point Operations Per Second, measuring a model's computational complexity.

In the first step, φ is fixed at 1, and then the optimal values of α, β, γ for EfficientNet-B0 are found using a small grid search based on Equation 3 and 4. In the second step, α, β, γ are fixed, and the baseline network is scaled up with increasing φ based on Equation 3 to obtain Efficient-B1 through B7. The EfficientNet-B0 is utilized in this paper to classify pistachio species.

3.3 Vision Transformer

Originally introduced for machine translation, transformers have since become the leading method for achieving state-of-the-art (SOTA) performance across various NLP tasks. Transformer was first applied to CV tasks in [12], where it was named Vision Transformer (ViT).

The standard Transformer accepts 1D sequences as input. In CV tasks, the 2D image $X \in \mathbb{R}^{H \times W \times C}$ are reshaped into

a sequence of flattened 2D patches $X_p \in \mathbb{R}^{N\times(P^2\cdot C)}$, where (H,W) represents the resolution of the image, C represents the channels' number, (P,P) represents the resolution of image patches, and $N = HW/P^2$ is the total patches number. The sequence of embedded patches ($z_0^0 = X_{class}$) is added with a learnable embedding, with its state at the output of the Transformer encoder (z_L^0) serving as the image representation Y. The explanation of ViT is shown as:

$$z_0 = \left[X_{class}; X_p^1 E; X_p^2 E; \ldots; X_p^N E\right] + E_{pos}, \quad E \in \mathbb{R}^{(P^2\cdot C)\times D}, E_{pos} \in \mathbb{R}^{(N+1)D} \tag{5}$$

$$z'_d = MSA\left(LN\left(z_{d-1}\right)\right) + z_{d-1}, d = 1\ldots D \tag{6}$$

$$z_d = MLP\left(LN\left(z'_d\right)\right) + z'_d, d = 1\ldots D \tag{7}$$

$$Y = LN\left(z_L^0\right) \tag{8}$$

D refers to the fixed size of the latent vector maintained across all layers. MSA denotes multi-headed self-attention, while MLP is made up of two layers incorporating GELU non-linearity. LN stands for Layernorm, which is applied before each block, and residual connections are used following each block. ViT has demonstrated leading performance on numerous image classification benchmarks, it is also utilized in this study for classifying pistachio species.

3.4 Confusion Matrix

The confusion matrix is a standard tool for assessing image classification performance, and it is employed in this study to develop performance metrics for the trained pistachio classifier. Table 1 presents the specifics of the two-class confusion matrix.

Table 1. Two-class confusion matrix

		True Class	
		Positive(P)	Negative(N)
Predicted Class	True(T)	TP	FN
	False(F)	FP	TN

1. TP: True Positive. The real class of the sample is 1, and it is correctly predicted as 1.
2. FN: False Negative. The real class of the sample is 1, but it is incorrectly predicted as 0.
3. FP: False Positive. The real class of the sample is 0, but it is incorrectly predicted as 1.
4. TN: True Negative. The real class of the sample is 0 and it is correctly predicted as 0.

3.5 Performance Metrics

Five Metrics are implemented in this study, all the metrics are derived from the two-class confusion matrix in Table 2.

Table 2. Calculation formulas of Performance Metrics

Performance Metrics	Calculation formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F-1 Score	$\frac{2TP}{2TP+FP+FN}$
Recall	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$
Specificity	$\frac{TN}{TN+FP}$

4. Experimental results and discussion

4.1 Pistachio Image Dataset

A dataset of pistachio images was collected from the Kaggle repository, featuring two varieties: Kirmizi and Siirt. The dataset comprises 2,148 images, including 1,232 of Kirmizi and 916 of Siirt. In this study, 60% of the data was designated for training the model, 20% was used for validation, and the remaining 20% was reserved for assessing the final model's performance.

4.2 Experiment I: Comparison of three models

In the first experiment, model training was performed with the actual dataset over ResNet, EfficientNet, and ViT. Fig.3 shows the confusion matrices, providing a comprehensive overview of their performance. By utilizing these matrices, other performance metrics are calculated, as shown in Table 3.

According to Table 3, ResNet achieved the highest classification success, with an accuracy of 99.30%. EfficientNet achieved an accuracy of 94.66%, but Vision Transformer only reached 83.06% accuracy. The inference time of all the models was also compared, results in Table 4 show that ResNet achieved the lowest inference time per image, at 0.0169 seconds. These results clearly show that CNNs are preferable for training pistachio species classifiers, ResNet is both accurate and efficient for classifying the two varieties of pistachios.

CNNs use convolutional layers to extract features and excel at capturing local spatial information, while ViT utilizes self-attention mechanisms to process image data and capture global context and relationships [13]. It is concluded that for pistachio species classification, focusing on local information is more important than capturing the global context.

(a) ResNet (b) EfficientNet (c) Vision Transformer

Fig.3 Confusion matrices for all the models

Table 3. Performance metrics for all the models

Performance Metrics	ResNet	EfficientNet	Vision Transformer
Accuracy	0.9930	0.9466	0.8306
F-1 Score	0.9939	0.9532	0.8630
Recall	0.9919	0.9474	0.9312
Precision	0.9959	0.9590	0.8042
Specificity	0.9946	0.9457	0.6957

Table 4. Inference Time of all models using GPU (seconds per image)

	ResNet	EfficientNet	Vision Transformer
Inference Time	0.0169	0.0202	0.1486

4.3 Experiment II: Augmented Data Training

The second experiment compares how models perform with actual and augmented datasets. Table 5 contains a detailed overview of the performance metrics, where both ResNet and EfficientNet are compared to enhance the credibility of the results.

The results indicate that data augmentation improves classification performance, but it also leads to increased in-

ference time. Thus, the decision to use data augmentation depends on a balance between accuracy and acceptable inference time, which is different in specific tasks.

Table 5. Performance comparison over the actual dataset and augmented dataset

Performance Metrics	Actual dataset		Augmented dataset	
	ResNet	EfficientNet	ResNet	EfficientNet
Accuracy	0.9930	0.9466	0.9930	0.9629
F-1 Score	0.9939	0.9532	0.9940	0.9683
Recall	0.9919	0.9474	1.0000	0.9879
Precision	0.9959	0.9590	0.9880	0.9494
Specificity	0.9946	0.9457	0.9837	0.9293
Inference Time	0.0169	0.0202	0.0316	0.0252

5. Conclusion

This study demonstrates that CNNs outperform Vision Transformers in classifying pistachio species. Among the CNN architectures evaluated, ResNet emerged as the most effective, achieving a high classification accuracy of 99.30% and the lowest inference time of 0.0169 seconds per image. Although ResNet performs best with the current dataset, results may vary with different datasets. This variability emphasizes the need for ongoing model assessment and adaptation to maintain accuracy. Additionally, while data augmentation enhances classification performance, it also increases inference time, a balance between accuracy and efficiency will be needed to meet market requirements.

References

[1] Mateos R, Salvador M D, Fregapane G, et al. Why Should Pistachio Be a Regular Food in Our Diet?. Nutrients, 2022, 14(15): 3207.

[2] Singh D, Taspinar Y S, Kursun R, et al. Classification and analysis of pistachio species with pre-trained deep learning models. Electronics, 2022, 11(7): 981.

[3] Ozkan IA, Koklu M, Saraçoğlu R. Classification of pistachio species using improved k-NN classifier. Health, 2021, 23: e2021044.

[4] Dini A, Zadeh H G, Rahimifard A, Fayazi A, Eftekhari M, Abbaszadeh M. Designing a hardware system to separate defective pistachios from healthy ones using deep neural networks. Iran. J. Biosyst. Eng, 2020, 51: 149-159.

[5] Kheiralipour K, Nadimi M, Paliwal J. Development of an intelligent imaging system for ripeness determination of wild pistachios. Sensors, 2022, 22(19): 7134.

[6] Rahimzadeh M, Attar A. Detecting and counting pistachios based on deep learning. Iran Journal of Computer Science, 2022, 5(1): 69-81.

[7] Duta I C, Liu L, Zhu F, et al. Improved residual networks for image and video recognition. International Conference on Pattern Recognition (ICPR) IEEE, 2021, 25: 9415-9422.

[8] Reddy A S B, Juliet D S. Transfer learning with ResNet-50 for malaria cell-image classification. International conference on communication and signal processing (ICCSP) IEEE, 2019, 0945-0949.

[9] Liang J. Image classification based on RESNET. Journal of Physics: Conference Series, 2020, 1634.

[10] Sarwinda D, Paradisa R H, Bustamam A, et al. Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. Procedia Computer Science, 2021, 179: 423-431.

[11] Koonce B. EfficientNet. Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization, 2021, 109-123.

[12] Khan S, Naseer M, Hayat M, et al. Transformers in vision: A survey. ACM computing surveys (CSUR), 2022, 54(10s): 1-41.

[13] Raghu M, Unterthiner T, Kornblith S, et al. Do vision transformers see like convolutional neural networks?. Advances in neural information processing systems, 2021, 34: 12116-12128.