

Article

Apple Varieties Classification Using Deep Features and Machine Learning

Alper Taner ^{1,*}, Maitem Teweldemedhin Mengstu ^{2,†}, Kemal Çağatay Selvi ¹, Hüseyin Duran ¹, İbrahim Gür ³ and Nicoleta Ungureanu ^{4,*}

¹ Department of Agricultural Machinery and Technologies Engineering, Faculty of Agriculture, Ondokuz Mayıs University, 55139 Samsun, Turkey; kcselvi@omu.edu.tr (K.Ç.S.); huseyin.duran@omu.edu.tr (H.D.)

² Department of Agricultural Engineering, Hamelmalo Agricultural College, Keren P.O. Box 397, Eritrea; maitem23@gmail.com

³ Fruit Research Institute, 32500 Isparta, Turkey; igur03@gmail.com

⁴ Department of Biotechnical Systems, Faculty of Biotechnical Systems Engineering, National University of Science and Technology Politehnica Bucharest, 060042 Bucharest, Romania

* Correspondence: alper.taner@omu.edu.tr (A.T.); nicoleta.ungureanu@upb.ro (N.U.)

† These authors contributed equally to this work.

Abstract: Having the advantages of speed, suitability and high accuracy, computer vision has been effectively utilized as a non-destructive approach to automatically recognize and classify fruits and vegetables, to meet the increased demand for food quality-sensing devices. Primarily, this study focused on classifying apple varieties using machine learning techniques. Firstly, to discern how different convolutional neural network (CNN) architectures handle different apple varieties, transfer learning approaches, using popular seven CNN architectures (VGG16, VGG19, InceptionV3, MobileNet, Xception, ResNet150V2 and DenseNet201), were adopted, taking advantage of the pre-trained models, and it was found that DenseNet201 had the highest (97.48%) classification accuracy. Secondly, using the DenseNet201, deep features were extracted and traditional Machine Learning (ML) models: support vector machine (SVM), multi-layer perceptron (MLP), random forest classifier (RFC) and K-nearest neighbor (KNN) were trained. It was observed that the classification accuracies were significantly improved and the best classification performance of 98.28% was obtained using SVM algorithms. Finally, the effect of dimensionality reduction in classification performance, deep features, principal component analysis (PCA) and ML models was investigated. MLP achieved an accuracy of 99.77%, outperforming SVM (99.08%), RFC (99.54%) and KNN (91.63%). Based on the performance measurement values obtained, our study achieved success in classifying apple varieties. Further investigation is needed to broaden the scope and usability of this technique, for an increased number of varieties, by increasing the size of the training data and the number of apple varieties.

Keywords: transfer learning; deep features; principal component analysis; machine learning; apple



Citation: Taner, A.; Mengstu, M.T.; Selvi, K.Ç.; Duran, H.; Gür, İ.; Ungureanu, N. Apple Varieties Classification Using Deep Features and Machine Learning. *Agriculture* **2024**, *14*, 252. <https://doi.org/10.3390/agriculture14020252>

Academic Editors: Francesco Marinello, Gniewko Niedbała, Sebastian Kujawa, Magdalena Piekutowska and Tomasz Wojciechowski

Received: 22 December 2023

Revised: 28 January 2024

Accepted: 1 February 2024

Published: 3 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Apples are among the most consumed and produced temperate fruit crops in the world; in 2022, over 95 million tons of apples were produced [1]. Due to their taste and nutritional value, there is an ever-increasing demand for apples that requires continuous production and supply. In this context, the implementation of innovative solutions in the systems of agricultural production and marketing is much required [2].

Identifying and sorting apples has several limitations because of its subjective nature, and it is a relatively complex problem due to the impressive number of apple varieties [3]. Human perception can be influenced when assessing the texture, color pattern, smell and other characteristics of apples. Likewise, manual inspections are highly dependent on the experience, training and duration of work of the personnel, as well as the environmental and psychological conditions, which can cause inconsistencies and variations in the results

or processing time [4,5]. However, automated recognition and classification systems can play a vital role in reducing labor costs and enhancing the economic efficiency of fruit, right from their harvest to the market [6]. Having the advantages of speed, suitability and high accuracy, computer vision has been effectively utilized as a non-destructive approach to automatically recognize and classify fruits and vegetables to meet the increased demand for food quality-sensing devices.

In the field of computer vision, image classification is a widely studied topic. In traditional machine learning, extracting features before training models on these features is required. Thus, the quality of the extracted features has a significant impact on a given classifier [7]. Deep learning has gained much popularity in image recognition and classification tasks of fruits and vegetables, as computing power and algorithms to process big data are emerging [8]. Because the convolutional neural network (CNN) model serves to automatically extract and classify features [9], it has been effectively utilized as a non-destructive approach to automatically classify fruits and vegetables to meet the increased demand for food quality-sensing devices [10], avoiding the need to manually or separately extract image features or representations [9].

In studies [11–15], convolutional neural networks (CNNs), which are deep learning-based, have shown excellent outcomes in a wide range of food and agricultural tasks, namely grading and sorting, varieties classification and disease detection. In grading bio-colored apples, a study [16] employed CNN, using multispectral images, to ensure the quality grading of apples. Likewise, CNN was also used to differentiate mature apples from immature ones in apple trees [17]. In fruit sorting, CNN models were applied to detect defective apples [18], while a similar work identified bruised apples in the investigation of automated sorting, by fusing deep features [19]. A great performance of the suppression mask R-CNN was reported in the classification of Gala and Blondee apple varieties [20].

Integrating CNNs and a convolution autoencoder, the authors of study [21] classified 26 different fruits, out of which nine classes were apples. In a similar classification problem, a CNN model was trained utilizing 30 types of leaf images from various growth periods [22]. Recently, the successful application of transfer learning to identify and classify 13 apple varieties using publicly available image datasets was reported [23].

The application of transfer learning, using models pre-trained on images from the internet, in diverse tasks such as the classification of crops and fruits, has been increasingly applicable and effective. As a result, applying a pre-trained network to learn new patterns with new data is beneficial. Furthermore, it is helpful when there are relatively small data to train a model. Thus, employing a pre-trained model is a typical solution [24]. This study firstly aims to test the performance of transfer learning in apple varieties and to investigate the impact of principal components (PC) of deep features coupled with traditional machine learning models. To the best of our knowledge, the integration of PC, deep features and machine learning has not been tested in apple varieties classification. Accordingly, this study aims to assess how these diverse components can jointly interact to enhance the accuracy and efficacy of apple variety classification.

The objectives of this study are: (I) to apply the transfer learning approach to develop an apple varieties classifier using pre-trained popular CNN architectures; (ii) to train and evaluate machine learning (ML) models using deep features obtained using best-performing models from the transfer learning approach; and (iii) to assess the effect of principal component analysis (PCA) on the performance of ML models trained using deep features.

2. Materials and Methods

2.1. Image Data

Ten apple varieties obtained from the Ministry of Agriculture and Fruit Research Institute of the Republic of Turkey were used in this study (Figure 1). The images were captured with a 20-megapixel resolution camera from a uniform distance and light set-up.

A total of 5808 images were captured from three views. Of these images, 70% were used in training, 15% in testing and 15% in validation, respectively.


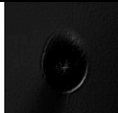
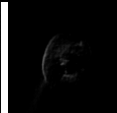
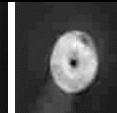
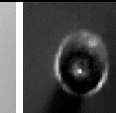
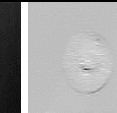
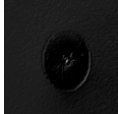
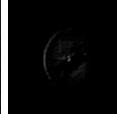
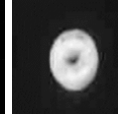
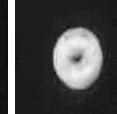
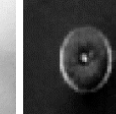
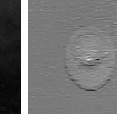

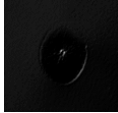
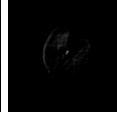
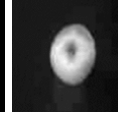
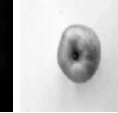
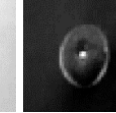
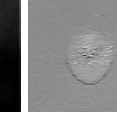

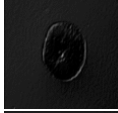

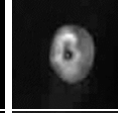
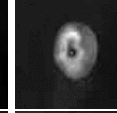
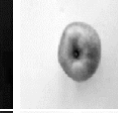
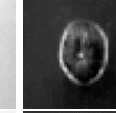
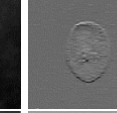
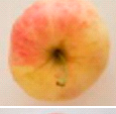
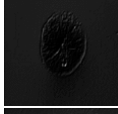
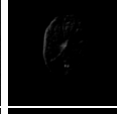
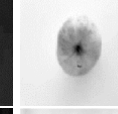
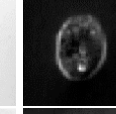


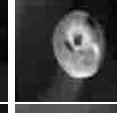
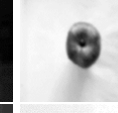
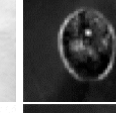
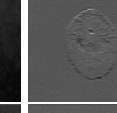
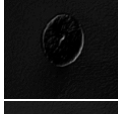
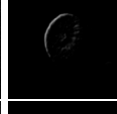
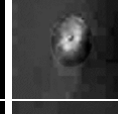
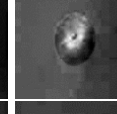
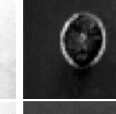

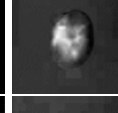
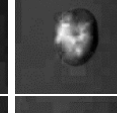
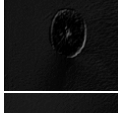


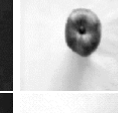
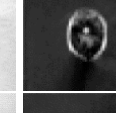
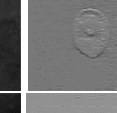


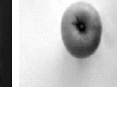
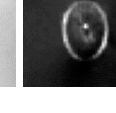

CNN Models								
Varieties		VGG16	VGG19	InceptionV3	Xception	Mobile Net	ResNet 150V2	DenseNet 201
FUJI								
GOLDEN REINDERS								
GRANNY SMITH								
KASEL 37								
MONDIAL GALA								
RED BREABURN								
RED CHIEF								
SCARLET SPUR								
STAR KRIMSON								
STARKSPUR C.DELICIOUS								

Figure 1. Activations of the first layer of all models.

2.2. Transfer Learning

Generally, the training of CNNs is performed using large amount of data. Their deeper and interconnected layers are the reason for their good performance. Nonetheless, CNNs are often trained on a small dataset, which can simply result in the overfitting of the networks, leading to poor and unconvincing performances [23]. Likewise, training deep CNNs contributes to significant weaknesses, namely high processing costs and slow-running processes [25]. Thus, to solve the aforementioned difficulties, transfer learning is widely adopted. Transfer learning is a deep learning approach that utilizes the information acquired from a well-established model to reinstate with a new problem [26]. Because it

transfers the relevant knowledge from the learned model to the new task, transfer learning reduces the training time [27].

Thus, in this study, a transfer learning approach was adopted; seven popular architectures (VGG16, VGG19, InceptionV3, MobileNet, Xception, ResNet150V2 and DenseNet201) were trained using the Keras library in Google Colab, keeping the weights of the convolutional base and modifying only the final output layer, according to the number of classes in our study. This method utilizes the knowledge gained by these architectures from previously learned tasks, enhancing the model's performance on the target task. The description of each CNN architecture is as follows: VGG16 comprises thirteen convolutional layers and five Maxpooling layers; and VGG19 has sixteen convolutional layers and five Maxpooling layers. The VGG networks are typically designed successively [28]. From the Inception family, the InceptionV3 and Xception networks were adopted, which are characterized by multiple parallel convolutional operations, known as inception modules [29], with 48 convolutional layers and 71 layers, respectively. MobileNet is known for its depthwise separable convolution, having 53 layers [30]. From networks characterized by residual connections, ResNet150V2, with 150 layers [31], and DenseNet201, with dense connectivity between layers [32], with 201 layers, were used. In all the above models, the convolutional base was adopted excluding the top layer. According to our task, one hidden layer of 256 neurons with ReLu activation and a final layer of 10 represents the number of the apple classes or the final model prediction.

To discern how different architectures handle different apple varieties, the activations of the first layer of each model used have been provided in Figure 1. The activations represent the availability of specific features in the input data that serve as building blocks for successive layers to detect more complex features. Every model was trained for the same training, validation and test proportions. During the training, the hyperparameters were fixed to 100 batch size, 100 patience and 200 epochs. Early stopping and model checkpoint callbacks were also adopted to stop the training and save the weights of the best model, while monitoring the validation loss for a given training argument. This strategy facilitates the training, as it stops the training before all specified epochs are undertaken when the validation loss stops to decrease for a considerable amount of time.

2.3. Deep Features

Deep features are a set of high-level representations of input data, such as images or text, that are extracted from a deep neural network (DNN) such as MobileNet and DenseNet that is pre-trained on a large dataset. DNN is typically a multi-layered neural network that learns to identify patterns and features in the input data by successively transforming it through multiple layers of non-linear functions. The final output of the DNN, which is often a classification decision, is produced by a classifier that is trained on the deep features extracted from the input data. Deep features are more robust and informative than traditional hand-crafted features, as they capture more complex and nuanced relationships between the input data and the output labels [33].

Deep features are useful in a variety of machine learning tasks, such as image classification, object detection and natural language processing. In this study, after the transfer learning strategy was implemented, we took the best-performing model, which was DenseNet201, extracted 1920 deep features and used them to train traditional machine learning models (Figure 2).

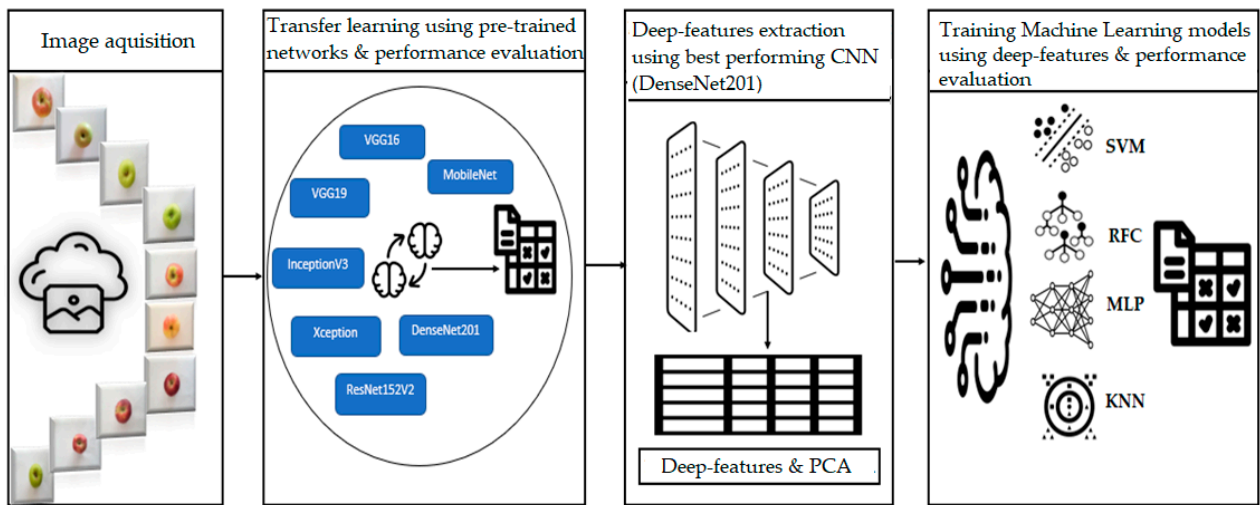


Figure 2. Diagram representation of the proposed approach.

2.4. Machine Learning Models

To train ML models, the 1920 deep features extracted using the best-performing CNN architecture, DenseNet201, were used. Four ML models, namely support vector machine (SVM), random forest classifier (RFC), multi-layer perceptron (MLP) and K-nearest neighbor (KNN) were trained with 10-fold stratified cross-validation (Skfold), as shown in Figure 3.

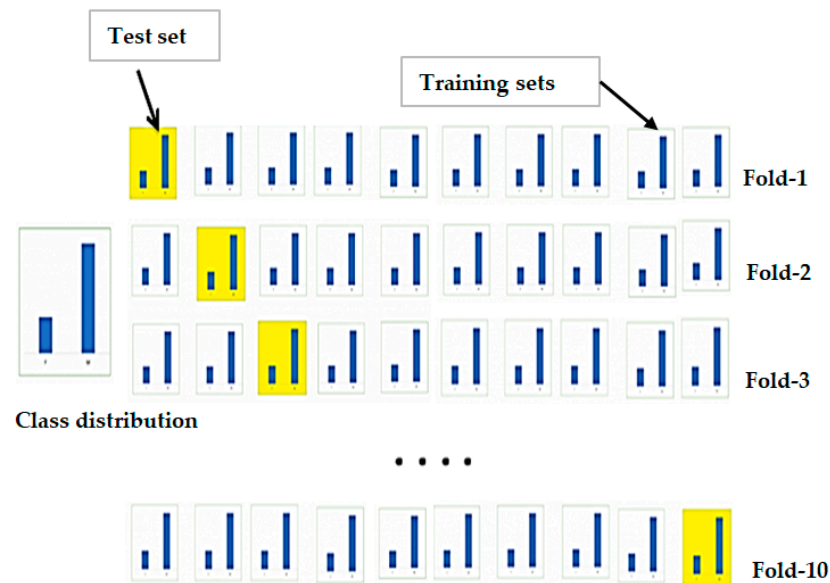


Figure 3. Tenfold stratified cross validation.

As depicted in Figure 3, the utilization of Skfold becomes vital when dealing with imbalanced data distributions among classes. In conditions where the frequency of different classes varies significantly, only K-fold cross-validation may result in unequal representations of classes across folds. Skfold, however, addresses this case by ensuring that each fold maintains a proportional representation of the various classes available in the dataset. This approach is mainly helpful in machine learning task where maintaining the balance of class distribution is decisive for model training. By stratifying the folds based on class labels, Skfold improves the robustness of the model evaluation, avoiding biased performance metrics that could appear from uneven class representation in traditional cross-validation. This approach promotes a more reliable assessment of the model’s generalization capabilities across diverse class distributions, ultimately contributing to a more accurate and

unbiased evaluation of the model's performance. To fine-tune the models, a grid search was implemented and final models were trained using the best parameters.

2.5. Principal Component Analysis

To investigate the effect of dimensionality reduction in classification performance, all ML models were also trained with deep features after implementing principal component analysis (PCA). PCA is a method that takes multi-dimensional data and gives it components by using the dependencies between the variables representing it in a more manageable and lower-dimensional form, without losing too much information. The fundamental idea of PCA is to minimize the dimensionality of a data set that is composed of a large number of interrelated variables, while retaining the variation present in the data to a feasible extent [34].

PCA is a linear transformation of data that minimizes the redundancy measured through covariance and maximizes the information that is measured through variance. PCA diminishes the number of given variables by reducing the last principal components that do not significantly contribute to the observed variability [35].

In the analysis, new elements, known as principal components and ranked by their Eigen values, are created. Principal components (PC) are new variables with two properties: firstly, each PC is a linear combination of the original variables; and secondly, PCs are uncorrelated to each other and the redundant information is removed [36].

As a multivariate unsupervised statistical procedure, PCA is widely used as a data exploratory tool in conditions that require feature selection such as data compression, image analysis, visualization, pattern recognition, regression and time series prediction. In our case, using the Python PCA library, we performed a PCA for the 1920 deep features. After examining the eigenvalues which told us the amount of variance explained by a single component, 262 features with eigenvalues greater than one were selected to train SVM, RFC, MLP and KNN models.

2.6. Performance Evaluation

To evaluate the performance of all models' accuracy, precision, recall, specificity, F1-score, Cohen's kappa, Matthews correlation coefficient, area under the receiver operating characteristic curve (AUC-ROC) the trade-off between true positive rate and false positive rate, area under the precision-recall curve (AUC-PR) and the trade-off between precision and recall for different classification thresholds, performance metrics were used in this study. The equations of all metrics are provided below, where TP: true positive, TN: true negative, FP: false positive, FN: false negative, Po: relative observed agreement among raters and Pe: relative observed agreement among raters.

1. Accuracy (Acc) is the ratio of the number of correctly classified samples to the total number of samples:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

2. Precision (Pre) is the proportion of true positives out of total predicted positives, also known as positive predicted value:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

3. Recall (Rec) is the proportion of positive samples classified as true. Recall is referred to as a true positive rate:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

4. Specificity (Spec) is the proportion of negative samples classified as true. Known as a true negative rate:

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

5. F1-Score (FS) is the harmonic mean of recall and precision:

$$\text{FS} = 2 \cdot \frac{\text{Pre} \cdot \text{Recall}}{\text{Pre} + \text{Recall}} \quad (5)$$

6. Cohen's kappa (K) is a measure of inter-rater agreement that considers the agreement that would be expected by chance:

$$\text{K} = \frac{\text{Po} - \text{Pe}}{1 - \text{Pe}} \quad (6)$$

7. Matthews correlation coefficient (MCC) is the correlation between the predicted and actual classifications:

$$\text{MCC} = \frac{[(\text{TP} \cdot \text{TN}) - (\text{FP} \cdot \text{FN})]}{\sqrt{[(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})]}} \quad (7)$$

3. Results

In this study, three methods were proposed: (i) the adoption of transfer learning using the popular CNN model; (ii) extracting deep features and training traditional ML models making use of the best performing CNN model, which was the DenseNet201; (iii) applying PCA to the deep features and training ML models using the PCs with eigenvalues greater than one. The performance metrics of each proposed method are presented in Tables 1–3, respectively.

Table 1. Performance of popular CNN models using Transfer Learning.

CNN Model	Performance Metrics								
	Acc	Pre	Rec	Spec	FS	AUC -ROC	AUC -PR	K	MCC
VGG16	91.87	92.449	91.87	99.09	91.85	99.72	97.97	90.95	91.02
VGG19	94.16	94.55	94.16	99.35	94.18	99.76	98.38	93.49	93.54
InceptionV3	92.44	92.74	92.51	99.16	92.45	99.77	98.13	91.59	91.63
MobileNet	96.45	96.61	96.45	99.61	96.44	99.96	99.67	96.05	96.07
DenseNet201	97.48	97.54	97.48	99.72	97.48	99.98	99.86	97.19	97.20
Xception	93.47	93.97	93.47	99.27	93.39	99.79	98.69	92.73	92.80
ResNet152V2	94.84	95.26	94.84	99.42	94.76	99.89	99.17	94.26	94.32

As depicted in Table 1, in the results of performance metrics of the seven pre-trained CNN models, DenseNet201 outperformed all other models with an accuracy of 97.48% when classifying 10 apple varieties.

Similarly, the performance of the machine learning models, namely SVM, RFC, MLP and KNN, which were trained and tested using 1920 features obtained from DenseNet201 CNN (presented in Table 2), showed that SVM obtained an accuracy of 98.28%, outperforming all other models. However, as PCA was applied and features were reduced to 262, MLP outperformed all models with a classification accuracy of 99.77% (presented in Table 3).

Table 2. Performance metrics of ML models.

Model	Performance Metrics								
	Acc	Pre	Rec	Spec	FS	AUC-ROC	AUC-PR	K	MCC
SVM	98.28	98.32	98.10	97.67	98.18	99.88	98.71	98.08	98.09
RFC	91.86	91.65	91.69	90.14	91.62	99.56	96.60	90.94	90.95
MLP	98.05	97.99	97.95	96.47	97.96	99.96	99.70	97.83	97.83
KNN	89.33	91.59	88.93	98.46	89.04	98.59	95.04	88.13	88.17

Table 3. Performance metrics of ML models trained with PCA.

Model	Performance Metrics								
	Acc	Pre	Rec	Spec	FS	AUC-ROC	AUC-PR	K	MCC
SVM	99.08	99.06	99.05	97.70	99.05	99.93	99.43	98.97	98.98
RFC	99.54	99.57	99.55	100	99.99	99.99	99.94	99.49	99.49
MLP	99.77	99.78	99.75	100	99.99	99.99	99.99	99.74	99.75
KNN	91.63	92.35	91.26	98.70	99.70	99.70	97.09	90.68	90.73

The confusion matrices containing the classification results of the test data of each model trained using 262 PCs are depicted in the figures below.

Applying the SVM model, a few misclassifications were observed in Fuji, Golden Reinders, Kasel37 and Mondial Gala apple varieties, as shown in Figure 4. The model scored an excellent classification performance of 99.08%.

SVM	Predicted										
	FUJI	GOLDEN REINDERS	GRANNY SMITH	KASEL 37	MONDIAL GALA	RED BREABURN	RED CHIEF	SCARLET SPUR	STARKRIMSON	STARKSPUR	G.DELICIOUS
FUJI	72	0	0	0	0	0	0	0	0	0	
GOLDEN REINDERS	1	92	0	0	0	0	0	0	0	0	
GRANNY SMITH	0	0	88	0	0	0	0	0	0	0	
KASEL 37	0	0	0	79	0	0	0	0	0	0	
MONDIAL GALA	0	0	0	0	72	0	0	0	0	0	
RED BREABURN	0	0	0	1	1	85	0	0	0	0	
RED CHIEF	0	0	0	0	0	0	91	1	0	0	
SCARLET SPUR	0	0	0	0	0	0	0	96	0	0	
STARKRIMSON	1	0	0	1	0	0	0	0	91	0	
STARKSPUR G.DELICIOUS	1	1	0	0	0	0	0	0	0	98	

Figure 4. SVM confusion matrix.

Figure 5 presents the confusion matrix according to the MLP model. Out of the four ML models, MLP has the highest performance of 99.77%; out of 782 instances of given test data, only four misclassifications were observed.

MLP	Predicted									
Actual	FUJI	GOLDEN REINDERS	GRANNY SMITH	KASEL 37	MONDIAL GALA	RED BREABURN	RED CHIEF	SCARLET SPUR	STARKRIMSON	STARKSPUR G.DELICIOUS
FUJI	72	0	0	0	0	0	0	0	0	0
GOLDEN REINDERS	0	92	0	0	0	1	0	0	0	0
GRANNY SMITH	0	0	88	0	0	0	0	0	0	0
KASEL 37	0	0	0	79	0	0	0	0	0	0
MONDIAL GALA	0	0	0	0	72	0	0	0	0	0
RED BREABURN	0	0	0	0	0	87	0	0	0	0
RED CHIEF	0	0	0	0	0	0	91	1	0	0
SCARLET SPUR	0	0	0	0	0	0	1	95	0	0
STARKRIMSON	0	1	0	0	0	0	0	0	92	0
STARKSPUR G.DELICIOUS	0	0	0	0	0	0	0	0	0	100

Figure 5. MLP confusion matrix.

With the highest confusion between Kasel37 and Red Braeburn, the RFC model performed with an overall classification accuracy of 99.54%, where eight misclassifications were observed (Figure 6).

RFC	Predicted									
Actual	FUJI	GOLDEN REINDERS	GRANNY SMITH	KASEL 37	MONDIAL GALA	RED BREABURN	RED CHIEF	SCARLET SPUR	STARKRIMSON	STARKSPUR G.DELICIOUS
FUJI	72	0	0	0	0	0	0	0	0	0
GOLDEN REINDERS	0	92	0	0	0	0	0	0	0	1
GRANNY SMITH	0	0	88	0	0	0	0	0	0	0
KASEL 37	0	0	0	76	0	3	0	0	0	0
MONDIAL GALA	0	0	0	0	72	0	0	0	0	0
RED BREABURN	0	0	0	0	0	86	0	0	0	1
RED CHIEF	0	0	0	1	1	0	90	0	0	0
SCARLET SPUR	0	0	0	0	0	0	0	95	1	0
STARKRIMSON	0	0	0	0	0	0	0	0	93	0
STARKSPUR G.DELICIOUS	0	1	0	0	0	0	0	0	0	99

Figure 6. RFC confusion matrix.

Among the four models, KNN appeared to have the lowest performance, with 90 misclassifications, which is very significant compared to the other three models. A total of 15 misclassifications were also observed between Fuji and Red Braeburn apple varieties, followed by seven misclassifications between Red Braeburn and Kasel37 varieties; or, out of the total misclassifications, 22 or 24% were wrongly classified as Red Braeburn, which shows that a close observation in the features of Red Braeburn variety is needed (Figure 7).

KNN	Predicted										
	FUJI	GOLDEN REINDERS	GRANNY SMITH	KASEL 37	MONDIAL GALA	RED BREABURN	RED CHIEF	SCARLET SPUR	STARKRIMSON	STARKSPUR	G.DELICIOUS
FUJI	47	0	0	4	1	15	1	0	1	3	
GOLDEN REINDERS	0	77	2	1	0	0	0	0	0	13	
GRANNY SMITH	0	1	86	0	0	0	0	0	0	1	
KASEL 37	0	0	1	73	0	2	2	0	0	1	
MONDIAL GALA	1	0	0	4	63	4	0	0	0	0	
RED BREABURN	0	1	0	7	0	77	0	0	2	0	
RED CHIEF	0	0	0	4	0	0	82	6	0	0	
SCARLET SPUR	0	0	0	2	1	0	5	88	0	0	
STARKRIMSON	0	0	0	3	0	1	4	2	83	0	
STARKSPUR G.DELICIOUS	0	2	0	0	0	0	0	0	0	98	

Figure 7. KNN confusion matrix.

4. Discussion

A deep learning-based convolutional neural network for apple classification was presented in a previous study [37]. The model used a network with four layers to classify unforeseen apple images, and the CNN model was trained and tested using images of 13 apple varieties. At the test phase, the model achieved an accuracy of 90%. In another study, by proposing an integrated CNN and certainty factor, a model with a dataset containing images of six apple varieties was trained, and the image classification performance scored an excellent result of 99.78% [38]. When classifying a group of 26 different fruits, which included nine apple classes, the authors of study [21] also obtained an accuracy of 95.67%, employing CNN and the autoencoder. In study [39], a shallow CNN was employed to classify six apple varieties; the authors collected and labeled apple images to train a model along with data augmentation. As the training and model parameter optimization were performed using the Caffee framework, at the test stage, the model accuracy performance was 92%. A CNN-based deep learning model was developed in study [22] to classify thirty apple varieties under complex natural environments, which contributed to the currently available means of apple variety classification; the model’s accuracy was 93.14% using the test set.

Applying the ability of the CNN model to extract features, a method to identify 14 apple varieties was proposed by the authors of study [40]. Compared to transfer learning techniques, such as ResNet50, VGG16, MobileNet and EfficientNetB0, their approach attained an improved test accuracy of 99.59%.

Recently, in the successful application of transfer learning to identifying and classifying 13 apple varieties using publicly available image datasets with accuracies of 96–100%, the use of different models was reported [23]. The summary of recently related works, including our proposed model, has been provided in Table 4.

Table 4. Performance comparison of the proposed methods with related works.

Task	Model and Accuracy	References
Classification of 13 apple varieties	CNN, 90%	[37]
Classification of 6 apple varieties	CNN and certainty factor, 99.78%	[38]
Classification of 26 fruits	CNN and autoencoder, 95.67%	[21]
Classification of 6 apple varieties	Shallow CNN, 92%	[39]
Classification of 30 apple varieties	CNN, 93.14%	[22]
Classification of 14 apple varieties	Lightweight CNN, 99.59%	[40]
Classification of 14 apple varieties	CNN, 96.1–100%	[23]
Classification of 10 apple varieties	Transfer learning, 97.48%	Proposed Method 1
Classification of 10 apple varieties	Deep features and ML, 98.28%	Proposed Method 2
Classification of 10 apple varieties	Deep features, PCA and ML, 99.77%	Proposed Method 3

In this study, compared to relevant and related works, it can be concluded that the results of our proposed models are in agreement with the classification accuracy indicating the relevance of all the techniques, and especially the third method, which is the integration of deep features, PCA and ML, which performed as excellently as state-of-the-art models and even outperformed some of them, likely because of the use of PCA.

Additionally, in the research of apple varieties' classifications, to the best of our knowledge, the integration of deep features and PCA has not yet been adopted; therefore, the approach of coupling deep features, PCA and machine learning models in the area of apple varieties classification is an innovative approach that this study has investigated and demonstrated. Furthermore, the adoption of Skfold validation is a new addition to the literature that this investigation has undertaken.

According to the performance metrics (Table 1), DenseNet201 has the highest classification accuracy (97.48%) among popular CNN models applied in the transfer learning approach, indicating its potential to be applied in apple varieties classification. The investigation into the integration of deep features and traditional ML models shows the benefits and advantages of coupling deep features and ML features to have improved classification accuracy; SVM scored 98.28%, but ultimately, the incorporation of PCA increased the model's performance up to 99.77%, as achieved by MLP model. Additionally, observing other performance metrics, MLP has the highest values—a precision of 99.78%, a recall of 99.75% and an F1 score is 99.76%. Besides, its discrimination power was observed in the results of AUC-ROC and AUC-PR, which were 99.99% and 99.99%, respectively. Moreover, Cohen's Kappa and MCC metrics, which examine the agreement between predicted and actual classes, are 99.75% and 99.74%, outperforming all the ML models, as shown in Table 3.

Thus, based on our proposed technique, integrating deep features, PCA and the ML models MLP and SVM can classify ten apple varieties with excellent performance.

5. Conclusions

In solving agricultural problems such as fruit varieties' classification and grading, the application of machine learning has a significant role. As indicated in our work, and also in other recent and related works, transfer learning is commonly used in classifying different fruit types, and specifically apple varieties. Our work trained and tested seven popular CNN architectures from different families, which all of them performed well, with above

90% classification accuracy. To further and discover a new dimension in apple varieties' classification, our investigation focused on the application of deep features coupled with PCA and traditional ML algorithms, for which the Skfold stratified cross-validation was adopted during the training.

All four models, namely SVM, RFC, MLP and KNN, were tested with separate test data and achieved classification accuracies of 99.08%, 99.54%, 99.77% and 91.63%, respectively. The study confirmed the increase in performance by coupling deep features and PCA for the given image dataset. This study was undertaken on ten apple varieties; however, to prove the efficiency of the proposed technique, increasing the size of the training data and increasing the number of apple varieties should be further studied and examined. Besides, to develop robust and versatile models, we suggest the use of random images, containing scattered apple images taken in real-time, such as on conveyors and other production environments. In future investigations, images captured in various acquisition setups and light conditions that represent actual and complicated environments should be considered. One of the challenges in apple varieties is the occurrence of variability within classes, which contributes to confusion and misclassification. Although it is difficult to provide conclusive reasons based on experiments, the misclassifications indicated in the results might be due to variability within classes. Hence, future research endeavors should explore and identify algorithms capable of handling such complexities. Prioritizing the discovery and implementation of algorithms that are robust to variations within classes will enhance the model's ability to accurately discriminate between intricate and closely related instances, contributing to an improved classification performance in challenging scenarios. On top of the varieties, the difference or similarities of agricultural products such as fruits is highly impacted by the stage of ripening. Two different varieties might be highly similar and highly different at various stages of their ripening, which significantly contribute to the challenges in the development of sorting, grading or classification models. Therefore, during data collection, such things need to be addressed and considered. Moreover, apart from classification models, the use of object detection algorithms, including prominent ones like YOLO (You Only Look Once), needs to be researched, to further strengthen and enhance the practical and real-time application of models. Object detection goes beyond classifying entire images, allowing for the precise identification and localization of multiple objects within an image.

Author Contributions: Conceptualization, A.T. and M.T.M.; methodology, A.T.; software, M.T.M.; validation, K.Ç.S., A.T. and H.D.; formal analysis, A.T. and M.T.M.; investigation, A.T.; resources, İ.G.; data curation, A.T.; writing—original draft preparation, A.T., N.U. and H.D.; writing—review and editing, N.U. and İ.G.; visualization, N.U. and M.T.M.; supervision, A.T., K.Ç.S. and H.D.; project administration, A.T.; funding acquisition, N.U. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by National University of Science and Technology Politehnica Bucharest, Romania, within the PubArt Program.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

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

References

1. Food and Agriculture Organization of the United Nations. *FAOSTAT Statistical Database*; FAO: Rome, Italy, 2022. Available online: <https://www.fao.org/faostat/en/#data/QCL/visualize> (accessed on 15 January 2024).
2. Rachmawati, E.; Supriana, I.; Khodra, M.L. Toward a new approach in fruit recognition using hybrid RGBD features and fruit hierarchy property. In *Proceedings of the 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Yogyakarta, Indonesia, 19–21 September 2017; pp. 1–6. [CrossRef]
3. Hameed, K.; Chai, D.; Rassau, A. A comprehensive review of fruit and vegetable classification techniques. *Image Vis Comput.* **2018**, *80*, 24–44. [CrossRef]

4. Barbelian, M.A.; Bălan, C.V. Fault tree event classification by neural network analysis. *UPB Sci. Bull. Series D Mech. Eng.* **2017**, *79*, 55–66.
5. Bhargava, A.; Bansal, A.; Goyal, V. Machine learning-based detection and sorting of multiple vegetables and fruits. *Food Anal. Methods* **2022**, *15*, 228–242. [[CrossRef](#)]
6. Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. Computer vision technology in agricultural automation—A review. *Inf Process Agric.* **2020**, *7*, 1–19. [[CrossRef](#)]
7. Taner, A.; Mengstu, M.T.; Selvi, K.Ç.; Duran, H.; Kabaş, Ö.; Gür, İ.; Karaköse, T.; Gheorghită, N.-E. Multiclass apple varieties classification using machine learning with histogram of oriented gradient and color moments. *Appl. Sci.* **2023**, *13*, 7682. [[CrossRef](#)]
8. Srivalli, D.S.; Geetha, A. Fruits, vegetable and plants category recognition systems using convolutional neural networks: A review. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2019**, *5*, 452–462. [[CrossRef](#)]
9. Iosif, A.; Maican, E.; Biriş, S.; Popa, L. Automated quality assessment of apples using convolutional neural networks. *INMATEH -Agric. Eng.* **2023**, *71*, 483–498. [[CrossRef](#)]
10. Abasi, S.; Minaei, S.; Jamshidi, B.; Fathi, D. Dedicated non-destructive devices for food quality measurement: A review. *Trends Food Sci. Technol.* **2018**, *78*, 197–205. [[CrossRef](#)]
11. Gikunda, P.K.; Jouandeau, N. State-of-the-art convolutional neural networks for smart farms: A review. In Proceedings of the Intelligent Computing Conference, London, UK, 16–17 July 2019; pp. 763–775. [[CrossRef](#)]
12. Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart Seed Classification System based on MobileNetV2 Architecture. In Proceedings of the 2nd International Conference on Computing and Information Technology (ICCIIT), Tabuk, Saudi Arabia, 25–27 January 2022; pp. 217–222. [[CrossRef](#)]
13. Kaur, P.; Harnal, S.; Tiwari, R.; Upadhyay, S.; Bhatia, S.; Mashat, A.; Alabdali, A.M. Recognition of leaf disease using hybrid convolutional neural network by applying feature reduction. *Sensors* **2022**, *22*, 575. [[CrossRef](#)]
14. Nagaraju, M.; Chawla, P.; Upadhyay, S.; Tiwari, R. Convolution network model based leaf disease detection using augmentation techniques. *Expert Syst.* **2022**, *39*, e12885. [[CrossRef](#)]
15. Nagaraju, M.; Chawla, P.; Tiwari, R. An effective image augmentation approach for maize crop disease recognition and classification. In *Computational Intelligence and Smart Communication; ICCISC 2022*. Communications in Computer and Information Science; Mehra, R., Meesad, P., Peddoju, S.K., Rai, D.S., Eds.; Springer: Cham, Switzerland, 2022; Volume 1672. [[CrossRef](#)]
16. Unay, D. Deep learning based automatic grading of bi-colored apples using multispectral images. *Multimed. Tools Appl.* **2022**, *81*, 38237–38252. [[CrossRef](#)]
17. Lu, S.; Chen, W.; Zhang, X.; Karkee, M. Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. *Comput. Electron. Agric.* **2022**, *193*, 106696. [[CrossRef](#)]
18. Fan, S.; Li, J.; Zhang, Y.; Tian, X.; Wang, Q.; He, X.; Zhang, C.; Huang, W. On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* **2020**, *286*, 110102. [[CrossRef](#)]
19. Hu, Z.; Tang, J.; Zhang, P.; Jiang, J. Deep learning for the identification of bruised apples by fusing 3D deep features for apple grading systems. *Mech. Syst. Signal Process.* **2020**, *145*, 106922. [[CrossRef](#)]
20. Chu, P.; Li, Z.; Lammers, K.; Lu, R.; Liu, X. Deep learning-based apple detection using a suppression mask R-CNN. *Pattern Recognit. Lett.* **2021**, *147*, 206–211. [[CrossRef](#)]
21. Xue, G.; Liu, S.; Ma, Y. A hybrid deep learning-based fruit classification using attention model and convolution autoencoder. *Complex Intell. Syst.* **2023**, *9*, 2209–2219. [[CrossRef](#)]
22. Chen, J.; Han, J.; Liu, C.; Wang, Y.; Shen, H.; Li, L. A deep-learning method for the classification of apple varieties via leaf images from different growth periods in natural environment. *Symmetry* **2022**, *14*, 1671. [[CrossRef](#)]
23. Yu, F.; Lu, T.; Xue, C. Deep learning-based intelligent apple variety classification system and model interpretability analysis. *Foods* **2023**, *12*, 885. [[CrossRef](#)]
24. Naranjo-Torres, J.; Mora, M.; Hernández-García, R.; Barrientos, R.J.; Fredes, C.; Valenzuela, A. A review of convolutional neural network applied to fruit image processing. *Appl. Sci.* **2020**, *10*, 3443. [[CrossRef](#)]
25. Canziani, A.; Paszke, A.; Culurciello, E. An analysis of deep neural network models for practical applications. *arXiv* **2016**, arXiv:1605.07678. [[CrossRef](#)]
26. Behera, S.K.; Rath, A.K.; Sethy, P.K. Maturity status classification of papaya fruits based on machine learning and transfer learning approach. *Inf. Process Agric.* **2020**, *8*, 244–250. [[CrossRef](#)]
27. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
28. Xiao, J.; Wang, J.; Cao, S.; Li, B. Application of a novel and improved VGG-19 network in the detection of workers wearing masks. *J. Phys. Conf. Ser.* **2020**, *1518*, 012041. [[CrossRef](#)]
29. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [[CrossRef](#)]
30. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [[CrossRef](#)]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

32. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
33. Lee, S.H.; Chan, C.; Mayo, S.J.; Remagnino, P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit.* **2017**, *71*, 1–13. [[CrossRef](#)]
34. Jolliffe, I.T. Principal Component Analysis. In *Springer Series in Statistics*; Springer: New York, NY, USA, 2002; ISBN 0-387-95442-2.
35. Ma, J.; Yuan, Y. Dimension reduction of image deep feature using PCA. *J. Vis. Commun. Image Represent.* **2019**, *63*, 102578. [[CrossRef](#)]
36. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the Science and Information (SAI) Conference, London, UK, 27–29 August 2014; pp. 372–378. [[CrossRef](#)]
37. Al-Shawwa, M.; Abu-Naser, S.S. Knowledge based system for apple problems using CLIPS. *Int. J. Acad. Eng. Res.* **2019**, *3*, 1–11.
38. Katarzyna, R.; Paweł, M. A vision-based method utilizing deep convolutional neural networks for fruit variety classification in uncertainty conditions of retail sales. *Appl. Sci.* **2019**, *9*, 3971. [[CrossRef](#)]
39. Li, J.; Xie, S.; Chen, Z.; Liu, H.; Kang, J.; Fan, Z.; Li, W. A shallow convolutional neural network for apple classification. *IEEE Access* **2020**, *8*, 111683–111692. [[CrossRef](#)]
40. Shruthi, U.; Narmadha, K.S.; Meghana, E.; Meghana, D.N.; Lakana, K.P.; Bhuvan, M.P. Apple varieties classification using light weight CNN Model. In Proceedings of the 4th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 21–23 December 2022; pp. 68–72. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.