Article

Not peer-reviewed version

A General Machine Learning Model for Assessing Fruit Quality Using Deep Image Features

Ioannis D. Apostolopoulos *, Mpesi Tzani, Sokratis Aznaouridis

Posted Date: 24 July 2023

doi: 10.20944/preprints202307.1552.v1

Keywords: Fruit Quality; Machine Learning; Deep Learning

Article

A General Machine Learning Model for Assessing Fruit Quality Using Deep Image Features

Ioannis D. Apostolopoulos [1,*], Mpesi Tzani [2] and Sokratis I. Aznaouridis [3]

[1] Department of Medical Physics, School of Medicine, University of Patras, 26504 Rio, Greece
[2] Department of Electrical and Computer Technology Engineering, University of Patras, 26504 Rio, Greece
[3] Department of Computer Engineering and Informatics, University of Patras, 26504 Rio, Greece
* Correspondence: ece7216@upnet.gr

Abstract: Fruit quality is a critical factor in the produce industry, affecting producers, distributors, consumers, and the economy. High-quality fruits are more appealing, nutritious, and safe, boosting consumer satisfaction and revenue for producers. Artificial Intelligence can aid in assessing the quality of the fruit using images. This paper presents a general machine-learning model for assessing fruit quality using deep image features. The model leverages the learning capabilities of the recent successful networks for image classification called Vision Transformers (ViT). The ViT model is built and trained with a combination of various fruit datasets and learned to distinguish between good and rotten fruit images. The general model demonstrated impressive results in accurately identifying the quality of various fruits such as Apples (with a 99.50% accuracy), Cucumbers (99%), Grapes (100%), Kakis (99.50%), Oranges (99.50%), Papayas (98%), Peaches (98%), Tomatoes (99.50%), and Watermelons (98%). However, it showed slightly lower performance in identifying Guavas (97%), Lemons (97%), Limes (97.50%), mangoes (97.50%), Pears (97%), and Pomegranates (97%).

Keywords: fruit quality; machine learning; deep learning

1. Introduction

Fruit quality refers to a fruit's overall characteristics that determine its desirability, nutritional content, and safety for consumption [1]. It is determined by the fruit's appearance, flavour, texture, nutritional value, and safety [2]. High fruit quality is crucial for the industry, consumers, and the economy for several reasons.

High-quality fruits benefit growers and sellers economically, promote healthy eating habits, reduce healthcare costs, positively impact the environment, ensure food safety, and promote international trade [3]. Promoting high fruit quality requires using sustainable farming practices, implementing food safety regulations, and promoting healthy eating habits [3]. Hence, it is crucial for all stakeholders involved in the fruit industry to work towards promoting high fruit quality.

Artificial Intelligence (AI) can aid in assessing the quality of the fruit using images [4]. AI-based technologies such as computer vision and machine learning algorithms can analyse the visual characteristics of the fruit and provide an objective quality assessment [5]. The AI algorithms can be trained using a large dataset of images of different fruits with varying quality. They can learn to identify the specific features that indicate the quality of the fruit [6].

The present study is the first to introduce the concept of a general Machine Learning (ML) model for visually assessing the fruit quality of various fruits. We considered the development of a Vision Transformer (ViT) network [7], a type of neural network architecture designed for image classification tasks that use the transformer architecture, introduced initially for natural language processing. In ViT, an image is first divided into fixed-size patches. These patches are then flattened and linearly projected into a lower-dimensional space, creating a sequence of embeddings. These embeddings are then fed into a multi-head self-attention mechanism, which allows the network to learn to attend to essential patches and relationships between patches.

The self-attention mechanism [8] is followed by a feedforward neural network, which processes the attended embeddings and outputs class probabilities. ViT also includes additional techniques,

 doi:10.20944/preprints202307.1552.v1

such as layer normalisation, residual connections, and token embedding, which help improve the network's performance. ViT allows for the effective use of self-attention mechanisms in image classification tasks, providing a promising alternative to traditional Convolutional Neural Networks (CNNs) [9].

A collection of fruit-quality datasets served to train the general model and inspect its performance against fruit-dedicated trained models.

The contributions of the study can be summarised as follows:

- We present a general ML model for determining the quality of various fruits based on their visual appearance
- The general model performs better or equal to dedicated per-fruit models
- Comparisons with the state-of-the-art reveal the superiority of ViTs in fruit quality assessment

2. Related Work

Recent studies have reported remarkable success in visually estimating the fruits' quality.

Rodríguez et al. [10] focused on identifying plum varieties during early maturity stages, a difficult task even for experts. The authors proposed a two-step approach where images are first processed to isolate the plum. Then, a deep convolutional neural network is used to determine its variety. The results demonstrate high accuracy ranging from 91 to 97%.

In [11], the authors proposed a CNN to help with sorting by detecting defects in mangosteen. Indonesia has identified mangosteen as a fruit with significant export potential, but not all are defect-free. Quality assurance for export is done manually by sorting experts, which can lead to inconsistent and inaccurate results due to human error. The suggested method achieved a classification accuracy of 97% in defect recognition.

During the growth process of apple fruit crops, there are instances where biological damage occurs on the surface or inside of the fruit. These lesions are typically caused by external factors such as the incorrect application of fertilisers, pest infestations, or changes in meteorological conditions such as temperature, sunlight, and humidity. Wenxue et al. [12] employed a CNN for real-time recognition of apple skin lesions captured by infrared video sensors, capable of intelligent, unattended alerting for disease pest. Experimental results show that the proposed method achieves a high accuracy and recall rate of up to 97.5% and 98.5%, respectively.

In [13], the authors proposed an automated method to distinguish between naturally and artificially ripened bananas using spectral and RGB data. They used a neural network on RGB data and achieved an accuracy of up to 90%. They used spectral data classifiers such as Random Forest, multilayer perceptron, and feedforward neural networks. They achieved accuracies of up to 98.74% and 89.49%, respectively. These findings could help ensure the safety of banana consumption by identifying artificially ripened bananas, which can harm human health.

In [14], hyperspectral reflectance imaging (400~1000 nm) was used to evaluate and classify three common types of peach diseases by analysing spectral and imaging information. Principal component analysis was used to reduce the high dimensionality of the hyperspectral images, and 54 imaging features were extracted from each sample. Three levels of peach decay (slight, moderate, and severe) were considered for classification using Deep Belief Network (DBN) and partial least squares discriminant analysis models. The results showed that the DBN model had better classification accuracy than the partial least squares discriminant analysis model. The DBN model, which utilised integrated information (494 features), had the highest accuracy of 82.5%, 92.5%, and 100% for slightly-decayed, moderately-decayed, and severely-decayed samples, respectively.

In [15] proposed developing a deep learning-based model called Fruit-CNN for recognising fruits and assessing their quality. The dataset used in the study includes 12 categories of six different fruits based on their quality. It comprises 12,000 images in real-world situations with varying backgrounds. The dataset is divided into training, validation, and testing subsets for training and evaluating the proposed CNN architecture. The Fruit-CNN model outperforms other state-of-the-art models, achieving an accuracy of 99.6% on a test set of previously unseen images.

In [16], the authors utilised a CNN to create an efficient fruit classification model. The model was trained using the fruits 360 dataset, which consists of 131 varieties of fruits and vegetables. The study focused on three specific fruits, divided into three categories based on quality: good, raw, and damaged. The model was developed using Keras and trained for 50 epochs, achieving an accuracy rate of 95%.

In [6], the authors used two banana fruit datasets to train and assess their presented model. The original dataset contains 2100 images categorised into ripe, unripe, and over-ripe, with 700 images in each category. The study employed a handcrafted CNN for the classification. The CNN model achieved an accuracy of 98.25% and 81.96% regarding the two datasets.

In [17], the authors developed a model to identify rotting fruits from input fruit images. The study used three types of fruits: apples, bananas, and oranges. The features of the fruit images were collected using the MobileNetV2 [18] architecture. The model's performance was evaluated on a Kaggle dataset, and it achieved a validation accuracy of 99.61%.

In [19], the authors proposed two approaches for classifying the maturity status of papaya: Machine Learning (ML) and transfer learning. The experiment used 300 papaya fruit images, with 100 images for each maturity level. The ML approach utilised Local binary pattern, histogram of directed gradients, grey level co-occurrence matrix, and classification approaches including k-nearest neighbours, support vector machine, and naive Bayes. In contrast, transfer learning utilised seven pre-trained models, including VGG-19 [20]. Both methods achieved 100% accuracy, with the ML method achieving this in 0.0995 seconds of training time and the transfer learning method achieving 100% accuracy.

Most related works have focused on building fruit-specific models. Subsequently, they utilised datasets containing fruits from a single variety. There is a need for general fruit quality prediction models, which are transferrable from industry to industry and are trained using large-scale datasets. Moreover, recent advances in deep learning models can be benchmarked for fruit quality assessment to investigate their performance.

3. Materials and Methods

3.1. Fruit Quality

Fruit quality is the measure of the characteristics that determine the value of fruit, including appearance, taste, texture, aroma, nutritional content, and safety [2]. The importance of fruit quality cannot be overstated, as it has significant implications for the industry, people, and the economy [3].

For the industry, fruit quality is critical for market competitiveness and profitability. The produce industry is highly competitive, and consumers are more discerning than ever, demanding high-quality fruits that meet their flavour, appearance, and nutrition expectations. Producers must, therefore, ensure that their fruits are of the highest quality to meet consumer demands and compete in the market effectively.

Furthermore, the reputation of producers and distributors depends on the quality of their products [3]. Consumers who are satisfied with the quality of fruit are more likely to become repeat customers and recommend the products to others, which can help to build a strong brand image and increase sales [3].

In addition, fruit quality is critical for food safety [1]. Poor-quality fruits are more prone to contamination by pathogens and spoilage microorganisms, leading to foodborne illness outbreaks and damaging the industry's reputation. Producers and distributors must ensure that the fruits they produce and sell are high-quality and safe for consumption.

For people, fruit quality is crucial because it determines the taste, nutritional value, and safety of their consumed fruits [1]. High-quality fruits are more nutritious, flavorful, and appealing, making them more likely to be consumed and incorporated into a healthy diet. Furthermore, high-quality fruits are less likely to contain harmful contaminants or spoilage microorganisms, reducing the risk of foodborne illness and promoting public health.

 doi:10.20944/preprints202307.1552.v1

For the economy, fruit quality is crucial because it impacts local and international trade. High-quality fruits are more likely to meet export standards and regulations, enabling producers to enter new markets and increase their revenue. Furthermore, high-quality fruits are more likely to fetch higher prices, boosting producers' income and contributing to economic growth.

In addition, fruit quality impacts the entire supply chain, from producers to distributors to retailers. High-quality fruits are less likely to spoil during transportation and storage, reducing waste and increasing profits for all parties involved. Furthermore, high-quality fruits are more likely to be sold at premium prices, increasing the value of the entire supply chain.

Several factors determine fruit quality, including variety, growing conditions, harvesting practices, transportation, and storage [1]. For example, the timing of the harvest can significantly impact fruit quality. Harvesting fruits too early can result in poor taste, texture, and aroma; harvesting them too late can lead to overripening, loss of nutrients, and spoilage.

Growing conditions like soil quality, irrigation, and pest management can also impact fruit quality. Fruits grown in nutrient-rich soil, with proper irrigation and pest management practices, are more likely to be of higher quality than those grown in poor soil conditions or with inadequate pest control measures.

Transportation and storage conditions are also crucial for maintaining fruit quality. Fruits must be transported and stored at optimal temperatures and humidity levels to prevent spoilage, maintain freshness, and preserve nutritional value.

3.2. Deep Learning Framework

We propose a ViT model for the classification task. The current section describes the fundamental concepts of the ViT model and the parameters of the proposed model.

3.2.1. Convolutional Neural Networks (CNNs)

CNNs are a class of neural networks designed explicitly for image-processing tasks [21]. CNNs use convolutional and pooling layers to extract features from an input image. Convolutional layers work by convolving a set of learnable filters (kernels) over the input image to produce feature maps [9]. The filters are designed to detect specific patterns in the image, such as edges or corners.

Pooling layers are used to downsample the feature maps produced by convolutional layers, reducing their size while retaining the most critical information. The most common type of pooling layer is max pooling, which takes the maximum value from each subregion of the feature map.

CNNs have succeeded highly in image classification tasks, achieving state-of-the-art performance on benchmark datasets such as ImageNet. However, they are limited in their ability to capture global relationships between different parts of an image

3.2.2. Transformers

Transformers are a type of neural network architecture initially developed for natural languages processing tasks, such as machine translation and text summarisation. Transformers use a self-attention mechanism [22] to capture relationships between different parts of an input sequence [23].

The self-attention mechanism works by computing a weighted sum of the input sequence, where the weights are learned based on the importance of each element to the other elements in the sequence. This allows the model to focus on relevant parts of the input sequence while ignoring irrelevant parts.

Transformers have been highly successful in natural language processing tasks, achieving state-of-the-art performance on benchmark datasets such as GLUE and SuperGLUE.

3.2.3. ViT model

ViTs are a type of deep learning model that combines the power of CNNs with the attention mechanism of transformers to process images. This hybrid architecture is highly effective for image

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 July 2023 doi:10.20944/preprints202307.1552.v1

classification tasks, as it allows the model to focus on relevant parts of an image while capturing spatial relationships between them.

ViTs use two main components: CNNs and transformer networks. The CNNs are used for feature extraction from the images, while transformer networks are used for attention mechanisms. CNNs are particularly good at capturing local image features like edges and corners. In contrast, transformer networks can capture the global structure of images by attending to relevant regions. By combining the two, visual transformer CNNs can capture local and global features, improving performance.

ViTs divide the input image into a grid of smaller patches, similar to how image segmentation works [7]. Each patch is flattened and passed through convolutional layers to extract features. The transformer network then processes these features, which attends to the most relevant features and aggregates them to generate a representation of the image. This representation is then passed through a series of fully connected layers to classify the image.

The proposed ViT model Figure 1 consists of multiple layers of self-attention and feedforward networks. The self-attention mechanism allows the network to attend to different input parts and weight them based on relevance. The feedforward network generates a new representation of the input, which is then used in the next self-attention layer.

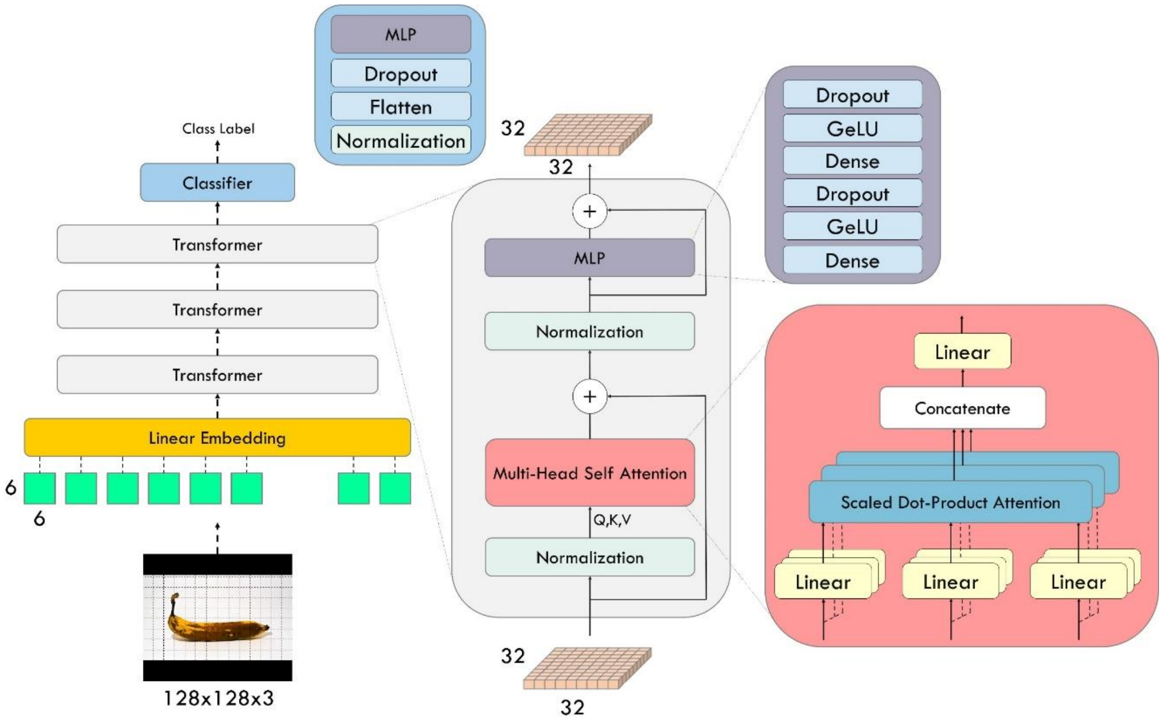

Figure 1. The proposed Vision Transformer Network.

The model takes input images of size (200,200,3) and returns a prediction of one of the two classes. The model's architecture consists of a series of transformer blocks, each with a multi-head attention layer and a multilayer perceptron (MLP) layer. The input images are first divided into patches and then fed into the transformer blocks. The model is trained using the Sparse Categorical Cross entropy loss function and the AdamW optimiser.

The model first processes the input images by dividing them into smaller patches. Each patch is then encoded using a Patch Encoder layer, which applies a dense layer and an embedding layer. The encoded patches are then passed through a series of transformer blocks. Each block applies a layer of multi-head attention followed by an MLP. The multi-head attention layer allows the model to attend to different image parts. In contrast, the MLP layer applies non-linear transformations to the encoded patches.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 July 2023 doi:10.20944/preprints202307.1552.v1

After the final transformer block, the encoded patches are flattened and fed into an MLP that produces the final classification. The MLP applies two dense layers with 500 and 250 units to the encoded patches, respectively. The output of the MLP is then passed through a dense layer with two units and a Softmax activation function to produce the final prediction.

The model is trained using the Sparse Categorical Cross-entropy loss function, which compares the predicted class probabilities to the actual class labels. The AdamW optimiser optimises the model, which applies weight decay to the model parameters. The model is evaluated using the Sparse Categorical Accuracy metric, which measures the proportion of correctly classified examples.

3.3. Datasets

3.3.1. Sources

We used various sources for collecting fruit images classified between quality-related categories. We used the extracted image collection to develop the study's large-scale dataset. The image sources are the following:

- FruitNet: Indian Fruits Dataset with Quality: https://www.kaggle.com/datasets/shashwatwork/fruitnet-indian-fruits-dataset-with-quality
- FruitQ dataset: https://www.kaggle.com/datasets/sholzz/fruitq-dataset
- Lemon Quality Dataset: https://www.kaggle.com/datasets/yusufemir/lemon-quality-dataset
- Mango Varieties Classification and Grading: https://www.kaggle.com/datasets/saurabhshahane/mango-varieties-classification

3.3.2. Characteristics

The datasets mentioned above were processed to create the study's dataset. The analysis identified 16 fruits.

We have followed the steps described below to create the dataset:

Step 1. Download all files from each source
Step 2. Create the initial vocabulary of examined fruits
Step 3. For each dataset, validate the availability of each fruit in the vocabulary
Step 4. For each dataset, exclude corrupted and low-resolution images
Step 5. Create a large-scale dataset that contains all available fruits
Step 6. Exclude fruits that are not labelled
Step 7. Define the two classes: Good Quality (GQ) and Bad Quality (BQ)
Step 8. Exclude fruits that include less than 50 images per class

Table 1 presents the image distribution between the classes of the final dataset, the total number of images per fruit, the initial image format and image size.

Table 1. Per-fruit characteristics of the study's dataset.

Datasets	Number of images representing good quality fruit	Number of images representing bad quality fruit	Total	Format	Image size (height, width)
Apple	1149	1141	2290	PNG	(192,256)
Banana	1292	1520	2812	PNG	(720,1280)
Cucumber	250	461	711	PNG	(720,1280)
Grape	227	482	709	PNG	(720,1280)
Guava	1152	1129	2281	JPEG	(256,256)
Kaki	545	566	1111	PNG	(720,1280)
Lemon	1125	951	2076	PNG	(300,300)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 July 2023 doi:10.20944/preprints202307.1552.v1

Lime	1094	1085	2179	JPEG	(192,256)
Mango	200	200	400	JPEG	(424,752)
Orange	1216	1159	2375	PNG	(256,256)
Papaya	130	663	793	PNG	(720,1280)
Peach	425	720	1145	PNG	(720,1280)
Pear	504	593	1097	JPEG	(720,1280)
Pomegranate	5940	1187	7127	JPEG	(256,256)
Tomato	600	1255	1855	PNG	(720,1280)
Watermelon	51	203	254	PNG	(720,1280)
Total (UD dataset)	15900	13315	29215	-	-

Apart from the 16 separate datasets, which have been organised to represent one fruit each, we created an ultimate dataset of all fruits for training the general model. This dataset will henceforth be addressed as Union Dataset (UD).

We also collected 200 images per fruit that serve the purpose of the external evaluation dataset. The characteristics of this dataset are presented in **Table 2**.

Table 2. Per-fruit characteristics of the study's external evaluation dataset.

External Dataset	Number of images representing good quality fruit	Number of images representing bad quality fruit	Total	Format	Image size (height, width)
Apple	100	100	200	JPEG	(192,256)
Banana	100	100	200	JPEG	(720,1280)
Cucumber	100	100	200	JPEG	(256,256)
Grape	100	100	200	PNG	(256,256)
Guava	100	100	200	JPEG	(256,256)
Kaki	100	100	200	PNG	(720,1280)
Lemon	100	100	200	PNG	(300,300)
Lime	100	100	200	JPEG	(192,256)
Mango	100	100	200	JPEG	(424,752)
Orange	100	100	200	JPEG	(256,256)
Papaya	100	100	200	PNG	(256,256)
Peach	100	100	200	JPEG	(256,256)
Pear	100	100	200	JPEG	(720,1280)
Pomegranate	100	100	200	JPEG	(256,256)
Tomato	100	100	200	PNG	(256,256)
Watermelon	100	100	200	JPEG	(720,1280)

Figure 2 illustrates the data collection and pre-processing steps for creating the datasets of the present study.

Dataset pre-processing includes sorting the images by fruit, excluding low-resolution and corrupted images, grouping the images into classes, resizing the images to fit in a black-background 200x200 pixel canvas, and normalisation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 July 2023 doi:10.20944/preprints202307.1552.v1

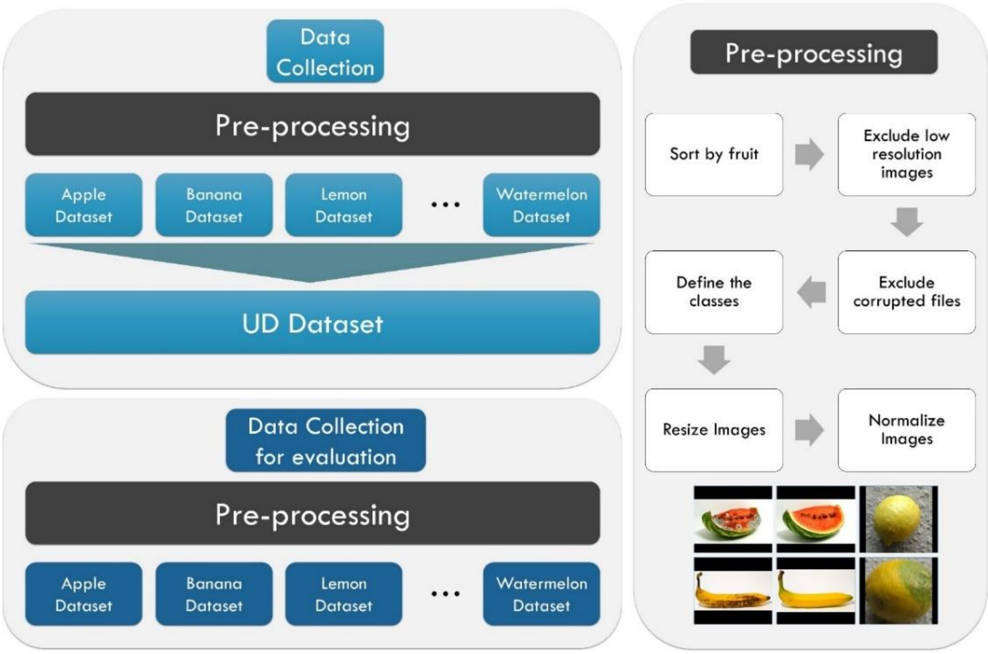

Figure 2. Data collection and processing procedure. The top-left box describes the process of creating the UD dataset. The low-left box presents the creating of the external evaluation datasets. Both datasets share the same pre-processing steps, visualized in the right box.

3.4. Experiment Design

Figure 3 illustrates the methodology of the present research study. We designed the experiments as follows:

a. Build a ViT network and perform a 10-fold cross-validation using the UD dataset.
b. Evaluate the model's per-fruit performance in detecting rotten- and good-quality fruits
c. Build ViT models for each fruit and perform 10-fold cross-validation using data from the specific fruit
d. Evaluate the models' performance in detecting rotten- and good-quality fruits

Figure 3 illustrates the methodology of the present research study.

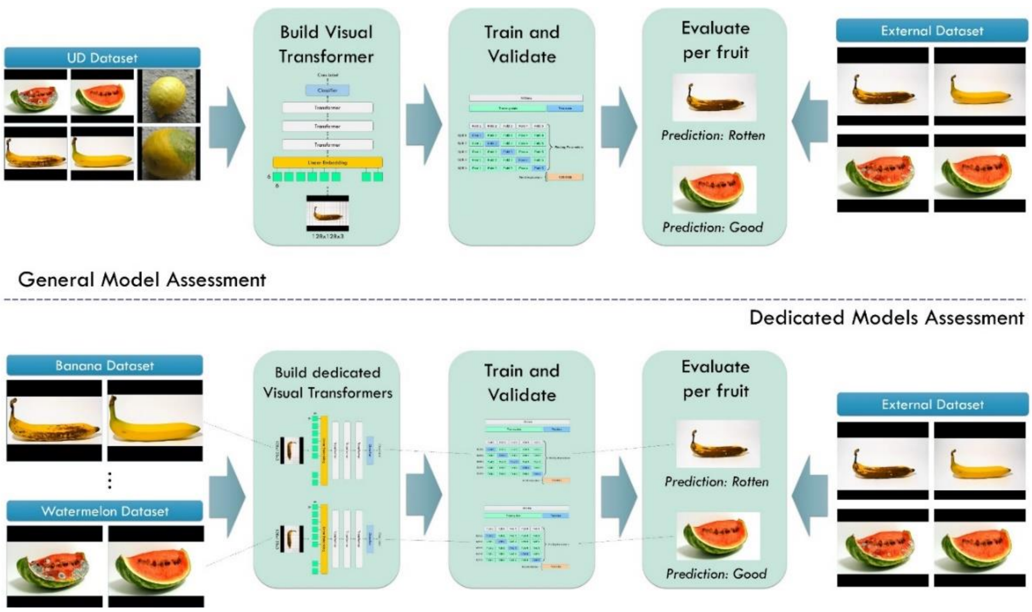

Figure 3. Research Methodology.

4. Results

4.1. General Model

In this section, we present the classification results of the general model, which was trained using the large-scale UD dataset.

4.1.1. Training and validation performance

Under the 10-fold cross-validation procedure, the general model achieves an accuracy of 0.9794. The latter is computed regardless of the fruit under examination. The model obtains 0.9886 Precision, 0.9733 Recall, and 0.9809 F1 score (**Table 2**).

Table 3. Results of the general model under a 10-fold cross-validation procedure.

Training Data	*Testing Data*	*Accuracy*	*Precision*	*Recall*	*F1*
UD	UD	0.9794	0.9886	0.9733	0.9809

The above scores represent the aggregated scores derived from each iteration over the ten folds. The model performs excellently in identifying the general condition of any fruit of the dataset. It yields 178 False-Good predictions and 424 False-Rotten predictions. Correct predictions include 15476 True-Good cases and 13137 True-Rotten cases.

4.1.2. External per-fruit evaluation

The general model has been evaluated using the external datasets of various fruits. The reader shall recall that each external dataset includes 100 good and 100 rotten fruit representations. **Table 4** presents the results.

Table 4. Results of the general model when testing with external data.

Training Data	*Testing Fruit*	*Accuracy*	*Precision*	*Recall*	*F1*
UD	Apple	0.9950	1.0000	0.9900	0.9950
UD	Banana	0.9800	0.9615	1.0000	0.9804
UD	Cucumber	0.9900	0.9804	1.0000	0.9901
UD	Grape	1.0000	1.0000	1.0000	1.0000
UD	Guava	0.9700	0.9796	0.9600	0.9697
UD	Kaki	0.9950	0.9901	1.0000	0.9950
UD	Lemon	0.9700	0.9608	0.9800	0.9703
UD	Lime	0.9750	0.9798	0.9700	0.9749
UD	Mango	0.9750	0.9897	0.9600	0.9746
UD	Orange	0.9950	0.9901	1.0000	0.9950
UD	Papaya	0.9800	0.9898	0.9700	0.9798
UD	Peach	0.9800	0.9706	0.9900	0.9802
UD	Pear	0.9700	0.9796	0.9600	0.9697
UD	Pomegranate	0.9700	0.9796	0.9600	0.9697
UD	Tomato	0.9950	0.9901	1.0000	0.9950
UD	Watermelon	0.9800	0.9706	0.9900	0.9802

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 July 2023 doi:10.20944/preprints202307.1552.v1

The general model shows remarkable performance in identifying the quality of Apples (Accuracy of 0.9950), Cucumbers (Accuracy of 0.99), Grapes (Accuracy of 1.00), Kakis (Accuracy of 0.9950), Oranges (Accuracy of 0.9950), Papayas (Accuracy of 0.98), Peaches (Accuracy of 0.98), Tomatoes (Accuracy of 0.9950), and Watermelons (Accuracy of 0.98).

Slight worse performance was recorded as far as Guavas (Accuracy of 0.9700), Lemons (Accuracy of 0.9700), Limes (Accuracy of 0.9750), Mangoes (Accuracy of 0.9750), Pears (Accuracy of 0.9700), and Pomegranates (Accuracy of 0.9700) are concerned.

It is worth noticing that the general model achieved equal or higher classification scores in the external datasets than the scores from the learning dataset (UD). This phenomenon is strong evidence of the generalisation capabilities of the model.

4.2. Dedicated Models

In this section, we present the results of the dedicated models. Each model is trained to distinguish between good and rotten images of a specific fruit. Subsequently, each model can operate using images of a single fruit variety.

4.2.1. Training and validation performance

Table 5 summarises the 10-fold cross-validation results of the dedicated models.

Table 5. Results of dedicated models under a 10-fold cross-validation procedure.

Training Data	*Testing Fruit*	*Accuracy*	*Precision*	*Recall*	*F1*
Apple	Apple	0.9948	0.9974	0.9922	0.9948
Banana	Banana	0.9904	0.9854	0.9938	0.9896
Cucumber	Cucumber	0.9887	0.9764	0.9920	0.9841
Grape	Grape	0.9661	0.9511	0.9427	0.9469
Guava	Guava	0.9965	0.9974	0.9957	0.9965
Kaki	Kaki	0.9928	0.9873	0.9982	0.9927
Lemon	Lemon	0.9981	1.0000	0.9964	0.9982
Lime	Lime	0.9991	0.9982	1.0000	0.9991
Mango	Mango	0.9625	0.9793	0.9450	0.9618
Orange	Orange	0.9971	0.9984	0.9959	0.9971
Papaya	Papaya	0.9546	0.7831	1.0000	0.8784
Peach	Peach	0.9965	0.9953	0.9953	0.9953
Pear	Pear	0.9909	0.9940	0.9861	0.9900
Pomegranate	Pomegranate	0.9964	0.9975	0.9981	0.9978
Tomato	Tomato	0.9957	0.9933	0.9933	0.9933
Watermelon	Watermelon	0.9055	0.6800	1.0000	0.8095

All models obtain high-performance metrics except for the Grape and Papaya models.

4.2.2. External per-fruit evaluation

Table 6 summarises the classification metrics of each dedicated model when predicting the classes of the external dataset.

Table 6. Results of dedicated models.

Training Data	*Testing Fruit*	*Accuracy*	*Precision*	*Recall*	*F1*
Apple	Apple	0.9950	1.0000	0.9900	0.9950
Banana	Banana	0.9950	0.9901	1.0000	0.9950

Cucumber	Cucumber	0.9850	0.9899	0.9800	0.9849
Grape	Grape	0.9900	0.9900	0.9900	0.9900
Guava	Guava	0.9850	0.9709	1.0000	0.9852
Kaki	Kaki	0.9900	1.0000	0.9800	0.9899
Lemon	Lemon	0.9950	1.0000	0.9900	0.9950
Lime	Lime	0.9800	0.9898	0.9700	0.9798
Mango	Mango	0.9500	0.9412	0.9600	0.9505
Orange	Orange	0.9950	1.0000	0.9900	0.9950
Papaya	Papaya	0.9500	0.9688	0.9300	0.9490
Peach	Peach	0.9800	0.9706	0.9900	0.9802
Pear	Pear	0.9650	0.9697	0.9600	0.9648
Pomegranate	Pomegranate	0.9950	0.9901	1.0000	0.9950
Tomato	Tomato	0.9800	0.9800	0.9800	0.9800
Watermelon	Watermelon	0.9550	0.9505	0.9600	0.9552

The dedicated models perform remarkably for apples (accuracy of 0.9950), bananas (accuracy of 0.9950), cucumbers (accuracy of 0.9850), grapes (accuracy of 0.99), kakis (accuracy of 0.99), lemons (accuracy of 0.9950), oranges (accuracy of 0.9950), and pomegranates (accuracy of 0.9950).

A slight decrease in accuracy is observed for the limes (accuracy of 0.98), peaches (accuracy of 0.98), and tomatoes (accuracy of 0.98).

The dedicated models show sub-optimal results in classifying mangos (accuracy of 0.95), papayas (accuracy of 0.95), pears (accuracy of 0.9650), and watermelons (accuracy of 0.9550).

We compared the results of the general model and the dedicated models (**Table 7**).

Table 7. Comparison between dedicated models and the general model in per-fruit accuracy measured over the external test set.

Fruit	*Dedicated Model*	*General Model*
Apple	0.9950	0.9950
Banana	0.9950	0.9800
Cucumber	0.9850	0.9900
Grape	0.9900	1.0000
Guava	0.9850	0.9700
Kaki	0.9900	0.9950
Lemon	0.9950	0.9700
Lime	0.9800	0.9750
Mango	0.9500	0.9750
Orange	0.9950	0.9950
Papaya	0.9500	0.9800
Peach	0.9800	0.9800
Pear	0.9650	0.9700
Pomegranate	0.9950	0.9700
Tomato	0.9800	0.9950
Watermelon	0.9550	0.9800

The general model is more effective than the dedicated models for predicting the quality of cucumbers, grapes, kakis, mangos, papayas, pears, tomatoes, and watermelons.

It yields equal classification accuracy in apples, oranges, and peaches. Subsequently, the dedicated models are better when built for bananas, guavas, lemons, limes, and pomegranates. Of the 16 fruits, the dedicated models performed better only in five (**Table 7**, Figure 4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 July 2023 doi:10.20944/preprints202307.1552.v1

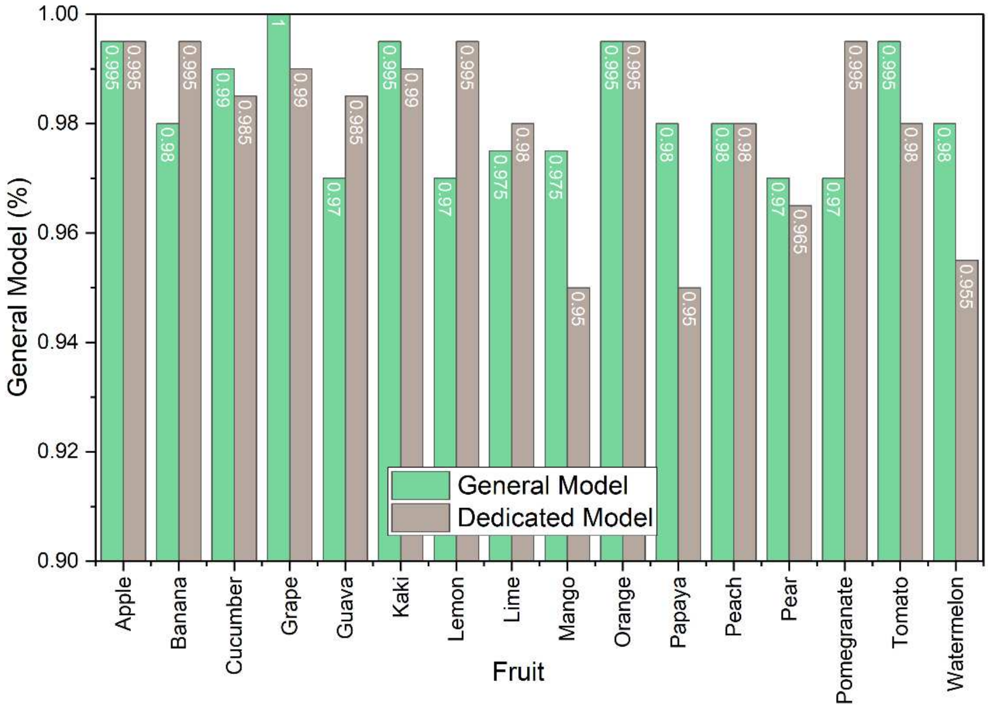

Figure 4. Column plot comparing the dedicated and the general models' per-fruit performance.

4.3. Comparison with state-of-the-art models under a 10-fold cross-validation procedure on the UD dataset

We oppose the proposed general model (ViT) to various state-of-the-art networks implemented using the Keras Python library. Each network was trained and evaluated under the same conditions. **Table 8** presents the obtained performance metrics.

Table 8. UD dataset classification of various state-of-the-art networks under a 10-fold cross-validation procedure.

Model	*Accuracy*	*Precision*	*Recall*	F1
Xception [24]	0.9524	0.9726	0.9390	0.9555
VGG16 [20]	0.9446	0.9647	0.9323	0.9482
VGG19 [20]	0.9671	0.9875	0.9516	0.9693
ResNet152 [25]	0.9785	0.9887	0.9716	0.9800
ResNet152V2 [25]	0.9606	0.9861	0.9409	0.9630
InceptionV3 [26]	0.9539	0.9711	0.9433	0.9570
InceptionResNetV2 [26]	0.9641	0.9796	0.9539	0.9666
MobileNet [18]	0.9536	0.9820	0.9319	0.9563
MobileNetV2 [18]	0.9624	0.9805	0.9499	0.9649
DenseNet169 [27]	0.9631	0.9669	0.9652	0.9660
DenseNet201 [27]	0.9598	0.9736	0.9519	0.9627
NASNetMobile [28]	0.9547	0.9819	0.9340	0.9574
EfficientNetB6 [29]	0.9660	0.9718	0.9655	0.9686
EfficientNetB7 [29]	0.9705	0.9842	0.9611	0.9725
EfficientNetV2B3 [29]	0.9591	0.9716	0.9526	0.9620
ConvNeXtLarge [30]	0.9732	0.9870	0.9634	0.9750
ConvNeXtXLarge [30]	0.9486	0.9651	0.9396	0.9522
Swin Transformer [31]	0.9632	0.9874	0.9445	0.9654
Perceiver Network [32]	0.9643	0.9711	0.9631	0.9671

Involutional Neural Network [33]	0.9635	0.9725	0.9601	0.9663
ConvMixer [7,34,35]	0.9591	0.9715	0.9529	0.9621
BigTransfer [36]	0.9574	0.9659	0.9555	0.9606
EANet [37]	0.9732	0.9874	0.9630	0.9750
FNet [23]	0.9690	0.9709	0.9722	0.9716
gMLP [38]	0.9597	0.9818	0.9435	0.9623
MLP-Mixer [36]	0.9564	0.9656	0.9539	0.9597
Attention VGG19 [39]	0.9644	0.9852	0.9489	0.9667
Visual Transformer (present study)	0.9794	0.9886	0.9733	0.9809

The top networks exhibiting equivalent performance include ResNet152 [25], ConvNeXtLarge [30], and EANet [37]. Figure 5 provides a visual comparison regarding the recorder accuracy of each model. The ViT model of the present study is slightly better than the rest. Further and extensive fine-tuning of other models may reveal that other models can perform equally well. However, the latter is beyond the scope of the present paper.

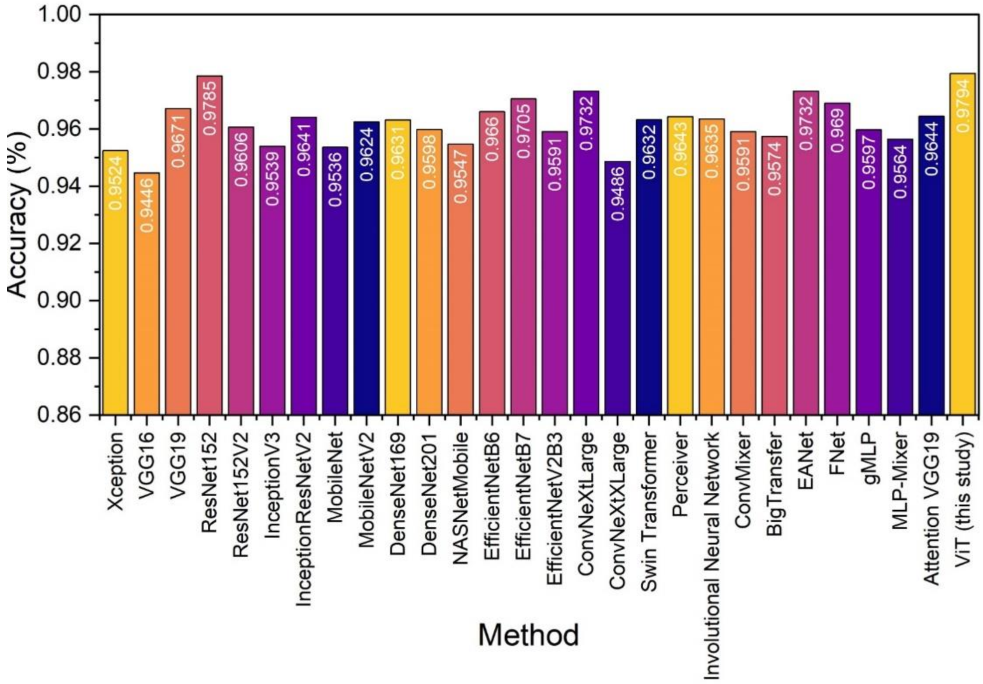

Figure 5. UD dataset classification of various state-of-the-art networks under a 10-fold cross-validation procedure.

Comparison with the literature

We collected recent literature employing either dedicated models and examining a single fruit or general models applied to various fruit representations. **Table 9** compares the general model of the study and models suggested by related works.

Table 9. Comparison with the literature.

Fruit	Study	Objective	Method(s)	Accuracy
Plum	[10]	Determination of plum maturity from images	Deep CNN	91-97%
Mangosteen	[11]	Quality assurance in	Deep CNN	97%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 July 2023 doi:10.20944/preprints202307.1552.v1

		mangosteen export		
Apple	[12]	Apple lesions identification	Deep CNN	97.5%
Banana	[13]	Distinguish between naturally and artificially ripened bananas	Neural Network	98.74%
Peach	[14]	Peach disease identification	Deep Belief Network	82.5-100%
Multiple (6)	[15]	Quality Assessment	Deep CNN	99.6%
Multiple (3)	[16]	Quality Assessment	Deep CNN	95%
Banana	[6]	Quality Assessment	Deep CNN	81.75% - 98.25%
Multiple (3)	[17]	Quality Assessment	Deep CNN	99.61%
Papaya	[19]	Quality Assessment	Deep CNN	100%
Pomegranate	[40]	Quality Assessment	Recurrent Neural Network	95%
Grapes	[41]	Quality Assessment	Artificial Neural Network	87.8%
Mango	[42]	Quality Assessment	SVM	98.6%
Apple	[42]	Quality Assessment	Deep CNN	98.6%

The comparison reveals that the suggested general model and the dedicated models are consistent with the literature and may exhibit better performance regarding specific fruits. More precisely, most studies report an accuracy between 97% and 99% in determining the quality of the fruits. The general model of this study reports per-fruit accuracies that vary between 97% and 100%.

The comparisons also verify that the general model is better than the dedicated models on many occasions.

5. Discussion

The quality of fruits is essential in determining their market value and consumer satisfaction. High-quality fruits are visually appealing, flavorful, and nutritionally dense. However, assessing fruit quality can be laborious and time-consuming, especially when done manually. This is where deep learning technology can be applied to automate and optimise the process of fruit quality assessment. By processing a large dataset of fruit images, deep learning algorithms can be trained to recognise specific patterns and features indicative of fruit quality. For instance, a deep learning model can be trained to identify specific colouration, texture, and shape characteristics that indicate freshness, ripeness, or maturity in a fruit. Deep learning can be used to assess the quality of fruits at different stages of production, from the farm to the market. Farmers can use deep learning algorithms to assess the quality of their products in real time, allowing them to make informed decisions on when to harvest or transport their fruits.

Additionally, food retailers can use deep learning technology to sort and grade fruits based on their quality, reducing waste and ensuring consistent product quality for consumers. Furthermore, deep learning can also be applied to preserve fruit quality during storage and transportation. By detecting and removing low-quality fruits before shipping, deep learning algorithms can reduce the chances of damage or spoilage during transportation, ensuring that consumers receive only high-quality fruits.

The research study presented a general ML model based on vision transformers for estimating the quality of fruit based on photographs. We proposed a general model that can be trained with multiple fruits and predict the quality of any fruit variety that participated in the training set. The general model was superior to dedicated models, where their training had been done using a single fruit variety. According to the results, a generalised model is more efficient in predicting the quality of cucumbers, grapes, kakis, mangos, papayas, pears, tomatoes, and watermelons than dedicated models. However, the classification accuracy of both the generalised and dedicated models is similar for apples, oranges, and peaches.

On the other hand, the dedicated models perform better for bananas, guavas, lemons, limes, and pomegranates. Out of the 16 fruits analysed, only five showed improved results when using dedicated models.

This suggests that while a generalised model may provide satisfactory results for most fruits, dedicated models tailored to specific fruits can significantly enhance the accuracy of the predictions, particularly for fruits with unique characteristics or qualities that are difficult to generalise.

The study has some limitations. Firstly, fruit quality can be evaluated based on several factors, including appearance, flavour, texture, and nutritional content. While the appearance of the fruit can be an indicator of quality, it is not always reliable. In some cases, the appearance of the fruit can provide some clues about its quality. For example, ripe fruit should have a bright and uniform colour, be free of bruises or blemishes, and have a firm and smooth texture. However, some exceptions exist to these guidelines, such as fruits like bananas, which develop brown spots as they ripen but are still perfectly edible. Other factors affecting fruit quality, such as flavour and nutritional content, cannot be assessed based on appearance alone. For example, a fruit may look perfectly fine but lack flavour or be low in certain nutrients. While some fruit characteristics such as colour, shape, and texture can be visually evaluated, other vital factors such as flavour, aroma, and nutritional content cannot be assessed visually. Moreover, the visual appearance of the fruit can be influenced by various factors such as lighting, the angle of the camera, and post-harvest treatments, which can affect the quality assessment. The latter can be considered a limitation of the present study.

Secondly, while studying 16 fruits provides valuable insights, it is essential to note that this sample size may not represent all fruit types. In order to fully assess the effectiveness of generalised versus dedicated models for predicting fruit quality, a more comprehensive and diverse dataset should be used.

Including a broader range of fruit varieties in future studies can help to identify patterns and trends across different types of fruit and further establish the efficacy of generalised and dedicated models. Additionally, expanding the sample size can provide more accurate and robust results, allowing for greater confidence in the findings and a better understanding of the strengths and limitations of these modelling approaches.

6. Conclusions

AI-based technologies can potentially revolutionise the fruit industry by providing objective and efficient quality assessment. The study introduced a general machine learning model based on vision transformers to assess fruit quality from images. The model outperformed dedicated models trained on single fruit types, except for apples, oranges, and peaches, where both had similar accuracy. Dedicated models were better for specific fruits like bananas and pomegranates. Overall, a generalised model worked well for most fruits, but dedicated models could improve accuracy for fruits with unique features. Fruit quality depends on multiple factors, including appearance, flavour, and nutrition. Appearance can be misleading and affected by various factors. The study has limitations in this regard. Finally, while the 16 fruits used in the study provide a valid starting point, future research should aim to include a more diverse and extensive range of fruit types to better evaluate the effectiveness of generalised and dedicated models in predicting fruit quality.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable

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

References

1. Prasad K, Jacob S, Siddiqui MW. Fruit Maturity, Harvesting, and Quality Standards. Preharvest Modul. Post-harvest Fruit Veg. Qual., Elsevier; 2018, p. 41–69. https://doi.org/10.1016/B978-0-12-809807-3.00002-0.
2. Pathmanaban P, Gnanavel BK, Anandan SS. Recent application of imaging techniques for fruit quality assessment. Trends Food Sci Technol 2019;94:32–42. https://doi.org/10.1016/j.tifs.2019.10.004.
3. Mowat A, Collins R. Consumer behaviour and fruit quality: supply chain management in an emerging industry. Supply Chain Manag Int J 2000;5:45–54. https://doi.org/10.1108/13598540010312963.
4. Zhou L, Zhang C, Liu F, Qiu Z, He Y. Application of Deep Learning in Food: A Review. Compr Rev Food Sci Food Saf 2019;18:1793–811. https://doi.org/10.1111/1541-4337.12492.
5. Dhiman B, Kumar Y, Kumar M. Fruit quality evaluation using machine learning techniques: review, motivation and future perspectives. Multimed Tools Appl 2022;81:16255–77. https://doi.org/10.1007/s11042-022-12652-2.
6. Aherwadi N, Mittal U, Singla J, Jhanjhi NZ, Yassine A, Hossain MS. Prediction of Fruit Maturity, Quality, and Its Life Using Deep Learning Algorithms. Electronics 2022;11:4100. https://doi.org/10.3390/electronics11244100.
7. d'Ascoli S, Touvron H, Leavitt M, Morcos A, Biroli G, Sagun L. ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases 2021. https://doi.org/10.48550/ARXIV.2103.10697.
8. Tian C, Xu Y, Li Z, Zuo W, Fei L, Liu H. Attention-guided CNN for image denoising. Neural Netw 2020;124:117–29. https://doi.org/10.1016/j.neunet.2019.12.024.
9. LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. Proc. 2010 IEEE Int. Symp. Circuits Syst., Paris, France: IEEE; 2010, p. 253–6. https://doi.org/10.1109/ISCAS.2010.5537907.
10. Rodríguez FJ, García A, Pardo PJ, Chávez F, Luque-Baena RM. Study and classification of plum varieties using image analysis and deep learning techniques. Prog Artif Intell 2018;7:119–27. https://doi.org/10.1007/s13748-017-0137-1.
11. Azizah LM, Umayah SF, Riyadi S, Damarjati C, Utama NA. Deep learning implementation using convolutional neural network in mangosteen surface defect detection. 2017 7th IEEE Int. Conf. Control Syst. Comput. Eng. ICCSCE, Penang: IEEE; 2017, p. 242–6. https://doi.org/10.1109/ICCSCE.2017.8284412.
12. Tan W, Zhao C, Wu H. Intelligent alerting for fruit-melon lesion image based on momentum deep learning. Multimed Tools Appl 2016;75:16741–61. https://doi.org/10.1007/s11042-015-2940-7.
13. Chowdhury A, Kimbahune S, Gupta K, Bhowmick B, Mukhopadhyay S, B.S. M, et al. Non-destructive method to detect artificially ripened banana using hyperspectral sensing and RGB imaging. In: Kim MS, Cho B-K, Chin BA, Chao K, editors. Sens. Agric. Food Qual. Saf. X, Orlando, United States: SPIE; 2018, p. 28. https://doi.org/10.1117/12.2306367.
14. Sun Y, Wei K, Liu Q, Pan L, Tu K. Classification and Discrimination of Different Fungal Diseases of Three Infection Levels on Peaches Using Hyperspectral Reflectance Imaging Analysis. Sensors 2018;18:1295. https://doi.org/10.3390/s18041295.
15. Kumar A, Joshi RC, Dutta MK, Jonak M, Burget R. Fruit-CNN: An Efficient Deep learning-based Fruit Classification and Quality Assessment for Precision Agriculture. 2021 13th Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshop ICUMT, Brno, Czech Republic: IEEE; 2021, p. 60–5. https://doi.org/10.1109/ICUMT54235.2021.9631643.
16. Bobde S, Jaiswal S, Kulkarni P, Patil O, Khode P, Jha R. Fruit Quality Recognition using Deep Learning Algorithm. 2021 Int. Conf. Smart Gener. Comput. Commun. Netw. SMART GENCON, Pune, India: IEEE; 2021, p. 1–5. https://doi.org/10.1109/SMARTGENCON51891.2021.9645793.
17. Chakraborty S, Shamrat FMJM, Billah MdM, Jubair MdA, Alauddin Md, Ranjan R. Implementation of Deep Learning Methods to Identify Rotten Fruits. 2021 5th Int. Conf. Trends Electron. Inform. ICOEI, Tirunelveli, India: IEEE; 2021, p. 1207–12. https://doi.org/10.1109/ICOEI51242.2021.9453004.
18. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv170404861 Cs 2017.
19. Behera SK, Rath AK, Sethy PK. Maturity status classification of papaya fruits based on machine learning and transfer learning approach. Inf Process Agric 2021;8:244–50. https://doi.org/10.1016/j.inpa.2020.05.003.
20. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv14091556 Cs 2015.
21. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. vol. 1. MIT press Cambridge; 2016.

22. Mou L, Zhu XX. Learning to Pay Attention on Spectral Domain: A Spectral Attention Module-Based Convolutional Network for Hyperspectral Image Classification. IEEE Trans Geosci Remote Sens 2020;58:110–22. https://doi.org/10.1109/TGRS.2019.2933609.
23. Lee-Thorp J, Ainslie J, Eckstein I, Ontanon S. FNet: Mixing Tokens with Fourier Transforms 2022.
24. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. ArXiv161002357 Cs 2017.
25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. ArXiv151203385 Cs 2015.
26. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, p. 1–9.
27. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. ArXiv160806993 Cs 2018.
28. Zoph B, Vasudevan V, Shlens J, Le QV. Learning Transferable Architectures for Scalable Image Recognition. ArXiv170707012 Cs Stat 2018.
29. Zhang L, Shen H, Luo Y, Cao X, Pan L, Wang T, et al. Efficient CNN Architecture Design Guided by Visualization 2022. https://doi.org/10.48550/ARXIV.2207.10318.
30. Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s 2022. https://doi.org/10.48550/ARXIV.2201.03545.
31. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows 2021. https://doi.org/10.48550/ARXIV.2103.14030.
32. Jaegle A, Gimeno F, Brock A, Zisserman A, Vinyals O, Carreira J. Perceiver: General Perception with Iterative Attention 2021. https://doi.org/10.48550/ARXIV.2103.03206.
33. Li D, Hu J, Wang C, Li X, She Q, Zhu L, et al. Involution: Inverting the Inherence of Convolution for Visual Recognition 2021. https://doi.org/10.48550/ARXIV.2103.06255.
34. Dai Z, Liu H, Le QV, Tan M. CoAtNet: Marrying Convolution and Attention for All Data Sizes 2021. https://doi.org/10.48550/ARXIV.2106.04803.
35. Hassani A, Walton S, Shah N, Abuduweili A, Li J, Shi H. Escaping the Big Data Paradigm with Compact Transformers 2021. https://doi.org/10.48550/ARXIV.2104.05704.
36. Tolstikhin I, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, et al. MLP-Mixer: An all-MLP Architecture for Vision 2021. https://doi.org/10.48550/ARXIV.2105.01601.
37. Guo M-H, Liu Z-N, Mu T-J, Hu S-M. Beyond Self-attention: External Attention using Two Linear Layers for Visual Tasks 2021. https://doi.org/10.48550/ARXIV.2105.02358.
38. Liu H, Dai Z, So DR, Le QV. Pay Attention to MLPs 2021.
39. Apostolopoulos ID, Aznaouridis S, Tzani M. An Attention-Based Deep Convolutional Neural Network for Brain Tumor and Disorder Classification and Grading in Magnetic Resonance Imaging. Information 2023;14:174. https://doi.org/10.3390/info14030174.
40. Mureşan H, Oltean M. Fruit recognition from images using deep learning. Acta Univ Sapientiae Inform 2018;10:26–42. https://doi.org/10.2478/ausi-2018-0002.
41. Pujari JD, Yakkundimath R, Byadgi AS. Recognition and classification of Produce affected by identically looking Powdery Mildew disease. Acta Technol Agric 2014;17:29–34. https://doi.org/10.2478/ata-2014-0007.
42. Khan MA, Akram T, Sharif M, Awais M, Javed K, Ali H, et al. CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. Comput Electron Agric 2018;155:220–36. https://doi.org/10.1016/j.compag.2018.10.013.