Smart Agricultural Technology 12 (2025) 101129

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Real-time freshness prediction for Apples and Lettuces using imaging recognition and advanced algorithms in a user-friendly mobile application

Chrysanthos Maraveas [a,*], George Kalitsios [b,1], Marianna I. Kotzabasaki [a,1], Dimitrios V. Giannopoulos [a], Kosmas Dimitropoulos [b], Anna Vatsanidou [c]

[a] Farm Structures Lab, Department of Natural Resources and Agricultural Engineering, Agricultural University of Athens, Greece
[b] Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), 57001 Thessaloniki, Greece
[c] Department of Agricultural Development, Agrofood and Management of Natural Resources, School of Agricultural Development, Nutrition & Sustainability, National and Kapodistrian University of Athens, Psahna, Evia, Greece

ARTICLE INFO

Keywords:
RGB imaging
Computer vision systems
Artificial intelligence
Mobile application
Freshness assessment
Sustainable agricultural practices

ABSTRACT

Over recent decades, consumer expectations for food quality and freshness have steadily increased. To meet these standards, fresh fruits and fresh-cut vegetables in supermarkets and other commercial outlets undergo rigorous sorting processes. Quality assessments typically focus on visible characteristics such as color, ripeness, shape uniformity, defect-free skin and flesh, and texture features like firmness, toughness, and tenderness. To automate real-time quality assurance of perishable agricultural products, we have developed a user-friendly smartphone application that enables freshness assessment of apples and lettuces using RGB data at multiple stages of the supply chain. This app utilizes image recognition technology, allowing for precise freshness assessment and estimated product lifespan. Nine deep algorithms were compared in the research for image classification including Vision Transformer (ViT), Swin Transformer, Residual Networks (ResNet), EfficientNet, ConvNeXt, DeiT, MobileNetV3, MaxViT, and TNT (Transformer in Transformer). The comparison considered three metrics, including accuracy (%), parameters (millions), and inference time (ms). Based on the findings, the MobileNetV3 was identified as the optimal deep learning architecture for the apple and lettuce classification because it maintained a good compromise between classification accuracy and mobile device resource constraints - (99.95 % and 2.5 ms for apple; 99.17 % and 2.5 million for lettuce). Such advancements offer valuable insights for policymakers, farmers, and stakeholders in making more informed decisions, thus supporting sustainable agricultural practices and improving food security across supply chains.

1. Introduction

In recent years, fresh and high-quality agricultural products have been increasingly demanded by consumers who focus on products that meet aesthetic and nutritional standards [1,2]. However, the consistency in the quality of agricultural products is affected by issues within the supply chain such as the implemented practices for harvesting, storage, and transportation. Two of the most popular foods that are vulnerable to deterioration in quality are lettuces and apples. With apples, their quality is challenged by issues such as over-ripening, blemishes on the external surface, and internal bruises due to factors such as improper storage, handling, or extended shelf-time, whereas the quality of lettuces is affected by other issues such as contamination by microbes, wilting, and discoloration due to poor handling and storage practices [3]. Subsequently, there is a need to guarantee the quality of such foods to ensure the reduction of losses from spoilage and failure to sell inventory [4,5] while improving the satisfaction of customers [1,2,6]. Additionally, predicting the shelf-life of perishable goods ensures minimal food waste levels [7,8].

In the past several years, the evaluation of fresh-cut vegetables and fruits was done using manual processes. Human workers were required to inspect food products visually alongside conveyor systems to identify defects such as injuries, bruises, diseases, and spoilage. However, the shortcoming associated with this method was its diminishing efficiency after several hours of inspecting the products continuously [9]. Therefore, it is essential to develop inspection systems that are manual, more

* Corresponding author.
E-mail addresses: c.maraveas@maraveas.gr, maraveas@aua.gr (C. Maraveas).
[1] Equal Contribution.

https://doi.org/10.1016/j.atech.2025.101129
Received 18 May 2025; Received in revised form 23 June 2025; Accepted 23 June 2025
Available online 24 June 2025

effective, accurate, and can detect defects in agricultural food products quickly to replace manual inspection systems.

A trend observed in the quality assessment of vegetables and fruits is the shift to using contactless and non-destructive techniques. These techniques are objective and evaluate the compositional and sensory features of food products while complementing tools employed within the supply chain. The benefits of adopting such methods are that they improve efficiency in terms of cost and time while enhancing operation monitoring to prevent the effects on the environment [10]. One emerging and contactless technique for the quality evaluation of vegetables and fruits is the computer vision system (CVS) [11,12]. The system employs RGB cameras to imitate human vision when inspecting fresh produce and fruits. Lorente et al. [13] observe that the RGB cameras use three monochromatic filters: red (R), green (G), and blue (B) at 700 nm, 546 nm, and 435 nm wavelengths respectively. The system also employs the conventional imaging approach (CVS-CI) when replicating human vision by evaluating images of the visible food surface to determine quality [14]. The appropriateness of the CVS-CI technique arises in that it is used in the assessment of quality features of foods including texture, shape, and size. The technique employs machine learning (ML) classification algorithms to extract dominant food features from the RGB images when assessing quality and grade.

The synthesis of diverse literature reveals how ML methods can be used to assess the quality of agricultural food products. Patel et al. [15] showcased a Support Vector Machine (SVM) when assessing the quality of orange fruits. Torkashvand et al. [16] used linear regression (LR) models and Artificial Neural Networks (ANN) to evaluate the quality of kiwi fruits by examining firmness features. Maheswaran et al. [17] showed the use of ANNs for evaluating vegetable quality [17]. Nishi et al. [18] demonstrated a convolutional neural network (CNN) when assessing the quality of fruits and vegetables using RGB-D (RGB and depth) images.

Further assessment of the literature reveals the combination of approaches using diverse models for the assessment of the quality of ML models and processing images when assessing vegetables [19]. Galal et al. [20] also combined three-band indices and decision trees when assessing the quality of orange fruits. The studies showed that employing a combinatory method enhanced the grading of fruits and vegetables.

In other studies, apps used in smartphones were adopted to evaluate the grade quality of fruits and vegetables. The work by Kaur et al. [21] showed that Android apps could be adopted to assess whether fruits were naturally or artificially ripened. Balpande et al. [22] demonstrated an app to evaluate calorie content using the processing of images to evaluate features from the RGB images. The insights from these studies indicated that diverse ML methods were adopted for the assessment of vegetable and fruit quality. However, Apostolopoulos, Tzani, and Aznaouridis [23] contradicted the views of Kaur et al. [21] and Balpande et al. [22], revealing that ML method accuracy was limited when detecting fruit quality and ripeness aspects. As such, the quality of apps was affected by constraints in hardware and sensors that degraded performance.

As opposed to previous research relying on lab-based inspection systems or using costly hyperspectral and multispectral imagery technologies, the current work is a low-cost applicable solution that allows for non-experts to assess the freshness of produce in real-time from RGB images acquired with a smartphone. While earlier studies have indicated the applicability of high-level imaging techniques and machine learning for quality assessment in a laboratory environment, it has not tried to address the imperative of a portable consumer-level platform that can be easily installed at the point of sale or in household situations.. This study fills that gap by developing and evaluating a mobile app that merges state-of-the-art deep learning architectures to classify apples and lettuces as fresh or not based on RGB images only, making advanced quality assessment accessible to ordinary consumers without a need for expert hardware or knowledge. By relying on only standard smartphone cameras and focusing on real-world applicability, this research significantly advances the field over lab-based or commercial-level solutions to provide access to effective and consistent produce quality evaluation for broad accessibility.

The research objectives are stated as follows:

i. To develop a mobile app that integrates image recognition technology to facilitate real-time quality evaluation along the supply chain, from the farms to the warehouse and retail outlets.
ii. To demonstrate a feasible solution that allows users to capture, process, and display RGB data in an intuitive format to enhance practicality and accessibility even where they have limited technical expertise.
iii. To streamline decision-making processes by providing real-time feedback on the quality of apples and lettuce.

The mobile app is developed to be portable and user-centric to ensure suitability in undertaking rapid quality checks in settings where speed and ease of use are integral. The solution also represents a significant advancement in democratizing quality evaluation, reducing reliance on manual methods of inspection, and facilitating the widespread uptake of data-driven and efficient decision-making. The app offers a practical and scalable solution and contributes to enhancing the efficiency of the supply chain, lowering the economic losses from spoilage, and reducing the environmental impact through improved produce freshness management.

The research paper advances current practice by demonstrating the effectiveness of state-of-the-art deep learning algorithms adopted for the real-time quality assessment of apple and lettuce produce that can replace manual and ineffective methods. Additionally, the outcomes from this research are integral to enhancing the literature on the research topic related to the use of machine learning for classifying fruits and vegetables. The research entailed the implementation of nine deep algorithms to classify the lettuce and apple images: Vision Transformer (ViT), Swin Transformer, Residual Networks (ResNet), EfficientNet, ConvNeXt, DeiT, MobileNetV3, MaxViT, and TNT (Transformer in Transformer). The comparison considered three metrics, including accuracy (%), parameters (millions), and inference speed(ms).

The paper is structured into five parts. The introduction briefs the reader about the background of the research and the main problem to be addressed. The adopted methodology is presented in the second section. In the third section, deep learning algorithms are described and compared in classifying apples and lettuce products. The fourth section is the discussion where the findings are analyzed and synthesized in relation to previous literature. The fifth section is the conclusion where key insights from the publication are outlined and recommendations to enhance practice and future work are showcased.

2. Methodology

2.1. Overview of study workflow

An image-based, AI-enabled system was developed and implemented in this study to automatically classify apples and lettuces into four and five quality grades respectively (Fig. 1). For apples, Class 4, the highest grade, represents the freshest produce, cropped on the first day and stored at room conditions of 21–23 °C and 30–50 % humidity. Each subsequent grade (Class 3 to Class 1) represents apples stored for five days longer than the previous class under identical conditions, with Class 1 denoting rejected produce due to significantly diminished quality. For lettuces, Class 5, the top grade, includes the freshest produce harvested on the first day and stored under the same conditions. Each subsequent grade (Class 4 to Class 1) reflects lettuces stored for two days longer than the previous class, with Class 1 comprising rejected samples.

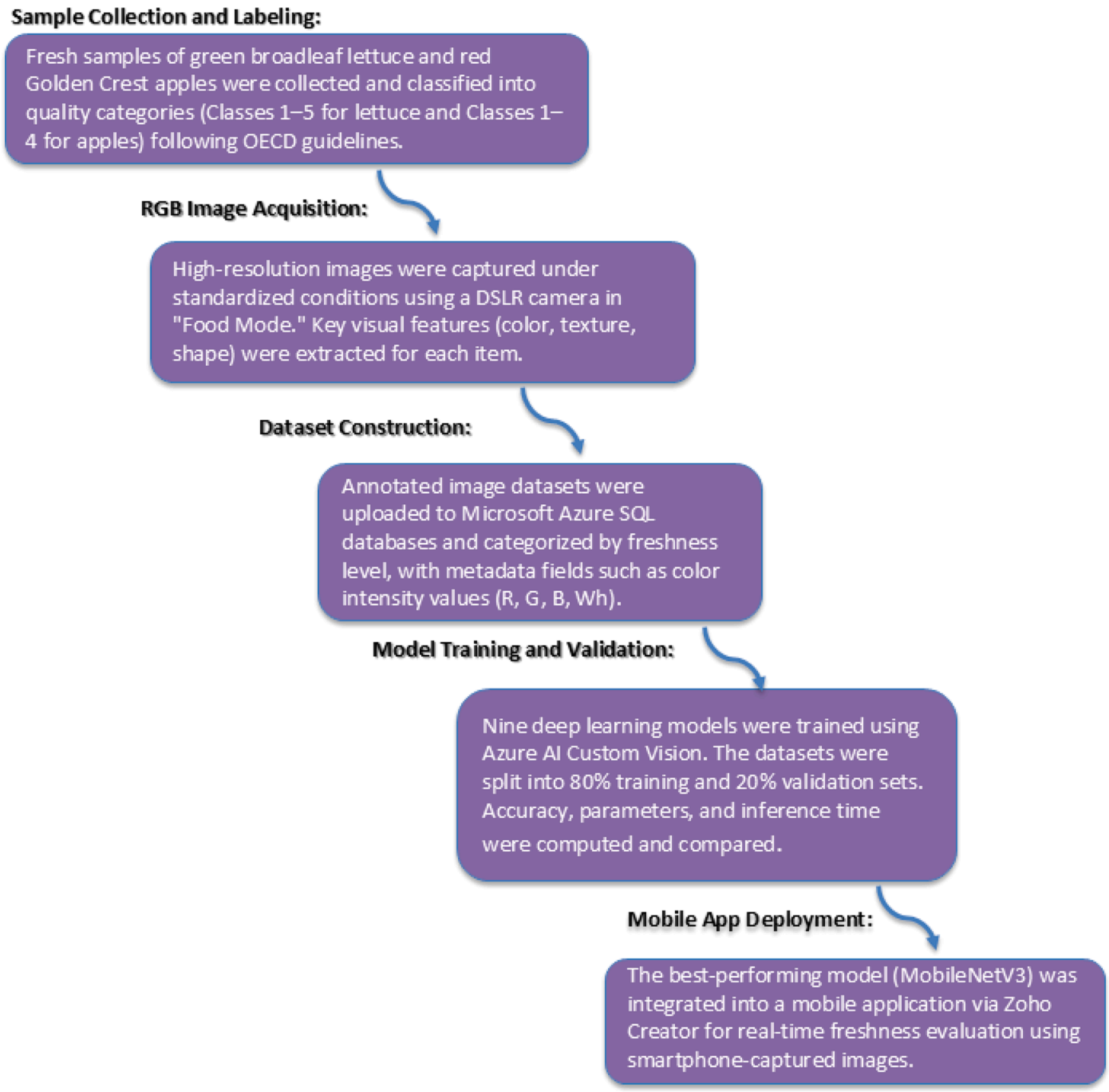


Fig. 1. Process diagram of the experimental design.

2.2. Experimental design

2.2.1. Samples collection

In this study, data samples of green broadleaf lettuce and red Golden Crest apple varieties in Greece were collected and utilized. The samples were gathered and pre-classified by experts compliant with the international standards for apples and lettuce classification [24,25]. A lot of different views or sides of the lettuces were captured for each class, making a total of three thousand six hundred (3600) instances. Samples were classified into Class 5 (for 5-star quality lettuces), Class 4 (for 4-star quality lettuces), Class 3 (for 3-star quality lettuces), Class 2 (for 2-star quality lettuces) and Class 1 (for 1-star quality lettuces with defects or‘the reject class’). Specifically, 720 RGB image instances were obtained per class for lettuces (Table 1). These instances include multiple images of the same product item, not 720 unique products per class. The representative samples of each class are depicted in Fig. 2

Table 1
Samples used for lettuce grade classification.

Class	Resolution	Number of Frames	Sample Size
Class 1	1920 × 1080	24	720
Class 2	1920 × 1080	24	720
Class 3	1920 × 1080	24	720
Class 4	1920 × 1080	24	720
Class 5	1920 × 1080	24	720

Lettuce and apple samples were captured using an RGB camera—an ordinary digital single-lens reflex camera (DSLR) (Canon EOS 850D manufactured in Tokyo, Japan, with 18.5 effective megapixels APS-C CMOS sensor) camera which was used for the extraction of external features of lettuce and apple samples. All the samples were captured using the camera’s automated "Food mode" setting.

2.2.2. RGB images acquisition

The RGB color model, also referred to as the Red, Green, and Blue model, combines three color channels—red, green, and blue—in varying proportions to generate a broad color spectrum. This model depends on devices and is employed for image sensing where it represents and displays them within electronic systems while facilitating applications involving image processing.

The RGB color space is one of the most widely used techniques in processing images and enhancing computer vision (CV). The research employed the RGB space to assess the color and texture features of the lettuce and apple products. For example, color data facilitated the analysis of the ripeness of the apples while the texture analysis revealed

Fig. 2. Lettuces classified within a range from 1-star to 5-star categories.A lot of different views or sides of the apples were captured for each class, making a total of two thousand eight hundred eighty (2880) instances. Samples were classified into Class 4 (for 4-star quality lettuces), Class 3 (for 3-star quality apples), Class 2 (for 2-star quality apples) and Class 1 (for 1-star quality apples with defects or 'the reject class'). There were 720 apple tiers per class (Table 2). The representative samples of each class are depicted below in Fig. 3.

Table 2
Samples used for apple grade classification.

Class	Resolution	Number of Frames	Sample Size
Class 1	1920 × 1080	24	720
Class 2	1920 × 1080	24	720
Class 3	1920 × 1080	24	720
Class 4	1920 × 1080	24	720

the discoloration or bruising on the peels while shape data was adopted to distinguish the apples from the lettuce.

The various samples were measured at a constant camera height which was important to determine their sizes. The same camera distance and illumination were reported in the entire process where RGB images were acquired. The same background color (white background) was used in all captured images to facilitate the apple and lettuce image segmentation. The resolution of the image samples, initially at 3984 × 2656, was reduced to an appropriate size (1920 × 1080) for use as input, with the JPEG file size limited to a maximum of 2 MB. The capture and preprocessing of image samples, including image size reduction, were performed using the built-in software of the DSLR camera, along with image analysis through the integrated Photoshop tool. The setup for getting the RGB images is shown in Fig. 4.

Color, texture, and shape were the main features extracted from the RGB JPEG file. In the RGB color space, various attributes were derived from each pixel. The apple peel's color was represented as a combination of the primary colors—red, green, and blue. 'Class 4′ apples were characterized by a red peel with minimal or no dark spots or blemishes, whereas 'reject class 1′ apples typically exhibited more dark or gray tones- features that could be effectively analyzed using the RGB color space. The RGB image files were used as input in the Azure AI Custom Vision platform (Microsoft Azure AI Custom Vision [26], Azure AI Custom Vision), an image recognition service designed for building, deploying, and improving image identifier models. An image identifier applies labels to images according to their visual characteristics. Each label represented a classification or object. A similar approach was applied to the lettuce RGB JPEC files.

2.2.3. Datasets construction

An SQL server was activated on the Azure cloud platform (Microsoft Azure SQL [27], Azure SQL). Two tables were created with mixed fields ("digits" and "files") for apples and lettuces, as follows: *"Sample_ID," "Date_Time," "R_value," "G_value," "D_value," "White (Wh)_value," "Freshness_category," and ".JPEG_image_file."* These tables correspond to SQL Database 1 (for lettuces) and SQL Database 2 (for apples). The SQL databases using the "Sample ID" field as the primary key, were published online and transformed into Azure Datalake mode (Microsoft Data Lake [28] Data Lake).

Fig. 3. Apples are classified within a range from 1-star to 4-star categories.

Fig. 4. Experimental setup for obtaining RGB images of apples using a white background and constant camera height of 20 cm above the samples.

Databases 1 & 2 have the following inputs:

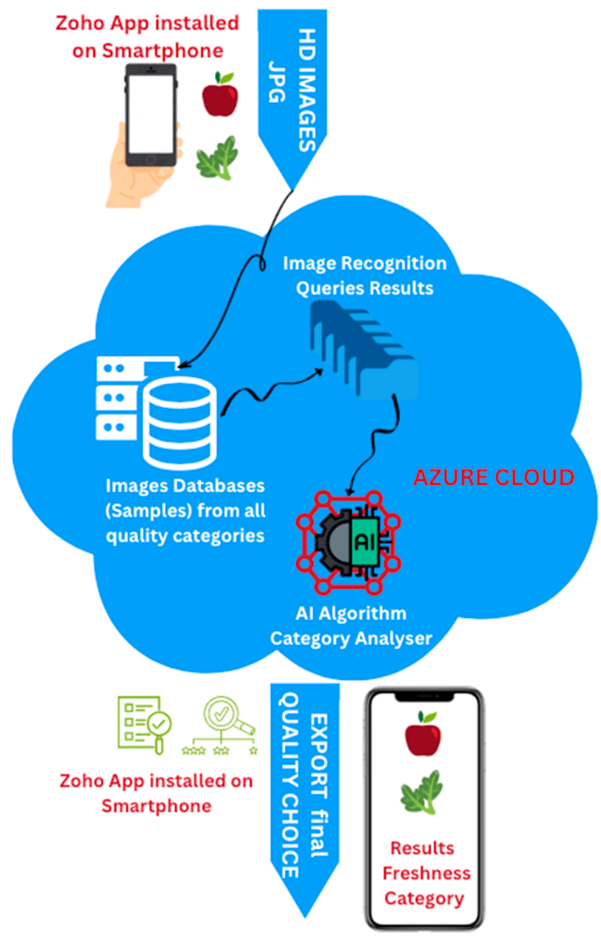
- Database 1 (Lettuces): 3600 lettuces data entries, including RGB image analysis (4-digit numbers) and real image files for each sample (max 2 MB per file), categorized into 5 freshness categories. (Online repository link: DATA SHARING | Mi4saferfood)
- Database 2 (Apples): 2880 apples data entries, including RGB image analysis (4-digit numbers) and real image files for each sample (max 2 MB per file), categorized into 4 freshness categories. (Online repository link: DATA SHARING | Mi4saferfood))

Databases 1 & 2, containing RGB image analysis data and the real freshness categories data, were directly processed using the Azure AI Custom Vision service (Microsoft Azure AI Custom Vision [26], Azure AI Custom Vision). This service was employed to train an object detector model on the available datasets, utilizing a pattern-matching algorithm. This algorithm isolates characters (glyphs) and compares them with stored templates on a pixel-by-pixel basis. An Azure AI Custom Vision (Training) resource was used to train the models, while an Azure AI Custom Vision (Prediction) resource was used to obtain predictions from the models.

2.2.4. Mobile application integration

The Zoho Creator platform (Zoho Corporation [29], Zoho Creator) was used that easily creates and publishes mobile applications for smartphone (iOS/Android) users. The design of the application involves a public online SQL query for Databases 1 & 2 (Fig. 5). The app communicates with the Microsoft Azure Platforms, and finally, Azure AI Custom Vision algorithms were used; Microsoft Azure SQL [27], Azure SQL), sending the user's food images and providing the final result to the user after the image analysis and AI processing on the platform, delivering the freshness rate and expected product life (Fig. 6). All these activities were performed using Java 12.0.3 edition [30].

More specifically, when the users photograph an apple or a lettuce with their smartphones, the application sends the photo to the Azure AI Custom Vision model via the cloud. The app returns the predicted class of freshness (for example,"Class 3 – Moderate Freshness"), along with a color-based indicator (green, yellow, red), and a relative freshness score (for example, 86/100) within seconds. The app also shows the estimated shelf-life remaining in days, derived from the data it has trained with, along with optional recommendations, for example,"consume within 2 days" or "not suitable for purchase." This feedback is delivered in an easy-to-read dashboard format, available for use by general consumers with no technical knowhow. Alerts are included to notify users with instructions such as storing or consuming the product within the right time.



5

Fig. 5. AI app structure.

3. Algorithms

3.1. Deep learning methods

Deep learning breakthroughs have impacted quality assessment and image processing in agriculture, thereby, allowing the automatic detection of complex visual data patterns [31]. The adoption of these models guarantees that the quality features from agricultural data including texture, shape, color, and defects can be determined quickly and accurately [32]. The models can assess features of the agricultural products including spoilage, ripeness, and an estimate of their shelf-life

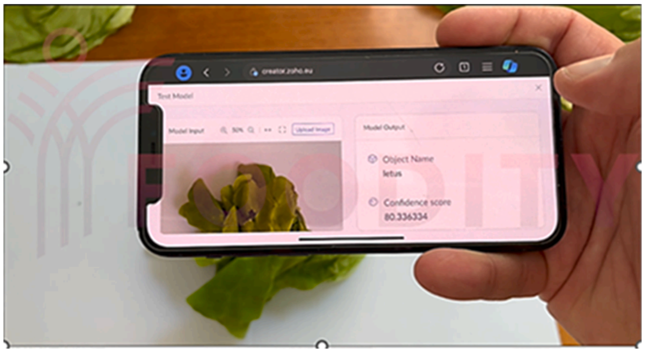


Fig. 6. The app communicates with the Microsoft Azure Platforms Microsoft Azure AI Custom Vision [26], Azure AI Custom Vision; Microsoft Azure SQL [27], Azure SQL) sending user images and providing the user with the final freshness result.

[33,34].

The adopted improvements ensure that evaluation is enhanced in real-time while ensuring solutions that are trustworthy and scalable can be developed in different application areas such as sorting and quality assurance examination [35]. Diverse deep learning architectures used for classifying lettuce and apple RGB images are outlined in the subsequent sections.

3.1.1. Vision transformer (ViT)

The model divides the target images into 16 × 16 pixel patches and translates them into vectors. Each of the vectors involves 12 transformer layers with 12 attention heads and 768 hidden units. The model employs feed-forward networks and multi-head self-attention to capture the complicated visual structures and long-range features of images.

Dosovitskiy et al. [36] show that the ViT model is appropriate in classification tasks where there is a requirement for global context awareness and local features. The method for processing images entails different patches that capture intricate and generic image features.

3.1.2. Swin transformer

Liu et al. [37] report that the Swin Transformer model employs a method of shifting windows and a structure that is hierarchical to calculate self-attention using non-overlapping windows. The generated images used for input are organized into smaller sizes of 224 × 224 pixels. The Swin Transformer model is also structured into unique phases where the resolution of the images reduces gradually while the model estimation is undertaken within local windows in each phase.

The design of the model ensures low levels of computational resources relative to the traditional shifting windows and complete self-attention that compromises efficiency when identifying local and global connections. The Swin Transformer model further ensures a balance between precision and computational efficiency.

3.1.3. ResNet

Residual Networks (ResNet) are effective when solving the problem of the vanishing gradient by employing residual connections that ensure gradients pass over different levels. He et al. [38] argue that the ResNet-50 model is divided into 50 layers structured into four phases and with filters that span between 64 and 2048. The various phases are organized into a pair of 3 × 3 convolutional layers while the model handles images scaled to 224 × 224 pixels. The model guarantees that the features of images can be learned at different abstraction levels while ensuring efficiency in resolving classification problems.

3.1.4. EfficientNet

Tan et al. [39] observe that the model entails compound scaling to adjust input resolution, width, and depth. The EfficientNet-B0 is a variation of the model that uses depth-wise separable convolutions and scaled images at 224 × 224 pixels to lower the parameter count while minimizing the burden of processing.

The model also uses 20 layers and is associated with an increase in filter numbers. Additionally, the model guarantees a balance between accuracy and computational resources to ensure it can be adopted in settings where performance is crucial in limited resource environments.

3.1.5. ConvNeXt

LIou et al. [40] posit that this model is a version of the CNN and uses methods inspired by transformer model optimization. The model improves the flow of gradient and efficiency by using depth-wise convolutions, residual connections, and layers that are normalized. The model's architecture is divided into 30 layers where each has 3 × 3 convolutions and assesses images scaled to 224 × 224 pixels.

3.1.6. Data-efficient image transformer (DeiT)

The model is an improved version of the ViT and it employs knowledge distillation to train smaller "student" models using larger "teacher" models. According to Touvron et al. [41], the model involves 12 transformer layers and uses feed-forward networks characterized by multi-head self-attention. The model scales images to 224 × 224 pixels when processing. The distillation from the teacher-to-student model ensures the DeiT functions appropriately with fewer labeled samples, guaranteeing that it is advantageous in settings involving sparsely labeled data. The model also performs effectively in classifying images while also maintaining higher training efficiency.

3.1.7. MobileNetV3

MobileNetV3 is a lightweight CNN adopted for image classification problems in scenarios where there are limited computational resources [42]. The design attains high efficiency and robust accuracy by combining squeeze-and-excitation modules with depth-wise separable convolutions. However, Wang and Sofla (2023) contradict Howard et al. [42], revealing that MobileNetV3 is challenged by difficulties in handling complex environments while the model is also likely to struggle to generalize the variability of images captured using mobile devices. As such, the constraints hinder the application of the MobileNetV3 model in addressing different classification problems. Upon integrating the Hard-Swish activation function, the MobileNetV3 model enhances computational speed and reduces overhead. The MobileNetV3 employs inverted residual blocks with linear bottlenecks that have been tuned using neural architecture search to develop its design.

3.1.8. MaxViT

The model's architecture involves a hybrid neural network and a multi-axis attention method with convolutional layers to promote the extraction of features when classifying images [43]. The model architecture ensures easier monitoring of patterns locally and globally while improving the capacity to detect visually complex aspects. The combination of the different methods ensures the model balances classification accuracy with computational efficiency. As such, MaxViT is suitable in areas requiring quick classifications while balancing scalability and high levels of performance.

3.1.9. Transformer in transformer (TNT)

The model employs two transformer setups to enhance their performance in image classification. Chen et al. [44] accentuate that the architecture of the TNT improves image classification by dividing them into patches and also embedding them in high-dimensional environments while utilizing attention methods in two levels. The inner transformers identify local details within each of the patches while the external transformers undertake global interactions between the patches. The TNT model utilizes 12 transformer layers within the outer part, where each has 768-dimensional embeddings and 16 attention heads that are enhanced using smaller transformer modules in their inner parts to improve processing at the patch levels. The hierarchical method enhances the extraction of features and improves performance on classification benchmarks while ensuring high efficiency.

3.2. Implementation details and results

In this research, two NVIDIA RTX 3090 Ti GPUs were adopted to train the ML models. To reduce the time for training and improve the use of resources, data parallelism was distributed. The input images were scaled to 224 × 224 pixels to ensure they resonated with existing pre-trained models identified from the time library [45]. The researcher adopted batch sizes of 64 that were distributed across the GPUs to guarantee optimal memory consumption. The research also adopted the AdamW optimizer with an initial learning rate of 1 × 10−4. A StepLR scheduler was adopted with a 5-epoch step size and a 0.1 decay factor which was dynamically modified with a rate of learning to improve overall convergence.

Mixed precision training was also used in the research to improve the efficiency of memory and computational performance levels. The researcher also implemented early stopping at 5 epochs to avoid issues such as overfitting while maintaining the best-performing models. Normalization and data augmentation were also ensured to guarantee the models were fine-tuned and models accurately assessed [46]. Refer to Table 3 where the comparison of the generated results based on different architectures was displayed.

In Table 3, the results showed how the various architectures compared in the classification of lettuce and apple RGB images. The differences were identified in the accuracy, inference speed, and parameters. Fig. 7 showcases a plot of these models based on accuracy and inference speeds.

In Fig. 6, the various models are plotted based on their average accuracy and inference speeds. The comparison based on these metrics facilitates the selection of the most appropriate model for deployment settings.

3.2.1. Lettuce classifier

ConvNeXt was the most accurate in classifying lettuce, with a score of 99.74 %. Swin Transformer and DeiT finished second and third, respectively, with an accuracy of 99.48 %. Despite their excellent classification abilities, the parameter counts of these models, about 28 million for ConvNeXt and Swin and 22 million for DeiT, may make them challenging to integrate into resource-constrained mobile apps.

Considering the significance of efficiency in mobile apps, MobileNetV3 was chosen as the best choice. It obtained an outstanding 99.17 % accuracy while keeping a small parameter count of around 2.5 million with rapid inference rates. This makes MobileNetV3 the optimal architecture for the lettuce classifier, as it maintains a good compromise between classification accuracy and mobile device resource constraints. Fig. 8 illustrates the lettuce classifier's confusion matrix, which clearly demonstrates the model's high accuracy in classifying lettuce into five categories.

3.2.2. Apple classifier

The evaluation of the Apple classifier models displayed different types of results. The findings indicated that the Swin Transformer was effective in settings requiring high precision as its accuracy was 100 %. However, the challenge with the model was its high resource usage with 28 million parameters being recorded. Further analysis showed the high accuracy of ConvNeXt and DeiT models (99.95 % accuracy) which provided alternatives to the Swin Transformer due to lower processing requirements of 28.6 and 22 million respectively. However, only the MobileNetV3 model ensured a compromise between efficiency (22.5 ms) and accuracy (99.95 %) for the deployment in mobile environments due to a low parameter count and faster inference speed.

MobileNetV3 was identified as the best option for the Apple classifier in mobile apps when taking accuracy and computational economy into account. This choice is perfect for real-time classification on mobile devices as it guarantees excellent performance while consuming the fewest resources possible. As demonstrated in Fig. 9, the confusion matrix for the apple classifier shows the model's performance across four unique classes, emphasizing its accuracy in correctly identifying each apple category.

3.3. Evaluation dataset and app testing protocol

For assessing the performance of the models correctly, a test set separate from the training and validation data was used. The test data were images which were not seen during learning by the models but used solely for the purpose of testing in order to deliver unbiased results.

Professional visual assessors provided test and training set ground truth labels consistent with the OECD classification standards. All image annotation was by expert agreement for consistency between classes (Classes 1–4 for apples; Classes 1–5 for lettuces).

Testing data for accuracy values and confusion matrices included here contained 1932 samples for apples and 1923 for lettuces, split evenly between respective freshness classes. Samples were collected in semi-controlled environments, with perfect camera conditions (white tablecloth, controlled lighting), but with real-world use cases similar to what the mass consumer would have with the mobile application.

User testing was performed with a pilot sample size of 200 non-expert participants. Users were given short directions via an in-app instructional guide on how to put the object on a light background, to use continuous or natural light, to avoid motion blur and to upload the images via the application interface.

These semi-controlled tests replicated field functionality of the application in market deployment-like environments but within non-field-realistic conditions. User inputs went through the application to the cloud-based AI service, and predictions were tracked and compared with expert-marked ground truth. While the app has not yet been deployed at scale, subsequent research will advance this pilot testing to progressively representative samples and uncontrolled real-world environments in an effort to more thoroughly investigate robustness and generalizability.

Table 3
Comparison of Accuracy, Parameters, and Inference Speed for Different SOTA Classification Models.

Model Architecture	Apple Classifier Accuracy (%)	Lettuce Classifier Accuracy (%)	Parameters (Millions)	Inference Speed (ms)
ViT	99.74	98.97	86.5	95.2
Swin Transformer	**100.0**	99.48	28.0	45.6
ConvNeXt	99.95	**99.74**	28.6	58.2
ResNet50	99.64	97.58	25.6	75.3
MaxViT	98.86	98.66	35.4	66.7
MobileNetV3	99.95	99.17	**2.5**	**22.5**
EfficientNet	99.64	98.76	5.3	68.4
TNT	99.90	99.48	33.1	56.0
DeiT	99.95	99.69	22.0	42.1

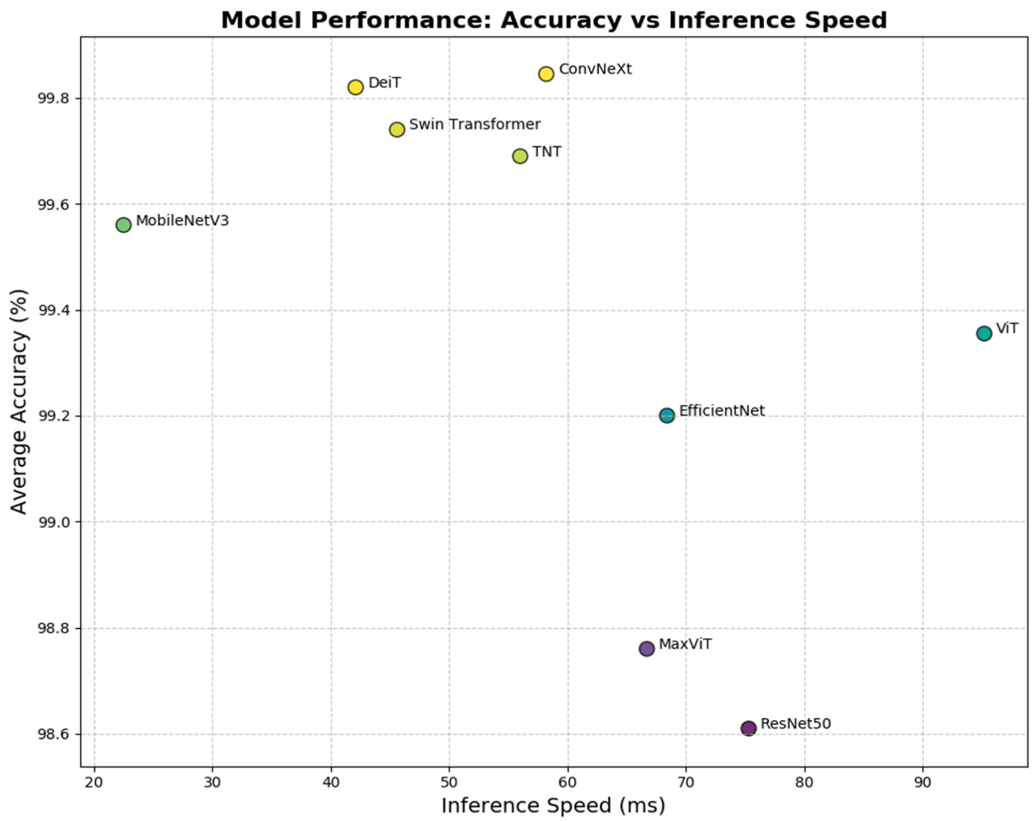


Fig. 7. Trade-off Between Accuracy and Inference Speed.

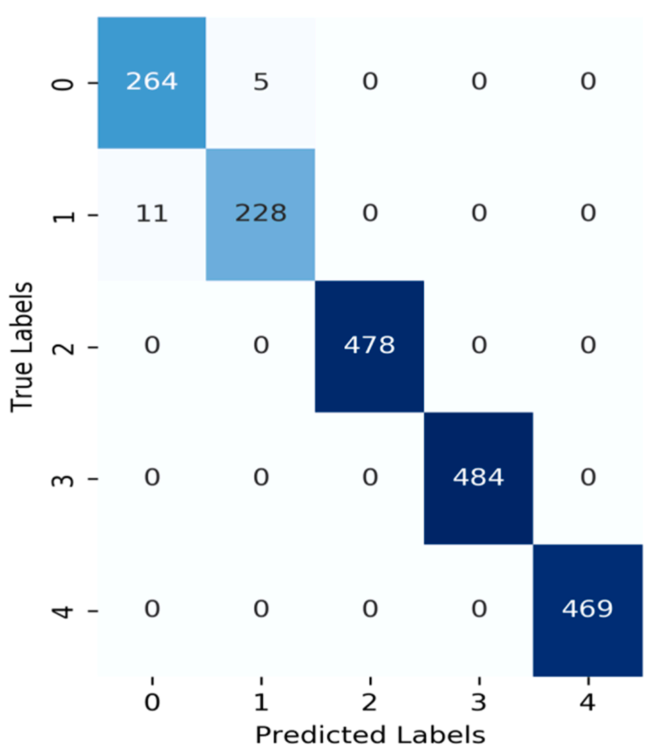


Fig. 8. MobileNetV3 Confusion Matrix for Lettuce Classification.

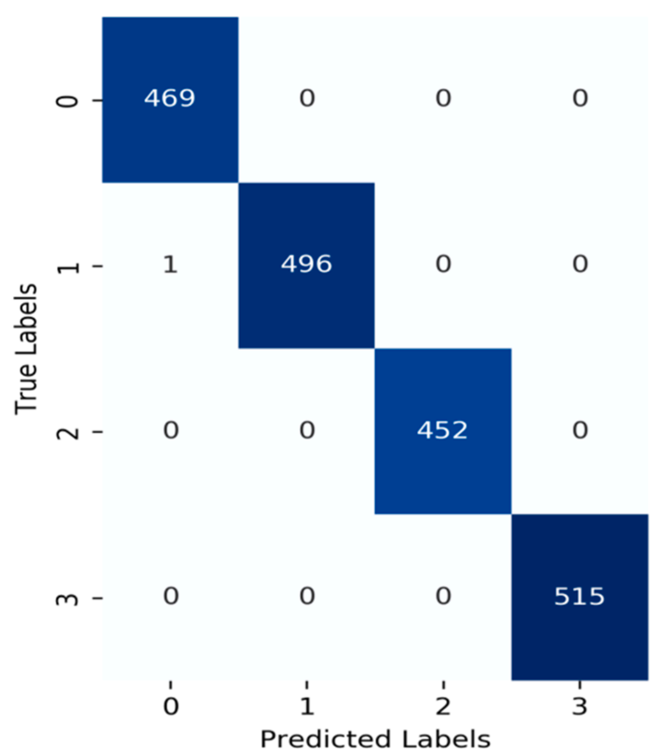


Fig. 9. MobileNetV3 Confusion Matrix for Apple Classification.

4. Discussion

The objectives of this research were to develop a mobile app that integrated image recognition technology to facilitate real-time quality evaluation along the supply chain, from the farms to the warehouse and retail outlets. The app would also allow non-technical users to capture, process, and display RGB data in an intuitive format to ensure streamlined decision-making processes by providing real-time feedback on the quality of apples and lettuce. The evaluation of the findings showed that the developed mobile app was effective in classifying images of lettuce and apples in real-time where all deep learning algorithms used demonstrated high accuracy of classification at 97–100 %. The analysis also indicated that although deep learning models such as the ViT demonstrated high accuracy of image classification at 99.74 % for the apples and 98.97 % for lettuce, its high computational resources at 86.5 million parameters disqualified it as the optimal solution. Further synthesis indicated that although the EfficientNet deep learning model had a high accuracy in classifying apples and lettuce (99.64 % and 98.76 % respectively), it had a higher resource requirement at 5.3 million parameters and higher inference speed at 68.4 ms. The interpretation showed that using the MobileNetV3 architecture would be feasible in a smartphone app where it would ensure high accuracy while consuming few computational resources. The findings reiterated past literature including Howard et al. [42] which emphasized that the MobileNetV3 model was appropriate for classifying images in areas with few computational resources. The implication was the recommendation for adopting the model in smartphone environments. However, Wang and Sofla (2023) contradicted Howard et al. [42] and revealed the limitation of the MobileNetV3 model based on challenges in image classification in complex environments. Subsequently, the challenge in using the model arose from the low level of accuracy in smartphones. The results which showed the accuracy of deep learning models in apple and lettuce classification resonated with the reviewed literature studies by Torkashvand et al. [16] and Kumar and Parkavi [19] that demonstrated the efficacy of ML algorithms in classifying the quality and grade of vegetables and fruits. The findings also reiterated past studies such as Nishi et al. [18] and Galal et al. [20] which revealed that RGB images could be classified when evaluating the grade quality of fruits and vegetables. The confluence between the findings and the past literature demonstrated the efficacy of ML models in classifying RGB images and generating highly accurate results when evaluating the quality of fruits and vegetables. However, the results contradicted a past study by Apostolopoulos, Tzani, and Aznaouridis (2023) which demonstrated that ML model accuracy when detecting the quality and ripeness of fruits was affected by constraints in hardware and sensors that degraded performance. Therefore, limitations in sensors and hardware were likely to affect the quality evaluation of vegetables and fruits. Despite the contradictions, the core aim of the research paper was addressed by demonstrating a feasible deep-learning solution to provide real-time feedback for the evaluation of the quality of lettuce and apples.

For consumers, the RGB image data from the developed databases offers an opportunity to better recognize truly fresh fruits and vegetables before purchasing, ultimately enhancing the quality of their diet. The developed mobile app was specifically designed with citizens and consumers in mind, involving their participation at every stage of development.

In order to test the possible real-world deployment of our mobile application, we organized several promotional/testing event, in fruit and super markets, where citizens tested the app directly in real conditions. During this event, participants captured 780 high-resolution images using their mobile phone cameras. The results were accurate, stable and were not sensitive to lighting conditions, distance between camera and object, colour background.

By leveraging the data and application, consumers can gain valuable knowledge about selecting high-quality fruits and vegetables in stores. Furthermore, the insights derived from our data and methods have the potential to improve social and environmental conditions. For example, the provided datasets can inform programs for marketing and producing fresh products to minimize food waste. The data also support the development of a Life Cycle Analysis (LCA) program, which serves as a cornerstone for applications like ours.

5. Conclusions and future research needs

The overarching aim of this research was to develop a mobile app that integrated image recognition technology to facilitate real-time quality evaluation along the supply chain, from the farms to the warehouses and retail outlets. By leveraging non-destructive techniques (CVS-CI), eight deep-learning classification models were compared in terms of accuracy in classifying apples and lettuces, parameters involved in computational resources, and inference speed. The comparison of the Vision Transformer (ViT), Swin Transformer, Residual Networks (ResNet), EfficientNet, ConvNeXt, DeiT, MobileNetV3, MaxViT, and TNT (Transformer in Transformer) showed that all models had a high accuracy in classification ranging 97–100 %. However, differences between the models emerged in the parameters and inference speed where the MobileNetV3 had the least values at 2.5 million and 22.5 ms respectively. Therefore, the research concluded that adopting the MobileNetV3 model was appropriate in the smartphone environment with limited computational resources.

The proposed smartphone app in this research categorized the freshness of lettuces and apples based on a 5-star and 4-star rating system respectively, and powered by an AI algorithm trained using 6480 images. By using RGB imaging technology, this guaranteed precise sorting, optimized use of resources, and sustainability of food supply chains. The central mechanism of the app was an AI algorithm used for visual recognition of food freshness using a simple photo and was thoroughly tested, showing highly reliable results.

By introducing a quick, app-based system for assessing the freshness of fruits and vegetables, this study addresses the demand for higher-quality food, reduces food waste, and fosters sustainable practices within food markets and packaging industries. Additionally, it empowers younger consumers to make informed choices about fresh produce, directly aligning with FOOD 2030′s mission to provide safe, high-quality nourishment.

The recommendation in this research is the adoption of the MobileNetV3 model for classifying apples and lettuce images within the developed smartphone app due to the low computational requirements and high inference speed. A further recommendation is to test the developed smartphone app in classifying other fruits and vegetables to compare its accuracy to the apples and lettuce.

In future work, scholars should also investigate the efficacy of using other deep learning models to compare their performance to the MobileNetV3 in terms of accuracy, parameters, and inference speed. Future research should also examine the contextual and environmental factors that are likely to affect the accuracy of classifying apples and lettuce using the developed smartphone app. Studies should examine whether factors such as time of day, lighting conditions, and specific fruit and vegetable varieties such as bananas and dragon fruit affect the efficacy of the smartphone app.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

CRediT authorship contribution statement

Chrysanthos Maraveas: Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **George Kalitsios:** Validation, Formal analysis, Data curation. **Marianna I. Kotzabasaki:** Writing – original draft, Project administration. **Dimitrios V. Giannopoulos:** Software, Resources, Investigation, Data curation. **Kosmas Dimitropoulos:** Visualization, Resources, Investigation, Formal analysis, Data curation. **Anna Vatsanidou:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Chrysanthos Maraveas reports financial support was provided by European Commission. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The MI4SaferFOOD project has indirectly received funding from the European Union's Horizon Europe research and innovation action programme, via the FOODITY – Open Call #1 issued and executed under the FOODITY project (Grant Agreement no. 101086105).

Data availability

Data base access is given from links included in manuscript

References

[1] S. Péneau, A. Linke, F. Escher, J. Nuessli, Freshness of fruits and vegetables: consumer language and perception, Br. Food J. 111 (3) (2009) 243–256, https://doi.org/10.1108/00070700910941453. Available at:.
[2] A. Saba, E. Moneta, M. Peparaio, F. Sinesio, M. Vassallo, F. Paoletti, Towards a multi-dimensional concept of vegetable freshness from the consumer's perspective, Food Qual. Prefer. 66 (2018) 1–12, https://doi.org/10.1016/j.foodqual.2017.12.008.
[3] Y. Sun, X. Zhao, Y. Ma, Z. Ma, Z. He, W. Zhao, P. Wang, S. Zhao, D. Wang, Investigation on the microbial diversity of fresh-cut lettuce during processing and storage using high throughput sequencing and their relationship with quality, Foods 11 (12) (2022) 1683, https://doi.org/10.3390/foods11121683.
[4] A. Omar, M.Z. MatJafri, Principles, methodologies, and technologies of fresh fruit quality assurance, in: A. Omar, M.Z. MatJafri (Eds.), Quality Assurance and Safety of Crops & Foods, Codon Publications, Singapore, 2013, p. 257, https://doi.org/10.3920/qas2012.0175.
[5] N.H. Sánchez, G.P.M. Cantero, A. Herrero-Langreo, Á.M. Herreros, Assessment of internal and external quality of fruits and vegetables, in: N.H. Sánchez, G.P. M. Cantero, A. Herrero-Langreo, Á.M. Herreros (Eds.), Food Engineering Series, Springer Nature, New York, 2016, p. 269, https://doi.org/10.1007/978-3-319-24735-9_9.
[6] S.R. Jaeger, L. Antúnez, G. Ares, An exploration of what freshness in fruit means to consumers, Food Res. Int. 165 (2023) 112491, https://doi.org/10.1016/j.foodres.2023.112491.
[7] C. Muriana, A focus on the state of the art of food waste/losses issue and suggestions for future research, Waste Manag. 68 (2017) 557–570, https://doi.org/10.1016/j.wasman.2017.06.047. Available at:.
[8] A. Spada, A. Conte, M.A. Del Nobile, The influence of shelf life on food waste: a model-based approach by empirical market evidence, J. Clean. Prod. 172 (2018) 3410–3414, https://doi.org/10.1016/j.jclepro.2017.11.071.
[9] Food Industry Hub. (2024, December 16). Know: visual inspection for quality control. Retrieved from [Food Industry Hub website].
[10] M.M.A. Chaudhry, F. Babellahi, M.L. Amodio, G. Colelli, A. Sahar, Image analysis, in: M.K.I. Khan (Ed.), Advances in Noninvasive Food Analysis, CRC Press, New York, 2020, pp. 200–201.
[11] A. Bhargava, A. Bansal, Fruits and vegetables quality evaluation using computer vision: a review, J. King Saud Univ. - Comput. Inf. Sci. 33 (2018) 10, https://doi.org/10.1016/j.jksuci.2018.06.002.
[12] H. Fan, J. Li, Y. Zhang, X. Tian, Q. Wang, X. He, C. Zhang, W. Huang, Online detection of defective apples using a computer vision system combined with deep learning methods, J. Food Eng. 286 (2020) 110102, https://doi.org/10.1016/j.jfoodeng.2020.110102.
[13] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O.L. García-Navarrete, J. Blasco, Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment, Food Bioproc. Tech. 5 (4) (2012) 1121–1142, https://doi.org/10.1007/s11947-011-0725-1. Available at:.
[14] A. Bhargava, A. Bansal, Fruits and vegetables quality evaluation using computer vision: a review, J. King Saud Univ. - Comput. Inf. Sci. 33 (2021) 243–257. https://api.semanticscholar.org/CorpusID:64682463.
[15] H. Patel, R. Prajapati, M. Patel, Detection of quality in orange fruit image using SVM classifier, in: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 74–78, https://doi.org/10.1109/ICOEI.2019.8862758.
[16] A.M. Torkashvand, A. Ahmadi, N.L. Nikravesh, Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR), J. Integr. Agric. 16 (7) (2017) 1634–1644, https://doi.org/10.1016/S2095-3119(16)61546-0. Available at:.
[17] S. Maheswaran, S. Sathesh, P. Priyadharshini, B. Vivek, Identification of artificially ripened fruits using smartphones, in: 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1–6, https://doi.org/10.1109/I2C2.2017.8321857. Available at:.
[18] T. Nishi, S. Kurogi, K. Matsuo, Grading fruits and vegetables using RGB-D images and convolutional neural network, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1–6.
[19] P.M. Kumar, A. Parkavi, Quality grading of fruits and vegetables using image processing techniques and machine learning: a review, in: J. Jayakumari, G. Karagiannidis, M. Ma, S. Hossain (Eds.), Advances in Communication Systems and Networks, Springer, Singapore, 2020.
[20] H. Galal, S. Elsayed, O. Elsherbiny, A. Allam, M. Farouk, Using RGB imaging, optimized three-band spectral indices, and a decision tree model to assess orange fruit quality, Agriculture 12 (10) (2022) 1558, https://doi.org/10.3390/agriculture12101558.
[21] M. Kaur, R. Sharma, ANN-based technique for vegetable quality detection, Int. J. Comput. Sci. Mob. Comput. 10 (2015) 62–70, https://doi.org/10.9790/2834-10516270.
[22] T. Balpande, N. Dhothkar, H. Satpute, N. Durbude, V. Chakole, S. Mohammad, Automatic fruit detection using artificial intelligence, SSRN Electron. J. 7 (2020) 86–89.
[23] I.D. Apostolopoulos, M. Tzani, S.I. Aznaouridis, A general machine learning model for assessing fruit quality using deep image features, AI 4 (4) (2023) 812–830, https://doi.org/10.3390/ai4040041.
[24] OECD, Apples, International Standards For Fruit and Vegetables, a, OECD Publishing, Paris, 2021, https://doi.org/10.1787/12ebba9f-en-fr.
[25] OECD, Lettuces, Curled-Leaved Endives and Broad-Leaved (Batavian) Endives, b, International Standards for Fruit and Vegetables, OECD Publishing, Paris, 2021, https://doi.org/10.1787/2fefdbbe-en-fr.
[26] Microsoft Azure AI Custom Vision, *2025, Microsoft Azure.* [Online] Available at: https://azure.microsoft.com/en-us/products/ai-services/ai-custom-vision [Accessed 05/10/2025].
[27] Microsoft Azure SQL, *2025 Family of SQL cloud databases | Microsoft Azure.* [Online] Available at: https://azure.microsoft.com/en-us/products/azure-sql/ [Accessed 05/10/2025].
[28] Microsoft Data Lake, *2025, Microsoft Azure.* [Online] Available at: https://azure.microsoft.com/en-us/solutions/data-lake/ [Accessed 05/10/2025].
[29] Zoho Corporation, n.d. *Zoho creator: best low-code custom application development platform.* [Online] Available at: https://www.zoho.com/creator/ [Accessed 05/10/2025].
[30] Oracle Corporation, 2019. *Java SE Development Kit 12.0.2 documentation.* [Online] Available at: https://www.oracle.com/java/technologies/javase/jdk12-archive-downloads.html [Accessed 05/10/2025].
[31] J. Zhou, Q. Zhang, X. Peng, X. Chen, J. Han, Deep learning for agriculture: a survey, Comput. Electron. Agric. 147 (2018) 1–13, https://doi.org/10.1016/j.compag.2018.02.014.
[32] D.R. Kouadio, Z. Shi, X. Zhou, J. Yang, Convolutional neural networks for automatic fruit recognition in agriculture, Comput. Ind. 130 (2021) 103451, https://doi.org/10.1016/j.compind.2021.103451.
[33] X. Liu, Y. Zhang, J. Wang, L. Zhao, Hybrid deep learning model for ripeness classification and quality evaluation of agricultural products, J. Food Eng. 243 (2019) 47–58, https://doi.org/10.1016/j.jfoodeng.2018.09.003.
[34] D. Wu, X. Zhao, J. Luo, W. Gao, Z. Li, Attention mechanisms in deep learning for agricultural image processing, Front Plant. Sci. 11 (2020) 620, https://doi.org/10.3389/fpls.2020.00620.
[35] J. Zhang, M. Li, L. Wang, Y. Zhang, Q. Xie, Advances in deep learning for agricultural image processing: a review, Front. Artif. Intell. 4 (2021) 39, https://doi.org/10.3389/frai.2021.00039.
[36] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, Image classification with vision transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2020, pp. 2423–2432, https://doi.org/10.1109/ICCV42600.2020.00246.
[37] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, Swin transformer: hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 10012–10022, https://doi.org/10.1109/ICCV48922.2021.00984.
[38] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[39] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: Proceedings of the IEEE/CVF International Conference on Machine Learning (ICML), 2019, pp. 6105–6114, https://doi.org/10.1109/ICML.2019.02046.
[40] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al., ConvNeXt: a ConvNet for the vision transformer era, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 5578–5588, https://doi.org/10.1109/CVPR52688.2022.00547.
[41] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou, Training data-efficient image transformers & distillation through attention. Proceedings of the International Conference on Machine Learning (ICML), 2021, pp. 10347–10357, https://doi.org/10.48550/arXiv.2012.12877.
[42] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, Q.V. Le, Searching for MobileNetV3, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1314–1324, https://doi.org/10.1109/ICCV.2019.00140.

[43] Z. Tu, H. Talebi, H. Zhang, F. Yang, A. Gholami, D. Mahajan, L. Li, MaxViT: Multi-axis vision transformer. Proceedings of the European Conference on Computer Vision (ECCV), 2022, pp. 1–19, https://doi.org/10.1007/978-3-031-19883-4_1.

[44] X. Chen, H. Touvron, M. Cord, TNT: Transformer in Transformer. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 5956–5965, https://doi.org/10.1109/ICCV48922.2021.00584.

[45] S. Ross, R. Girshick, timm: PyTorch image models. arXiv preprint arXiv 2003, 2021, p. 13772, https://doi.org/10.48550/arXiv.2003.13772.

[46] E.D. Cubuk, B. Zoph, J. Shlens, Q.V. Le, AutoAugment: Learning augmentation strategies from data, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 113–123, https://doi.org/10.1109/CVPR.2019.00021.