Classification Of Strawberry Diseases and Quality Using Different Machine Learning Methods

Kimia Aghamohammadesmaeiletabforoosh , Soodeh Nikan , Giorgio Antonini , Joshua Pearce *

Posted Date: 10 May 2024

doi: 10.20944/preprints202405.0710.v1

Keywords: computer vision; monitoring; strawberries; yield monitoring; image classification; machine learning; Vision Transformers; MobileNetV2; ResNet18

Article

Classification of Strawberry Diseases and Quality Using Different Machine Learning Methods

Kimia Aghamohammadesmaeiletabforoosh [1], Soodeh Nikan [1], Giorgio Antonini [1] and Joshua M. Pearce [1,2,*]

[1] Department of Electrical & Computer Engineering, Western University, London ON, Canada; kaghamoh@uwo.ca; snikan@uwo.ca; gantonin@uwo.ca
[2] Ivey Business School, Western University, London ON, Canada
* Correspondence: joshua.pearce@uwo.ca

Highlights

- Machine learning to recognize and classify strawberries images.
- New image dataset created for strawberries, merging 2 different ones.
- Vision Transformer, MobileNetV2, ResNet18 used to improve state-of-the-art results.
- Accuracy reached 98% and the most misclassified classes' accuracy was improved by 5%.

Abstract: Machine learning and computer vision have proven to be valuable tools for farmers to streamline their resource utilization to lead to more sustainable and efficient agricultural production. These techniques have been applied to strawberry cultivation in the past with limited success. To build on this past work, in this study two separate sets of strawberry images, along with their associated diseases, were collected and subjected to, resizing, and augmentation. Subsequently, a combined dataset consisting of 9 classes was utilized to fine-tune three distinct pre-trained models: Vision Transformer (ViT), MobileNetV2, and ResNet18. To address the imbalanced class distribution in the dataset, each class was assigned weights to ensure nearly equal impact during the training process. To enhance the outcomes, new images were generated by removing backgrounds, reducing noise, and flipping them. The performances of ViT, MobileNetV2, and ResNet18 were compared after being selected. Customization specific to the task was applied to all three algorithms, and their performances were assessed. Throughout this experiment, none of the layers were frozen, ensuring all layers remained active during training. Attention heads were incorporated into the first 5 and last 5 layers of MobileNetV2 and ResNet18, while the architecture of ViT was modified. The results indicated accuracy factors of 98.4%, 98.1%, and 97.9% for ViT, MobileNetV2, and ResNet18, respectively. Despite the data being imbalanced, the precision, which indicates the proportion of correctly identified positive instances among all predicted positive instances, approached nearly 99% with the ViT. MobileNetV2 and ResNet18 demonstrated similar results. Overall, the analysis revealed that the Vision Transformer model exhibited superior performance in strawberry ripeness and disease classification. The inclusion of attention heads in the early layers of ResNet18 and MobileNet18, along with the inherent attention mechanism in ViT, improved the accuracy of image identification. These findings offer the potential for farmers to enhance strawberry cultivation through passive camera monitoring alone, promoting the health and well-being of the population.

Keywords: computer vision; monitoring; strawberries; yield monitoring; image classification; machine learning; Vision Transformers; MobileNetV2; ResNet18

1. Introduction

The field of machine learning (ML) and computer vision (CV) are rapidly expanding within agriculture, offering a multitude of applications, including precision farming, crop monitoring, and disease detection (Shorif Uddin and Bansal, 2021). These technologies provide real-time, precise data regarding agricultural yields, and equipping farmers and agribusinesses with the necessary insights to make informed decisions in crop management (Hadipour-Rokni et al., 2023). This

 doi:10.20944/preprints202405.0710.v1

encompasses optimizing irrigation and fertilization strategies (Zhou et al., 2020). Computer vision proves invaluable in identifying issues such as diseases, pest infestations, and fruit damage (Hadipour-Rokni et al., 2023; Lello et al., 2023), facilitating timely intervention and enhancing overall crop quality (Suryanarayana 2016). Furthermore, it empowers farmers to streamline their resource utilization, encompassing water, fertilizer, and labor, ultimately leading to more sustainable and efficient agricultural practices (Chen et al. 2019).

One area that these benefits have yet to reach their full potential is in strawberry growing. Strawberries are popular fruit in Canada, with the majority of the crop being grown in Quebec, Ontario and British Columbia (Government of Canada Publications, 2019). In addition to being a tasty and nutritious fruit, strawberries also have a significant economic impact on the Canadian agricultural industry (Government of Canada Publications, 2019).

There have been four core studies applying machine vision to strawberries. First, Zheng et al. revealed that vegetable recognition and size detection could be effectively achieved using a stereo camera in conjunction with a key point detection method (Chen et al., 2019). Another study with a focus on vegetable health, (Rahamathunnisa et al., 2020) aimed to detect diseases in vegetables through the utilization of a combination of K-means clustering and support vector machines (SVMs). Transitioning to the context of strawberries, Afzaal et al. (2021) successfully detected diseases in strawberries and their leaves (Afzaal et al., 2021). This achievement was made possible through the application of classic deep learning techniques and the implementation of Region-based Convolutional Neural Networks (R-CNN) (Afzaal et al., 2021). Moreover, Puttemans et al. 2016 conducted a separate investigation that employed object detection methods to distinguish between ripe and unripe strawberries (Puttemans et al., 2016). Their methodology not only facilitated the differentiation of strawberries based on ripeness but also enabled the individual isolation of each strawberry from its cluster (Puttemans et al., 2016). In this project, an additional set of data from StrawDI ("StrawDI Dataset," 2020), introduced by Borrero (Pérez-Borrero et al., 2020) was utilized to complement the dataset from the Afzaal et al (Afzaal et al., 2021). project (Afzaal et al., 2021), allowing for a broader comparison that extends beyond diseases. Necessary modifications were subsequently made, and three pre-existing classification models were specifically trained for this task and comparison. The primary objective of this project is to provide farmers with valuable guidance concerning the most effective method for classifying their images.

In this study, two separate sets of strawberry images, along with their associated diseases, were collected and subjected to resizing and augmentation. Subsequently, a combined dataset consisting of 9 classes was utilized for fine-tuning three pre-trained classification algorithms. To address the imbalanced class distribution in the dataset, each class was assigned weights to ensure nearly equal impact during the training process. To improve accuracy, augmentation and attention layers were employed, proving particularly effective in addressing major misclassifications. All three algorithms underwent task-specific customization, and their performance was compared at the conclusion of the study.

2. Methodology

2.1. Dataset and Preparation

Two datasets were merged with the goal of identifying diseases in strawberries. The initial dataset as illustrated in Figure 1, comprises seven distinct types of strawberry diseases: angular leafspot, anthracnose fruit rot, blossom blight, gray mold, leaf spot, powdery mildew fruit, and powdery mildew leaf. Initially, the classes were separated in the training file. The number of images in each class was 287, 64, 146, 332, 452, 90, and 380, respectively. The images are RGB and sized 419 x 419 pixels. They were captured by a SAMSUNG Galaxy Note 5 (Afzaal et al., 2021; Ketabforoosh, 2023). The second dataset consists of cluttered images of strawberries that required cropping and labeling for seamless integration. For the other two classes, images from the StrawDI ("StrawDI Dataset", 2020) repository were used. These two classes, ripe and unripe, were included to enhance the completeness of the dataset. The number of images for each class in the training set is 202 for ripe and 208 for unripe. This process results in the creation of two distinct classes to distinguish between

 doi:10.20944/preprints202405.0710.v1

ripe and unripe strawberries, shown in Figure 2. Subsequently, two datasets were merged to form a more comprehensive data set.

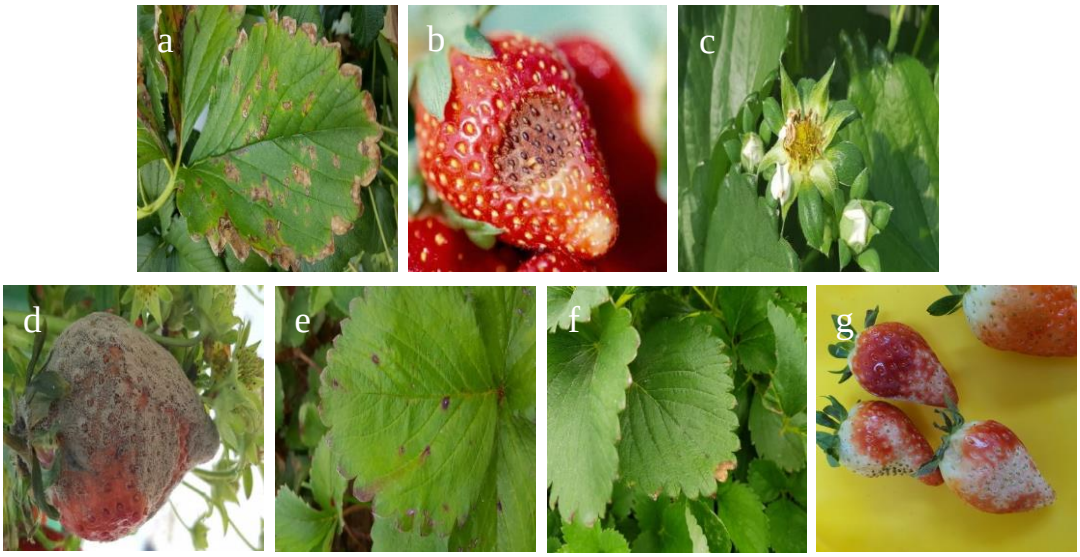


Figure 1. Images of the seven diseases in strawberries and their leaves a. Angular leafspot b. Anthracnose fruit rot c. blossom blight d. gray mold e. Leaf spot f. Powdery mildew leaf g. Powdery mildew fruit.

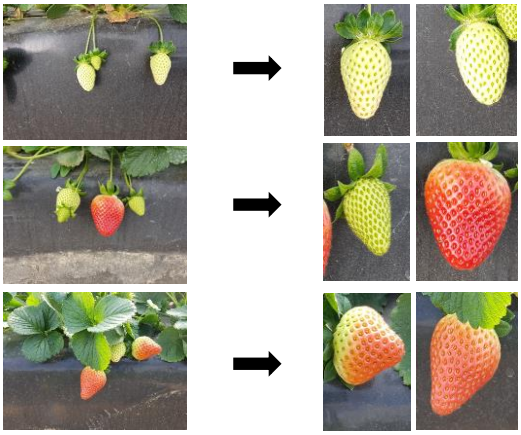


Figure 2. Strawberries dataset before & after cropping.

For this dataset, transformations are applied to ensure uniformity in the images, including resizing to 256 x 256 pixels, converting to tensors, and normalizing their pixel values based on predetermined zero mean and unit standard deviation values.

An imbalance in class distribution is initially observed in the dataset, as illustrated in Table 1. The first approach undertaken to address this issue involved generating additional images for the classes with the fewest number of images, namely anthracnose and powdery mildew fruit. Sets of new images were generated through background removal, flipping, and blurring of the existing images. Due to the limited number of images, approximately 70, achieving a balanced dataset through image generation, however, was not feasible. The other attempt to address this challenge was to adopt a technique involving a weighted sampling function, where a higher representation will have a smaller weight (Rezaei-Dastjerdehei et al., 2020). This approach assigns higher weights to the minority class samples and lower weights to the majority class samples during the model training process, amplifying the impact of the minority classes on prediction (Rezaei-Dastjerdehei et al., 2020). First the dataset is split into 0.8 and 0.2 for training and testing, respectively. Addressing the initial dataset's imbalance issue, weights were calculated and assigned to each class via the 'WeightedRandomSampler' in PyTorch ("PyTorch documentation — PyTorch 2.1 documentation,"

 doi:10.20944/preprints202405.0710.v1

2023). Table 2 shows the calculated weights for each class and the details of the augmentation step. Figure 3 schematically shows the pre-processing steps in this study. The details of the augmentation are depicted in Table 2.

Table 1. Preprocessing details.

Preprocessing	**Value**
Resize	(256, 256)
Center Crop	(224, 224)
Normalize (mean)	[0.485, 0456, 0.406]
Normalize (std)	[0.229, 0.224, 0.225]

Table 2. Distribution of each class with the weights mapped.

Class name	**Angular leafspot**	**Anthracnose fruit rot**	**Blossom blight**	**Gray mold**	**Leaf spot**	**Powdery mildew fruit rot**	**Powdery mildew leaf**	**Ripe strawberries**	**Unripe strawberries**
No. of Original	245	54	117	255	382	80	319	230	243
After Addition	245	100	150	255	382	151	319	230	243
Class weights	0.8569	3.8847	1.7481	0.7438	0.5406	2.6757	0.6490	1.0724	1.0385

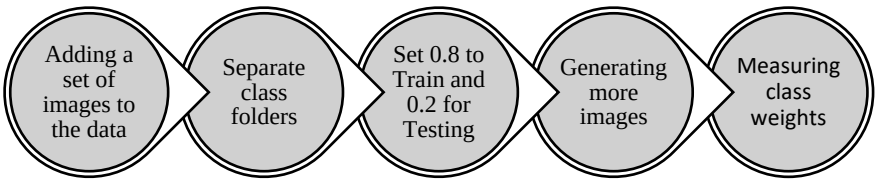


Figure 3. Pre-processing steps.

 doi:10.20944/preprints202405.0710.v1

2.2. Methodology

The prepared data is fed to three distinct pre-trained models: Vision Transformers (Dosovitskiy et al., 2021) MobileNetV2 (Sandler et al., 2019), and ResNet18 (He et al., 2016). Each of these models undergoes specific adjustments to make them suitable for the intended task, as elaborated in the following paragraphs. The Vision Transformer has demonstrated that models trained on big and varied datasets can effectively grasp fundamental visual concepts. This leads to better performance across various tasks and areas, showing improved adaptability and understanding (Caron et al., 2021). The streamlined structure of MobileNetV2 facilitates faster convergence in training, expediting both model development and deployment. Integrating MobileNetV2 into transfer learning leverages this accelerated training, resulting in more efficient utilization of computational resources (Howard et al., 2017). ResNet18, being a widely used algorithm, serves as a viable benchmark for comparison purposes in image classification applications.

Table 3 presents the specifications of the models. It is important to highlight that, to ensure fair comparison among the models, the parameters are identical and are outlined in Table 3.

Table 3. Parameters chosen for the models.

Parameter	**Value**
Optimizer	SGD
Batch size	32
Learning rate	0.001
epoch	200
momentum	0.9
Training GPU	Digital Alliance Canada (sharcnet) A100 (Google Colab)

2.2.1. Vision Transformer

A standard transformer architecture is used to process both token embeddings and 2D image data. Images are converted into sequences of flattened patches, which are then mapped to a fixed-size vector. Positional information is maintained using standard 1D position embeddings. The Transformer encoder consists of alternating layers of self-attention and multi-layer perceptron (MLP) blocks, with layer normalization and residual connections. This setup allows the model to effectively represent and process both text and image data (Vaswani et al., 2023). ViT differs from CNNs in its inductive bias, utilizing the two-dimensional neighborhood structure sparingly, primarily by dividing the image into patches. Adjustments to position embeddings are made during fine-tuning for images of different resolutions. Unlike CNNs, ViT's position embeddings contain no initial information about patch positions, requiring the model to learn spatial relationships between patches from scratch (Dosovitskiy et al., 2021). Instead of raw image patches, the input sequence can be made from feature maps of a CNN. In this hybrid model, patches from the CNN feature map undergo patch embedding projection. Patches with a spatial size of 1x1 flatten the feature map's spatial dimensions, projecting it to the Transformer dimension. Classification input and position embeddings are added as described (LeCun et al., 1989). The encoder part of the original Transformer architecture is employed by the ViT, and the decoder is not utilized. A sequence of embedded image patches, with a learnable class embedding prepended to the sequence, is taken as input to the encoder, which is augmented with positional information. The self-attention mechanism, a key component of the transformer architecture, is employed. Importance scores are assigned to patches by the model,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2024 doi:10.20944/preprints202405.0710.v1

allowing it to understand the relationship between different parts of an image and focus on the most relevant information. This aids in better comprehension of the image and enables the model to perform various computer vision tasks. Following this, a classification head attached to the output of the encoder receives the value of the learnable class embedding to output a classification label. Figure 4 illustrates all of these processes.

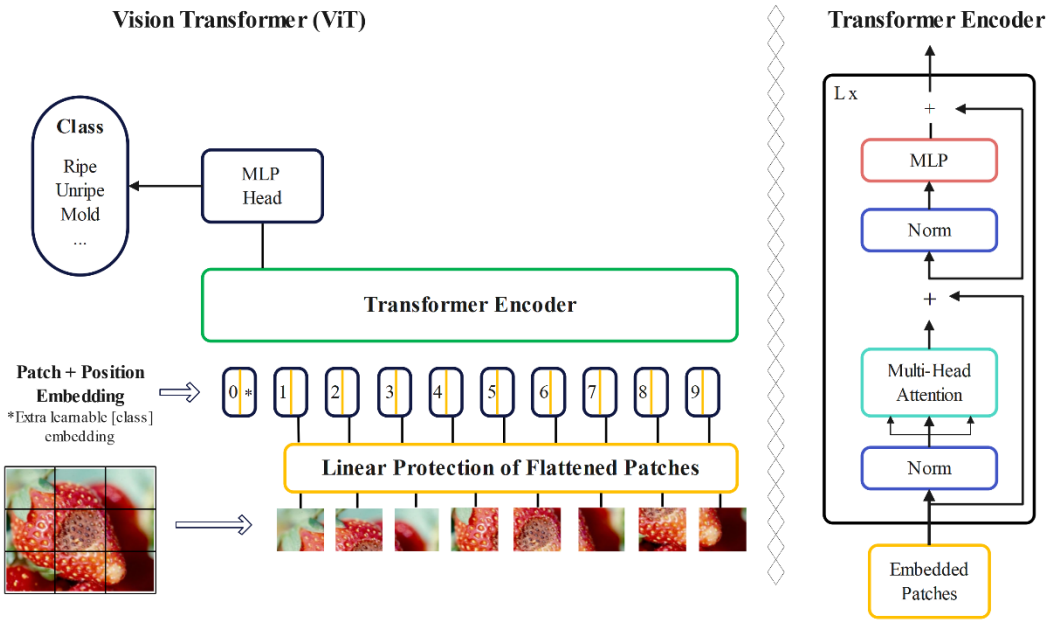


Figure 4. Architecture of a ViT.

2.2.2. MobileNetV2

MobileNetV2 is a specialized type of CNN designed for a range of visual tasks, particularly useful in agriculture (Sandler et al., 2019). Its standout feature is efficiency, crucial in scenarios with limited computational resources, (Chen et al., 2021). Its ability to achieve high accuracy with a reasonable number of parameters make it suitable for real-time applications such as crop monitoring and disease identification in agriculture.

In MobileNetV2, two types of blocks are present: one being a residual block with a stride of 1, and the other, a block with a stride of 2 for downsizing. Both types consist of three layers each. Firstly, a 1×1 convolution layer followed by ReLU6 activation is applied. Subsequently, the depth-wise convolution layer is employed, followed by another 1×1 convolution layer without any non-linearity. The computational cost and parameters of the primary network, with a width multiplier of 1 and a resolution of 224×224, are noted to be 300 million multiply-adds and 3.4 million parameters, respectively. However, the performance trade-offs are further explored for various input resolutions ranging from 96 to 224 and width multipliers from 0.35 to 1.4, leading to computational costs up to 585M multiply-adds and model sizes between 1.7M and 6.9M parameters. Notably, the removal of ReLU6 at the output of each bottleneck module results in improved accuracy. Additionally, incorporating shortcuts between bottlenecks yields better performance compared to shortcuts between expansions or those without any residual connections.

 doi:10.20944/preprints202405.0710.v1

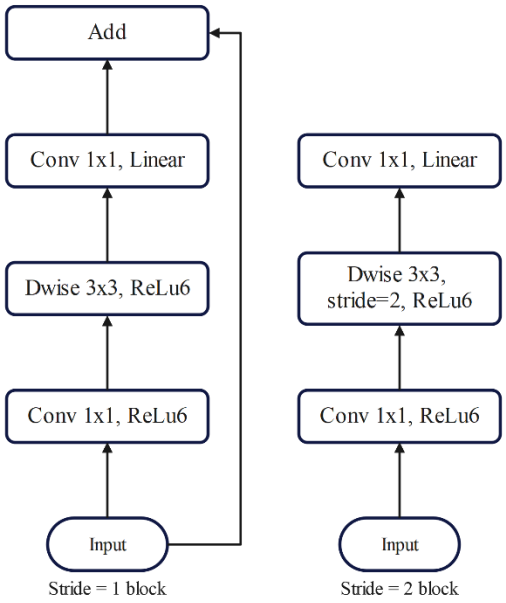


Figure 5. Original MobileNetV2's architecture.

2.2.3. ResNet18

ResNet18's key feature is its depth with many layers, which helps to extract distinctive features in complex image classification tasks (He et al., 2016). It deals with a common problem called "vanishing gradient", which can make training difficult. To address this, "residual connections" allow the network to skip certain layers during training, making it easier to train deep models (He et al., 2016). This approach demonstrated that deeper networks achieved improved optimization and accuracy (Szegedy et al., 2014). Traditional deep networks faced difficulties in training, however, and increasing the number of layers did not guarantee better learning outcomes. As deeper networks converged, accuracy would plateau and then rapidly decline. Residual learning tackled this issue by learning residual mappings rather than direct mappings. The original ResNet-18 architecture comprises eighteen layers, known as residual including convolutional layers with 3×3 filters and down sampling layers with a stride of 2. These blocks play a crucial role in improving how the network learns intricate features from input data. Throughout the network, residual shortcut connections were inserted between layers, either maintaining the same dimensions or adjusting for dimensionality changes. Residual block operation can be expressed as follows:

$$F(x) = \mathrm{H}(x) - x \tag{3}$$

where, F(x) is the residual function to be learned, x is the input to the block, and H(x) is the underlying mapping. The residual connection facilitates the learning of the residual function, mitigating the vanishing gradient problem. Another aspect is the use of global average pooling, a method that simplifies information before making final predictions. This pooling helps to reduce the spatial dimensions of feature maps (He et al., 2016).

 doi:10.20944/preprints202405.0710.v1

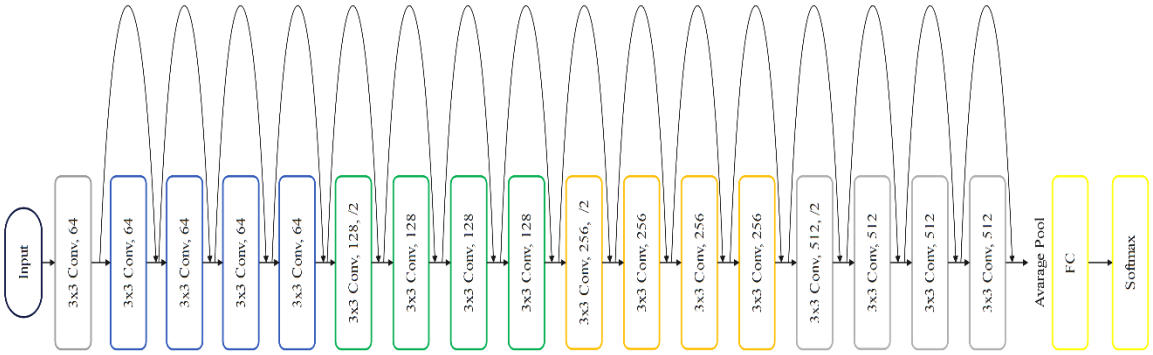


Figure 6. Original ResNet18's architecture.

2.3. Hyperparameter Optimization & Attention Mechanism

Machine learning algorithms often rely on hyperparameters which have to be chosen through an automatic hyperparameter optimization (HPO) in order to get reliable and reproducible results (Andonie, 2019). GridSearch (GS) is a method involving the systematic evaluation of hyperparameter combinations by discretizing their ranges. Numeric and integer hyperparameter values are typically evenly spaced within their specified constraints, with the number of distinct values per hyperparameter termed the grid's resolution. Optimization of categorical hyperparameters involves considering either a subset or all possible values. HPO methods streamline the process of finding optimal hyperparameter configurations, enhancing the performance and reproducibility of ML models. In this research, GridSearch (Great Learning Team, 2023) was used in every algorithm to select the best learning rate to fine-tune the model during validation. To determine the suitable learning rate, a gradual approach was taken. Initially, a low learning rate of 0.00001 was implemented for warm-up purposes, followed by a gradual increase to 0.1. Subsequently, after optimization, a value of 0.001 was identified as the optimal learning rate, which was then employed consistently across all models to ensure fair comparison. Cross Entropy Loss and the SGD optimizer were employed, with a learning rate of 0.001. The model underwent 5-fold cross-validation, to enhance models' generalization and reduce the risk of overfitting.

To understand feature maps and introduce attention mechanisms, it is essential to recognize that patterns and combinations of low-level features are captured by intermediate layers. A balance between low-level and high-level information is provided by extracting features from these layers. High-level semantic information, contributing to the understanding of more abstract concepts, is captured by later layers and residual blocks.

Attention mechanisms were introduced to specific layers responsible for these misclassifications, directing the model's focus to distinct parts of input images (Niu et al., 2021).

 doi:10.20944/preprints202405.0710.v1

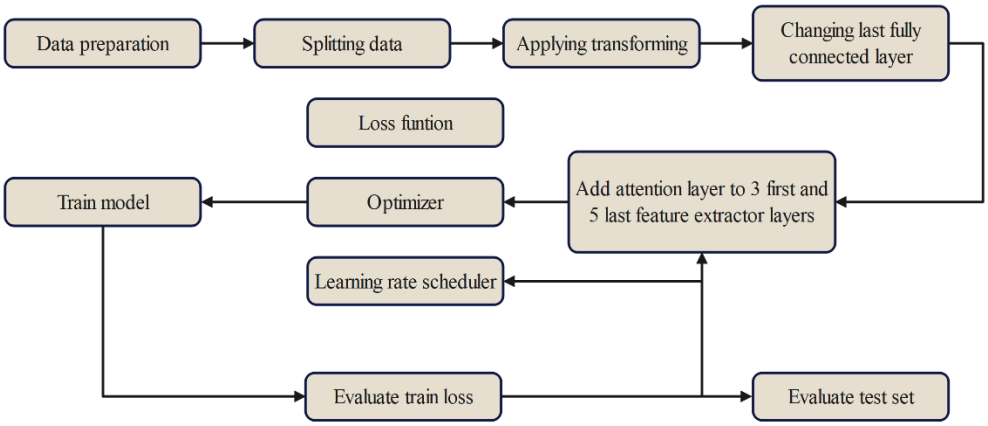


Figure 7. Procedure for the algorithms.

2.4. Evaluation Metrics

Accuracy measures the fraction of correctly classified instances in the total number of instances and is deemed effective when class distribution is balanced. Precision gauges the fraction of accurately classified positive instances in relation to all instances classified as positive, indicating how many of the predicted positive instances are truly positive. It is a suitable metric in scenarios where it is crucial to minimize the occurrence of false positives (Gad 2020). Recall, also referred to as sensitivity or true positive rate, assesses the fraction of accurately classified positive instances relative to all the positive instances, indicating the number of actual positive instances that were correctly identified as positive. It is advantageous in situations where missing a positive instance has significant consequences. On the other hand, F1-score, which is the harmonic mean of precision and recall, is a combined metric that balances the values of precision and recall, thus providing a single score that is useful when both false positives and false negatives need to be considered. Accuracy, precision, recall, and F1-score, defined in the following equations, are commonly employed evaluation metrics in machine learning for quantifying the performance of a classifier.

$$Accuracy = \frac{True\ Negative + True\ Positive}{True\ Negative + False\ Positive + True\ Positive + False\ Negtive} \quad (4)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (5)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (6)$$

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

3. Results

The implementations were performed using the library PyTorch (Torch 2.1) of Python with the support of the Digital Alliance of Canada ("Digital Research Alliance of Canada," 2023) and Google

 doi:10.20944/preprints202405.0710.v1

Colaboratory which provided the GPU resources to accommodate the enhanced processing demands posed by the extensive dataset and prolonged training epochs.

The results over the three models are compared using the accuracy, precision, recall and F1-score metrics as shown in Table 4. Based on the work of (Afzaal et al., 2021), an average precision of 82.43% was attained. In this investigation the precision in each class was improved. Additionally, all three algorithms utilized are different from ResNet101, as employed in the original study.

In this study, to enhance feature representation in models based on convolutional neural networks (MobileNetv2 and ResNet18) attention mechanisms were employed. Custom modules were employed for efficient feature extraction while minimizing computational complexity. Additionally, attention mechanisms were integrated into the CNN architecture through a module named AttentionModule, allowing dynamic adjustment of feature importance. AttentionModules are embedded into key feature extraction layers of a pre-trained CNN model, specifically the first five and last layers of convolutional layers. For the Vision Transformer, the ViT feature extractor is loaded to extract features from images. A collate function is defined to convert batches of data into tensors. The model was trained, evaluated, and metrics were logged. Optionally, a model card was created with information about the fine-tuning process, dataset, and tags. The trained model was pushed to the Hugging Face Model Hub. Each step contributed to the comprehensive process of loading, preparing, training, and evaluating the model for image classification.

The models were evaluated on the 20% dataset used as test set, with attention mechanisms impacting feature representation and overall model performance being assessed through experimental validation.

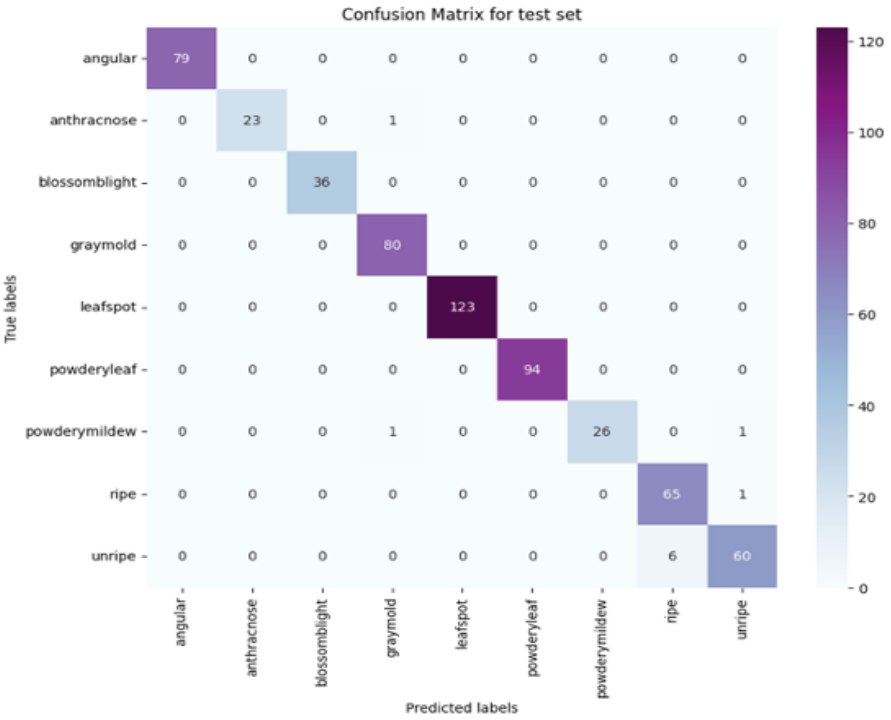
Table 4. - Evaluation results for each model.

Model	**Precision**	**Recall**	**F1-score**	**Accuracy**
Vision Transformer	0.983	0.983	0.983	0.984
MobileNetV2	0.980	0.979	0.979	0.981
ResNet18	0.979	0.978	0.978	0.979

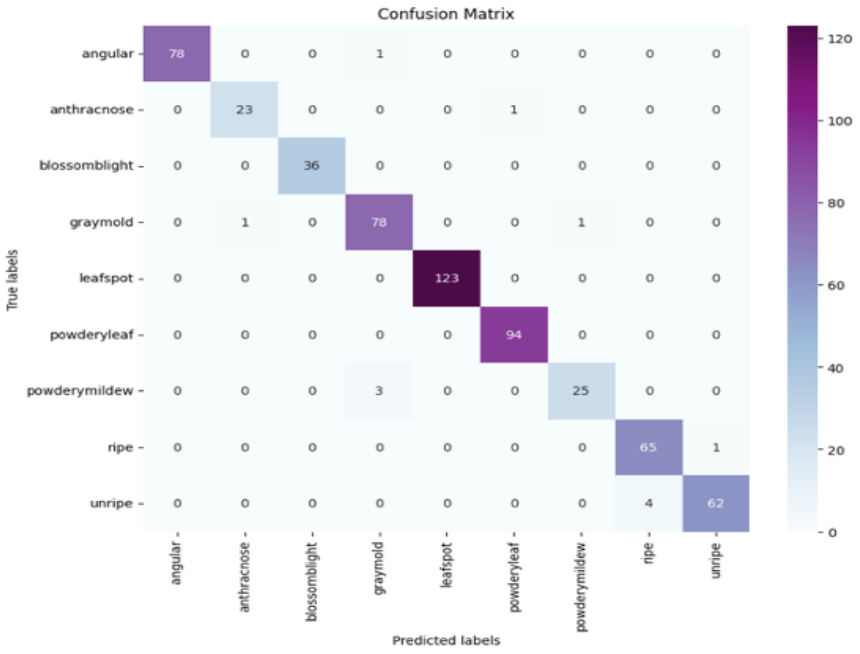
Moreover, these metrics were utilized to analyze the confusion matrix of the predictions. The confusion matrix shows the number of true positives, false positives, true negatives, and false negatives for each class. By analyzing the confusion matrix, we can identify which classes appeared to be more difficult to predict and focus on improving the performance of the model on those classes. Figure 8 displays the confusion matrices for ViT, MobileNetV2, and ResNet18.

 doi:10.20944/preprints202405.0710.v1

a)



b)



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2024 doi:10.20944/preprints202405.0710.v1

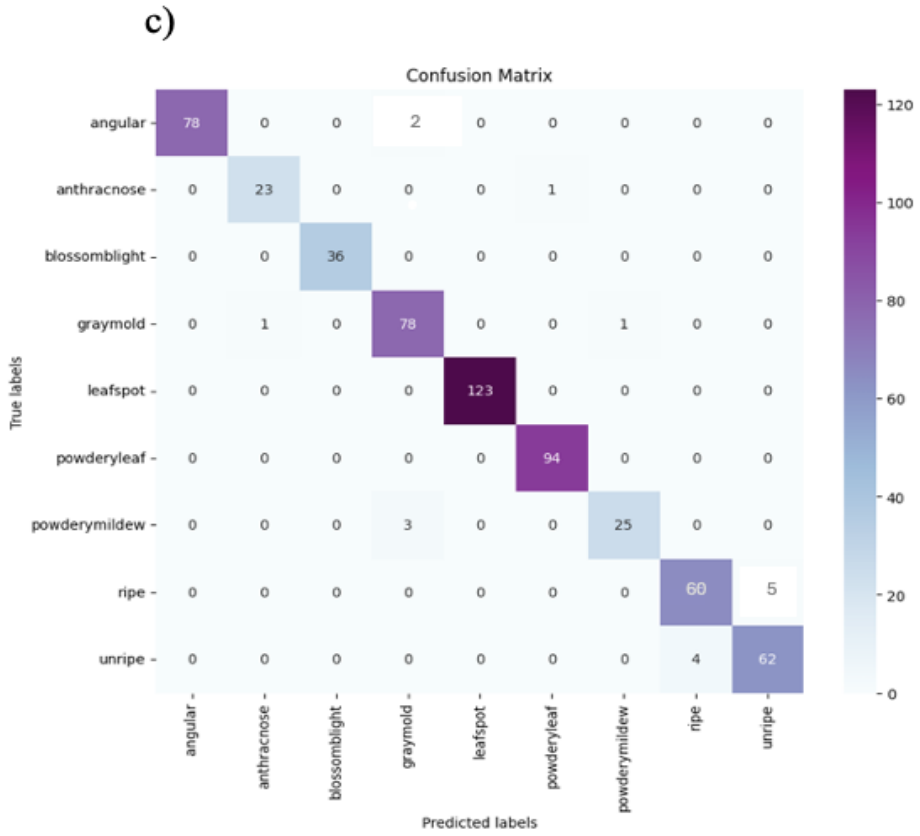


Figure 8. Confusion matrix for a) Vision Transformer, b) MobileNetV2, and c) ResNet18.

Upon an analysis of the confusion matrices, it is evident that, for this specific task, after the modifications, the ViT is the more appropriate selection. The observation can be made that there are more true positives, and the occurrences of false negatives and positives are reduced. In order to ensure the model's ability to distinguish between ripe and unripe strawberries, an unripe image was deliberately placed in the ripe folder for testing purposes. The confusion matrix generated for the ViT demonstrates that only one unripe image was misclassified, providing insight into the model's performance in correctly identifying ripe and unripe strawberries.

4. Discussion

From the results shown above, the Vision Transformer model presents better performance overall. ViT, MobileNetV2, ResNet18, reached their highest accuracy values of 98.4%, 98.1%, 97.9% respectively; in this particular context though, where the class distribution is relatively not balanced, accuracy loses reliability. In spite of this, the precision, defined in Equation 5 indicating how many of the predicted positive instances are truly positive, reached almost 98% with the ViT. It is a metric often used to minimize the occurrence of false positives (Gad, 2020) resulting, in this case, in a more reliable and less waste of good and healthy strawberries. Food waste is a major issue (Girotto et al., 2015) and reducing food waste (Dhiman, 2020) could help feed many of those suffering from food insecurity and starvation unnecessarily (Denkenberger and Pearce, 2015; Meyer and Pearce, 2022).

Overall, effective discrimination between strawberries and non-strawberries was achieved by all three of the algorithms. The classification task faced increased difficulty with images of diseased leaves due to their higher visual complexity and crowding. Additionally, addressing one of the challenges encountered in this project, the imbalanced and small-sized nature of the image dataset necessitated careful consideration, augmentation, and the assignment of appropriate weights.

 doi:10.20944/preprints202405.0710.v1

5. Future Work

Finally, it should be pointed out that a more balanced data set could potentially alter the results. In addition, by fixing the distance between camera and strawberries, future work could also enable the determination of size of the strawberries for automatic yield monitoring in both conventional (Kouloumprouka et al., 2024; Oğuz et al., 2022) and agrivoltaics-based crop systems (Dinesh and Pearce, 2016; Wydra et al., 2023; Widmer et al., 2024). This would be the key step in a fully autonomous system for strawberry harvesting (Woo et al., 2020).

By leveraging a more extensive dataset featuring high-quality images, a larger and balanced dataset could be generated, enabling the implementation of more sophisticated algorithms with appropriate modifications to facilitate the generalization of the models. The work presented represents the inception of a project aimed at the integration of machine learning into the quality control process for berries, particularly strawberries. The objective was to develop a model, such as MobileNetV2, capable of advancing the current standards in the classification and recognition of strawberry status through images. This model can be integrated with a vertical grow wall designed for strawberry cultivation. The wall system organizes plants in rows and facilitates water circulation from a reservoir, moving along rails to the top before descending. Throughout the growth cycle on these vertical walls, daily images can be captured to monitor the progress and health of the strawberries. Additionally, through the utilization of cameras and the application of image preprocessing methods to mitigate the impact of sunlight, these algorithms can be effectively extended for outdoor farming applications.

6. Conclusions

Although all three models outperformed the evaluation of strawberries presented in the past, the MobileNetV2 model presents better performance overall with an accuracy of 96%. Throughout the application of this novel method, improvements were achieved in accurately identifying various classes, with the ability to discern the diseased type even in instances of misclassification, although they were specific to the categorization of disease types. Importantly, none of the images depicting disease were erroneously classified as healthy. While acknowledging that these enhancements are tailored specifically to the current dataset, the groundwork has been laid for the establishment of a larger and more comprehensive database, which could prove invaluable for all strawberry growing in the future.

Acknowledgments: This work was supported by the Weston Family Foundation through the Homegrown Challenge, Carbon Solutions @ Western, and the Thompson Endowment.

References

1. Afzaal, U., Bhattarai, B., Pandeya, Y.R., Lee, J., 2021. An Instance Segmentation Model for Strawberry Diseases Based on Mask R-CNN. Sensors 21, 6565. https://doi.org/10.3390/s21196565
2. Andonie, R., 2019. Hyperparameter optimization in learning systems. J. Membr. Comput.
3. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging Properties in Self-Supervised Vision Transformers. https://doi.org/10.48550/arXiv.2104.14294
4. Chen, J., Zhang, D., Suzauddola, M., Zeb, A., 2021. Identifying crop diseases using attention embedded MobileNet-V2 model. Appl. Soft Comput. 113, 107901. https://doi.org/10.1016/j.asoc.2021.107901
5. Chen, Y., Lee, W.S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., He, Y., 2019. Strawberry Yield Prediction Based on a Deep Neural Network Using High-Resolution Aerial Orthoimages. Remote Sens. 11, 1584. https://doi.org/10.3390/rs11131584
6. colab.google [WWW Document], 2024. colab.google. URL https://colab.google/ (accessed 3.24.24).
7. Denkenberger, D., Pearce, J.M., 2015. Feeding everyone no matter what: managing food security after global catastrophe. Academic Press, London.
8. Dhiman, S., 2020. Sustainable Social Entrepreneurship: Serving the Destitute, Feeding the Hungry, and Reducing the Food Waste, in: Marques, J., Dhiman, S. (Eds.), Social Entrepreneurship and Corporate Social Responsibility, Management for Professionals. Springer International Publishing, Cham, pp. 193–208. https://doi.org/10.1007/978-3-030-39676-3_13

 doi:10.20944/preprints202405.0710.v1

9. Digital Research Alliance of Canada [WWW Document], 2023. . Digit. Res. Alliance Can. URL https://alliancecan.ca/en/node/10 (accessed 12.6.23).
10. Dinesh, H., Pearce, J.M., 2016. The potential of agrivoltaic systems. Renew. Sustain. Energy Rev. 54, 299–308. https://doi.org/10.1016/j.rser.2015.10.024
11. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929
12. Gad, A.F., 2020. Accuracy, Precision, and Recall in Deep Learning [WWW Document]. Pap. Blog. URL https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/ (accessed 12.6.23).
13. Girotto, F., Alibardi, L., Cossu, R., 2015. Food waste generation and industrial uses: A review. Waste Manag., Urban Mining 45, 32–41. https://doi.org/10.1016/j.wasman.2015.06.008
14. Government of Canada Publications, 2019. Crop Profile for Strawberry in Canada [WWW Document]. URL https://publications.gc.ca/site/archivee-archived.html?url=https://publications.gc.ca/collections/collection_2009/agr/A118-10-17-2005E
15. Great Learning Team, 2023 Hyperparameter Tuning with GridSearchCV. URL https://www.mygreatlearning.com/blog/gridsearchcv/ (accessed 12.6.23).
16. Hadipour-Rokni, R., Askari Asli-Ardeh, E., Jahanbakhshi, A., Esmaili paeen-Afrakoti, I., Sabzi, S., 2023. Intelligent detection of citrus fruit pests using machine vision system and convolutional neural network through transfer learning technique. Comput. Biol. Med. 155, 106611. https://doi.org/10.1016/j.compbiomed.2023.106611
17. He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
18. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. https://doi.org/10.48550/arXiv.1704.04861
19. Ketabforoosh, K., 2023. Strawberry Images.
20. Kouloumprouka, Z., Monaghan, J., Bromley, J., Vickers, L., 2024. Opportunities and challenges for strawberry cultivation in urban food production systems. PLANTS PEOPLE PLANET.
21. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput. 1, 541–551. https://doi.org/10.1162/neco.1989.1.4.541
22. Lello, F., Dida, M., Mkiramweni, M., Matiko, J., Akol, R., Nsabagwa, M., Katumba, A., 2023. Fruit fly automatic detection and monitoring techniques: A review. Smart Agric. Technol. 5, 100294. https://doi.org/10.1016/j.atech.2023.100294
23. Meyer, T.K., Pearce, J.M., 2022. How Easy is it to Feed Everyone? Economic Alternatives to Eliminate Human Nutrition Deficits. Food Ethics 8, 3. https://doi.org/10.1007/s41055-022-00113-3
24. Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091
25. Oğuz, İ., Halil, İ., Nesibe Ebru, K., 2022. Strawberry Cultivation Techniques. IntechOpen.
26. Pérez-Borrero, I., Marín-Santos, D., Gegúndez-Arias, M.E., Cortés-Ancos, E., 2020. A fast and accurate deep learning method for strawberry instance segmentation. Comput. Electron. Agric. 178, 105736. https://doi.org/10.1016/j.compag.2020.105736
27. Puttemans, S., Vanbrabant, Y., Tits, L., Goedemé, T., 2016. Automated visual fruit detection for harvest estimation and robotic harvesting. https://doi.org/10.1109/IPTA.2016.7820996
28. PyTorch documentation — PyTorch 2.1 documentation [WWW Document], 2023. URL https://pytorch.org/docs/stable/index (accessed 12.6.23).
29. Rahamathunnisa, U., Nallakaruppan, M.K., Anith, A., Kumar KS, S., 2020. Vegetable Disease Detection Using K-Means Clustering And Svm. IEEE, pp. 1308–1311. https://doi.org/10.1109/ICACCS48705.2020.9074434
30. Rezaei-Dastjerdehei, M.R., Mijani, A., Fatemizadeh, E., 2020. Addressing Imbalance in Multi-Label Classification Using Weighted Cross Entropy Loss Function, in: 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME). Presented at the 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME), pp. 333–338. https://doi.org/10.1109/ICBME51989.2020.9319440
31. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2019. MobileNetV2: Inverted Residuals and Linear Bottlenecks. https://doi.org/10.48550/arXiv.1801.04381
32. Shorif Uddin, M., Bansal, J.C., 2021. Computer Vision and Machine Learning in Agriculture. Algorithms for Intelligent Systems. Springer Singapore.
33. StrawDI Dataset [WWW Document], 2020. URL https://strawdi.github.io/ (accessed 1.21.24).

34. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going Deeper with Convolutions. https://doi.org/10.48550/arXiv.1409.4842
35. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention Is All You Need. https://doi.org/10.48550/arXiv.1706.03762
36. Widmer, J., Christ, B., Grenz, J., Norgrove, L., 2024. Agrivoltaics, a promising new tool for electricity and food production: A systematic review. Renew. Sustain. Energy Rev. 192, 114277. https://doi.org/10.1016/j.rser.2023.114277
37. Woo, S., Uyeh, D.D., Kim, J., Kim, Y., Kang, S., Kim, K.C., Lee, S.Y., Ha, Y., Lee, W.S., 2020. Analyses of Work Efficiency of a Strawberry-Harvesting Robot in an Automated Greenhouse. Agronomy 10, 1751. https://doi.org/10.3390/agronomy10111751
38. Wydra, K., Vollmer, V., Busch, C., Prichta, S., Wydra, K., Vollmer, V., Busch, C., Prichta, S., 2023. Agrivoltaic: Solar Radiation for Clean Energy and Sustainable Agriculture with Positive Impact on Nature. IntechOpen. https://doi.org/10.5772/intechopen.111728
39. Zhou, C., Hu, J., Xu, Z., Yue, J., Ye, H., Yang, G., 2020. A Novel Greenhouse-Based System for the Detection and Plumpness Assessment of Strawberry Using an Improved Deep Learning Technique. Front. Plant Sci. 11.