horticulturae

MDPI

Article

A Novel Damage Inspection Method Using Fluorescence Imaging Combined with Machine Learning Algorithms Applied to Green Bell Pepper

Danial Fatchurrahman [1], Noelia Castillejo [1,2,*], Maulidia Hilaili [3], Lucia Russo [1], Ayoub Fathi-Najafabadi [1] and Anisur Rahman [4]

[1] Dipartimento di Scienze Agrarie, Alimenti, Risorse Naturali e Ingegneria (DAFNE), Università di Foggia, Via Napoli 25, 71122 Foggia, Italy; danial.fatchurrahman@unifg.it (D.F.); lucia.russo@unifg.it (L.R.); ayoub.fathinajafabadi@unifg.it (A.F.-N.)
[2] Departamento de Tecnología de Alimentos—FoodUPV, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain
[3] Laboratory of Bio-Sensing Engineering, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan; maulidiahilaili99@gmail.com
[4] Department of Farm Power and Machinery, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh; anis_fpm@bau.edu.bd
* Correspondence: ncasmon@upvnet.upv.es

Abstract: Fluorescence imaging has emerged as a powerful tool for detecting surface damage in fruits, yet its application to vegetables such as green bell peppers remains underexplored. This study investigates the fluorescent characteristics of minor mechanical damage, specifically 5 × 5 mm cuts in the exocarp of green bell peppers, which conventional digital imaging techniques fail to classify accurately. Chlorophyll fluorescence imaging was combined with machine learning algorithms—including logistic regression (LR), artificial neural networks (ANN), random forests (RF), k-nearest neighbors (kNN), and the support vector machine (SVM) to classify damaged and sound fruit. The machine learning models demonstrated a high classification accuracy, with calibration and prediction accuracies exceeding 0.86 and 0.96, respectively, across all algorithms. These results underscore the potential of fluorescence imaging as a non-invasive, rapid, and cheaper method for assessing mechanical damage in green bell peppers, offering valuable applications in quality control and postharvest management.

Keywords: *Capsicum annuum* L.; early detection; defective; sorting; fruit quality

Citation: Fatchurrahman, D.; Castillejo, N.; Hilaili, M.; Russo, L.; Fathi-Najafabadi, A.; Rahman, A. A Novel Damage Inspection Method Using Fluorescence Imaging Combined with Machine Learning Algorithms Applied to Green Bell Pepper. *Horticulturae* **2024**, *10*, 1336. https://doi.org/10.3390/horticulturae10121336

Received: 5 October 2024
Revised: 8 December 2024
Accepted: 11 December 2024
Published: 13 December 2024

1. Introduction

Green bell pepper (*Capsicum annuum* L.) has become the preferable vegetable to be consumed during the last few years due to its suspected links to cardiovascular disease prevention; atherosclerosis, cancer, and hemorrhage prevention; the slowing of the aging process; the avoidance of cholesterol; and the enhancement of physical resistance, as it contains a high amount of vitamins and antioxidants [1]. Bell peppers are considered non-climacteric vegetables, typically harvested either during the green edible stage (green bell pepper) or during the ripe color stage (red, yellow, or orange). In Japan, the increasing demand for this vegetable increased the country's production, reaching 148.500 tons in 2021 from only around 140.400 tons in 2015 [2]. Green bell peppers have a limited shelf life, making them perishable products. Their postharvest quality is primarily affected by several physiological factors, such as disease-related issues, moisture loss, and chilling injury [3].

Bell peppers are primarily sorted by hand and evaluated by a production supervisor to comply with the specified quality standards. This evaluation includes sensory techniques or methods grounded in the physicochemical and biochemical alterations in green bell peppers; consequently, the outcomes may be subjective, and the methods can be invasive [4].

Moreover, before or during the sorting, the fruit can be affected by impact vibration, compression, friction, and puncture forces, causing mechanical damage and a decline in quality [5]. In some cases, detecting mechanical damage on the surface of bell peppers is very difficult due to their small size, and their showing almost the same color as the surface of the fruit. Therefore, differentiating bell peppers according to physical damage would be paramount for preparing high-quality products for storage and packaging.

Nowadays, nondestructive techniques are gaining more attention because they can obtain information related to damage in a short amount of time and do not compromise the integrity of the fruits [6–8]. These nondestructive methods are highly effective and have shown promising results when applied to fruit quality evaluation. Over the past few years, researchers have been focusing on investigating quick and nondestructive methods for detecting damage to fruits. One of these techniques is fluorescence imaging, developed in recent years to characterize plant tissues, fruits, and vegetables [9]. Fluorescence imaging is a cutting-edge technology that combines fluorescence spectroscopy with imaging techniques. This technology can display spatial information regarding the distribution of fluorescence signals and can be implemented in either hyperspectral or multispectral modes. Excitation light sources, optical filters, cameras, and computer equipment are the primary components that make up this system [6].

Examples of such fluorescence imaging systems include peel defect detection in citrus [10,11] and assessment of the quality of persimmon calyxes [12]. After being excited by a light beam, chlorophylls, the most abundant fluorophores in fruit tissues, produce chlorophyll fluorescence in the 650 to 850 nm [13]. As a result, chlorophyll fluorescence is a potential tool for identifying damage to fruits, given that damage to fruits is typically accompanied by a change in the amount of chlorophyll present.

Numerous studies have shown that chlorophylls, carotenoids, and anthocyanins are the three primary pigments responsible for vegetable and fruit color [14]. The concentration of these substances varies across different parts of the fruit and may serve as a potential probe for quality assessment [15]. Chlorophylls are among the primary pigments that have been associated with green pepper. Higher plants typically have a ratio of approximately 3:1 between chlorophyll *a* and the minor components chlorophyll *b*. According to [16], both types of chlorophyll exhibit typical fluorescence spectra when they are isolated pigments in an organic solution. These spectra have a maximum peak emission in the red region (690 nm) and a shoulder in the far-red region (730–740 nm).

Artificial intelligence-based modeling techniques have recently been extensively utilized in nondestructive testing. New methods, such as machine vision technology, have been used in various industrial processes across multiple fields to evaluate product quality. Image processing and analysis procedures are applied to the images acquired to accomplish this goal [17]. Techniques that include machine learning are utilized most frequently for image analysis and recognition of crops [18]. The use of machine learning for grading is based on knowledge of data mining and computational intelligence techniques. Classification and clustering are the two methods of grading that are used everywhere. Clustering is an unsupervised learning method, whereas classification is a supervised learning method. During the classification method, the models acquire knowledge from a sizable data set that has already been categorized, referred to as the training set. After that, the models that have been developed are utilized to classify new items and data. Numerous classification models have been developed, including logistic regression (LR) [19], random forest (RF), artificial neural network (ANN) [20], support vector machine (SVM) [21], and k-nearest neighbors (kNN) [22], which are extensively utilized in the nondestructive evaluation of food materials.

Until now, there has been no literature about using fluorescence images to detect damage in green bell peppers and applying machine learning algorithms to evaluate the scale of mechanical damage. Thus, the first objective of this research was to identify in vitro chlorophyll fluorescence components from the exocarp of both damaged and sound green

bell peppers. Then, we classified and differentiated the damaged fruit from the sound ones by applying different machine learning models.

2. Materials and Methods

2.1. Sample of Green Bell Peppers

Green bell peppers (*Capsicum annuum* L.) of the *Kyohikari* variety were obtained from a farmer association in Miyazaki Prefecture, Japan. The peppers were harvested from five different orchards on 16 January 2016, with approximately 120 fruits collected from each location. The fruits were transported to the laboratory under cold storage at 7 °C on the same day. The following day, after being stored in these conditions, 5 experts were tasked to select only sound fruit without defects, followed by machine vision technology to assess and finally select only 450 high-quality fruits with a uniform color, with around 90 fruits from each orchard. Details regarding the selection based on the color process are described in the image acquisition system. The 450 green bell peppers were then divided into two groups of 225 each. In one of the groups, a small incision measuring 5 × 5 mm was made manually with a scalpel blade n°10 (Merck, Darmstadt, Germany), removing a thin layer of the fruit's skin. This cut is challenging to detect visually, as the affected area closely resembles the undamaged part in color. Small cuts are difficult to detect because the color and texture of the damaged part are very similar to those of the healthy surface. This makes it challenging for standard image-based techniques to detect enough visual differences to distinguish the cuts clearly. Standard images captured by digital cameras, which use the visible red, green, and blue (RGB) spectrum, cannot capture the subtle changes that occur at the chemical or physiological level, such as pigment variations (e.g., chlorophyll) in the damaged areas. Mechanical damage may alter the internal structure of the tissue without producing visible changes that are easily detectable by the human eye or RGB sensors. These challenges explain why more advanced techniques, such as fluorescence imaging, are necessary. Fluorescence imaging can detect invisible changes by highlighting fluorescent signals generated by compounds like chlorophyll in the damaged areas. The selection of a 5 × 5 mm damage size for inspection in this study is motivated by its representativeness and the inherent challenges of visually detecting small damages. Mechanical damages, such as cuts and abrasions measuring 5 × 5 mm, frequently occur while handling and transporting fruits and vegetables. Evaluating the effectiveness of the inspection system on this scale of damage not only simulates real-world conditions but also assesses its sensitivity in identifying imperfections that may be imperceptible to the naked eye or standard inspection methods. On the other hand, smaller or subtler damage, like the 5 × 5 mm cuts in this study, might be more challenging to identify, which would test the limits of the sensitivity of the detection system. The positions and shapes of the damage during the sample production process are not always the same; this is to mimic real mechanical damage, which can vary in shape, position, and size, and detection systems must be flexible to handle these variations.

2.2. Image Acquisition System

An image acquisition setup was designed, as shown in Figure 1, featuring a DSLR camera (Canon E.O.S. Kiss X2, Tokyo, Japan) equipped with an EF-S 18-55 mm, F3.5-5.6 IS lens. The camera settings included an ISO of 100, a shutter speed of 1/13 s, an aperture (F-number) of 4, and a focal length of 50 mm. A 58 mm wideband circular polarizing (C-PL) filter was attached to the lens to reduce glare and reflections. The setup also utilized four halogen lamps (12V50W, KLS, Miyagi-ken, Japan) fitted with near-infrared (NIR), infrared (IR), and 62 mm wideband C-PL filters. Halogen lamps were selected due to their broad-spectrum output, stability, and high color rendering index (CRI), providing superior color accuracy compared to LED lights. NIR and IR filters (Edmund optics, York, UK) were used to manage heat, and each lamp was equipped with cooling fans. Images were captured with a light intensity of 8700 lux, a color temperature of 5000 K, and a resolution of 4272 × 2417 pixels at 72 dpi.

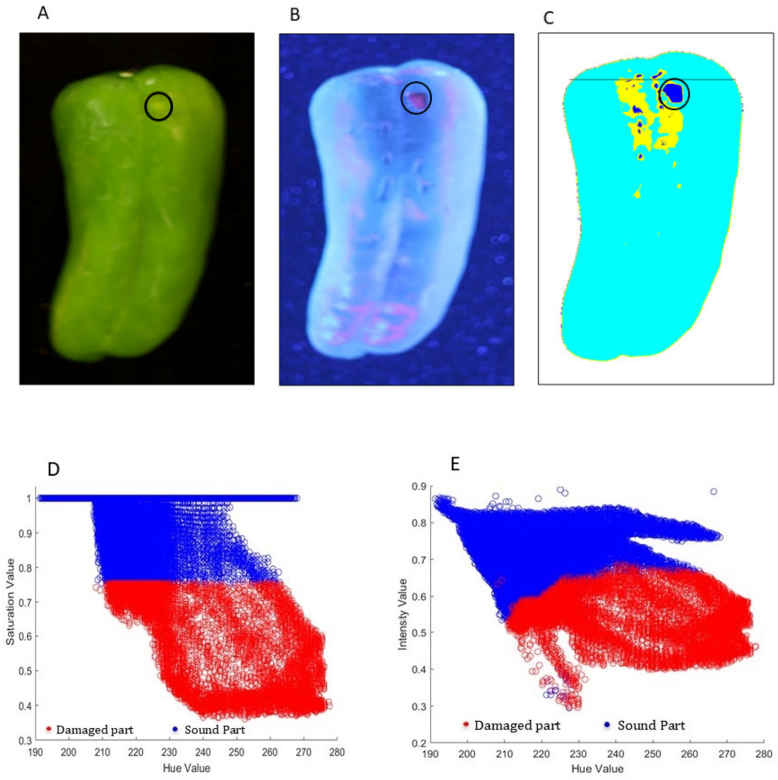

Figure 1. The appearance of white digital image (**A**), fluorescence image (**B**), pixel-based k-mean clustering image (**C**), pixel-based scatter plot of hue° vs. saturation (**D**), and pixel-based hue° vs. intensity (**E**). The damaged part was indicated with a black circle.

Images of green bell peppers were sorted according to their G channel values in the RGB image, ranging from 65 to 76. For investigating the chlorophyll fluorescence response in the damaged areas of the peppers, a UV light source (Philips, Osram PHILIPS Black Light UV Lamp 8 W, Hamburg, Germany) emitting at an excitation wavelength of 365 nm was employed. In the standard white light digital image (Figure 1A), the damaged regions were not easily identifiable through color (B, G, R, L, a*, b*) or texture features, such as contrast, dissimilarity, homogeneity, energy, correlation, and angular second moment (ASM) obtained from the gray-level co-occurrence matrix (GLCM) (Table 1). However, the fluorescence image (Figure 1B) displayed distinct characteristics with red chlorophyll emission, which differentiated the damaged areas from the healthy tissue. A K-means clustering algorithm was then applied to the pixels of the fluorescence image, resulting in a clustered representation (Figure 1C). Additionally, the pixels representing sound and damaged tissue were plotted according to Hue° versus saturation (Figure 1D) and Hue° versus intensity (Figure 1E), illustrating the efficacy of fluorescence imaging in distinguishing damaged tissue.

Table 1. Color and texture feature of green bell pepper (n = 60), the significance difference was assessed by using one-way ANOVA $p < 0.001$.

Classes	B	G	R	L	a*	b*	Con.	Diss.	Hom.	Ene.	Cor.	A.S.M.
Damaged	5.3 [a]	71.2 [a]	38 [a]	67.5 [a]	104.3 [a]	159.4 [a]	4.1 [a]	1.1 [a]	0.6 [a]	0.2 [a]	1.0 [a]	0.1 [a]
Sound	5.4 [a]	70.1 [a]	36.4 [a]	65.8 [a]	105.2 [a]	157.8 [a]	4.3 [a]	1.2 [a]	0.6 [a]	0.2 [a]	1.0 [a]	0.1 [a]

Different lowercase letters indicate significant differences between damaged and sound fruit.

Thus, this study compares standard digital and fluorescence images for classifying damaged and undamaged bell peppers using various machine learning algorithms, which will be detailed further in the manuscript.

2.3. Assessment of Fluorescence Excitation and Emission Matrix (EEM) of Fruit

The fluorescence spectra were obtained using a Jasco Spectrofluorometer (FP-8300, Tokyo, Japan) from circle-shaped Ø1.5 cm intact pericarp sections (sound samples) and pericarp sections with the exocarp removed (damaged samples). The representative of five replicates was used for each sample type from 5 random green bell peppers. The device was set with a response time of 10 ms and a scan speed of 5000 nm/min. Spectra were recorded over a 200–750 nm wavelength, utilizing a 5 nm bandwidth for excitation and emission. The excitation wavelengths were scanned at intervals of 10 nm, while emission wavelengths were measured at 1 nm intervals. The raw excitation–emission matrix (EEM) data were collected in arbitrary units (AU), allowing for a comparative analysis of the fluorescence spectra between the intact and damaged regions. In this study, we didn't quantify the concentration of fluorescence compounds, allowing sufficient measurement using AU. Therefore, for quantitative fluorescence measurement, Raman unit (RU) is advisable [23]. In this study, the correction factor for conversion was 0.021.

2.4. Machine Learning for Classification

The proposed system's methodology involves using and comparing five machine learning algorithms: LR, ANN, RF, kNN, and SVM [24]. These algorithms were employed to detect and classify the freshness of green bell peppers based on their features throughout storage. The machine learning algorithms were run under Orange Data Mining software version 3.37.0 on the Python platform, utilizing an approach that focuses on the extraction of image features from green paper at the pixel level.

As shown in Figure 2, the dataset, consisting of 900 images of green bell peppers, was obtained from the fluorescence image acquisition system only due to unreliable accuracy of below 0.4 by using white digital images. A data augmentation method using mirroring was applied to increase the number of samples in the original dataset, improving the model's generalization capability and avoiding overfitting [25]. Each image was mirrored on the vertical axis, obtaining 900 images as the total number of samples, including the original images. We utilized a pixel-based approach for our analysis. Specifically, we collected a total of 900 data points of fruits. These points were hand-selected or labeled, with 450 corresponding to pixels where incisions were made on the peppers (damage regions) and 450 randomly selected from healthy peppers (non-damage regions). This balanced selection aimed to ensure that our model was trained on both damaged and undamaged pixels, allowing it to effectively learn the differences between these regions. To assess the robustness of our model across different orchards, we employed a leave-one-population-out cross-validation (LOPOCV) approach. In this method, each orchard data is treated as an independent 'population'. Our dataset comprises 900 fruit samples, with equal samples from each of the five orchards.

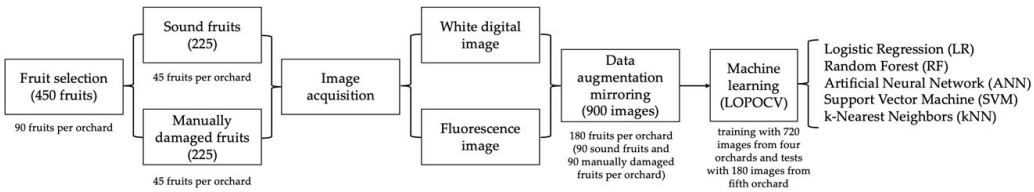

Figure 2. Flowchart of the classification model of sound and damaged green bell pepper using machine learning algorithms.

For each fold in the LOPOCV process, the model is trained on the data from four orchards (720 samples) and tested on the fifth orchard (180 samples) left out of the training process. This cycle is repeated five times, with each orchard serving once as the test set, ensuring that every population is systematically used for both training and testing.

This approach allows us to evaluate how well the model trained on the data from certain orchards performs on an unseen orchard. By cycling through each population as the test set, LOPOCV provides insight into the model's robustness and its ability to generalize

across different growing conditions, fruit characteristics, or orchard management practices that may vary from one orchard to another.

LOPOCV strengthens the validation of our model, helping to ensure that it is not overly reliant on specific population characteristics and has greater potential for application across diverse orchards. The overall procedure is depicted in Figure 2.

2.4.1. Logistic Regression (LR)

LR is a mathematical technique that models the relationship between a categorical dependent variable and one or more independent variables. It is most often applied in binary classification scenarios [19]. Due to its simplicity, ease of interpretation, and effectiveness in handling binary outcome predictions, this model is commonly used. For LR, we focused on regularization strength by adjusting the C parameter (inverse of regularization strength) over a logarithmic scale. After testing a range from 0.01 to 10, we found that C = 1.0 with an L2 penalty provided the best performance, effectively balancing regularization and accuracy.

2.4.2. Random Forest (RF)

As explained by [20], RF is a versatile and powerful ensemble learning method frequently applied to classification and regression problems in machine learning. It works by creating multiple decision trees during the training process and combining their results by using the mode for classification tasks or the average prediction for regression tasks [24]. In the RF model, we used max_features = sqrt(N), where N is the total number of features in the dataset. Specifically, sqrt selects the square root of the number of input features as the maximum number of features considered for each split within the forest. This parameter optimizes performance by balancing tree diversity and model interpretability. We tested the number of trees (n_estimators) between 50 and 300 and the maximum tree depth from 5 to 20 to avoid overfitting. The best performing model used 200 trees with a maximum depth of 15. We also examined the minimum samples required to split a node (1 to 5) and minimum samples per leaf (1 to 5). We found that minimum samples split = 2 and minimum samples leaf = 1 yielded the optimal results. The max features parameter was set to "sqrt", which improved model stability and interpretability.

2.4.3. Artificial Neural Network (ANN)

ANNs are machine learning models inspired by the structure and function of biological neural networks found in animal brains [20]. These networks are composed of interconnected neurons or units arranged into layers: an input layer, one or more hidden layers, and an output layer. Each neuron in a layer receives input from the neurons of the previous layer, processes this information using an activation function, and passes the result to neurons in the next layer. For the ANN, we varied the number of hidden layers between 1 and 3, with neurons per layer ranging from 16 to 128. The best configuration included two hidden layers, with 64 neurons in the first layer and 32 in the second layer, providing an effective balance of complexity and generalization. The learning rate was set to 0.001, and we selected ReLU as the activation function for the hidden layers, with softmax for the output layer. The Adam optimizer was chosen for its efficient convergence. For the ANN model, we utilized 50 epochs and a batch size of 32 during training. These values balance training time and model accuracy, ensuring effective learning without excessive computation or overfitting.

2.4.4. Support Vector Machine (SVM)

SVM is a highly effective supervised learning algorithm for classification and regression tasks [25]. SVM operates by finding the optimal hyperplane that best separates the data into different classes to maximize the margin between the hyperplane and the nearest data points from each class, referred to as support vectors. This robustness makes it popular in image classification. In the SVM model, we tested the regularization parameter (C) over

a range from 0.1 to 10 and gamma values from 0.001 to 1. The optimal settings were found with C = 1.0 and gamma = 0.01 using the RBF kernel, which provided the best classification accuracy by maximizing the margin. The regularization parameter C for SVM was determined using a linear search with a step size of 0.1, iterating over a range from 0.1 to 10. This method allowed for a thorough search across potential values, optimizing the classification performance.

2.4.5. k-Nearest Neighbors (kNN)

The kNN algorithm is a straightforward, yet powerful, supervised learning technique for classification and regression tasks [22]. In classification, the algorithm assigns the class of the query point based on a majority vote among its k closest neighbors. The simplicity and ease of implementation of kNN made it an intuitive model. For kNN, we varied k (number of neighbors) from 3 to 15 and tested both Euclidean and Manhattan distance metrics. The optimal configuration was found with k = 7 using Euclidean distance, balancing simplicity and accuracy.

3. Results and Discussion

3.1. Fluorescence EEM of Sound and Damaged Skin

Fluorescence EEM revealed significant differences between sound and damaged green bell peppers, as illustrated in Figure 3. The analysis identified three areas of interest in the EEM, with two primary peaks observed at Ex/Em 380/689 nm and Ex/Em 280/310 nm. The third area was attributed to scattering and considered noise, showing no discernible differences between the studied samples. The two prominent EEM peaks exhibited significant intensity variations between sound and damaged samples, especially Ex/Em 380/689 depicted in Figure 4, providing valuable insights into the condition of the green bell pepper. These peaks are associated with fluorescent substances, indicating that damage to the fruit triggers an increase in the concentration of these substances. The most abundant fluorophores in fruit and vegetables are amino acids, nucleic acids, chlorophylls, and polyphenols [26]. According to [27], the peak at Ex/Em 380/689 nm may be associated with chlorophylls and the peak at Ex/Em 280/310 nm with flavonoids.

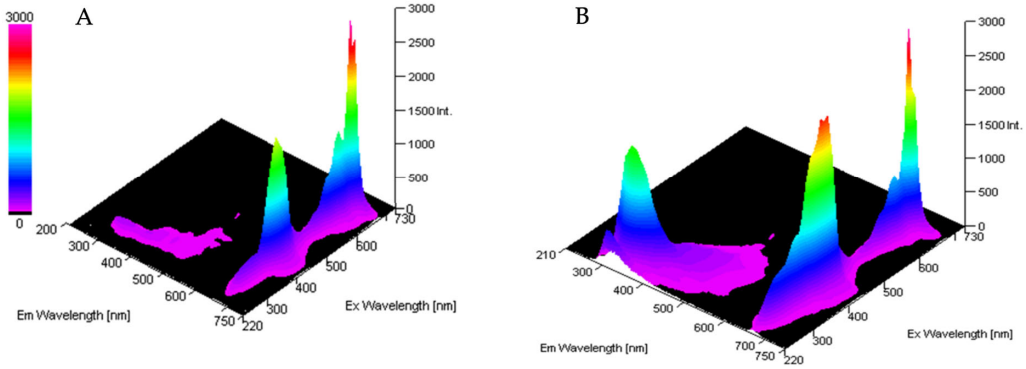

Figure 3. Excitation and emission matrix of sound (**A**) and damaged (**B**) green bell pepper. Scale bar from 0 to 3000 corresponding to fluorescence intensity is expressed in arbitrary units, AU.

Abiotic stresses have been shown to stimulate the production of various phytochemicals in plants. Ref. [28] demonstrated that even low doses of UV radiation can induce invisible damage to plant tissues, triggering a metabolic response that leads to increased synthesis of carotenoids and phenolic compounds.

Similarly, mechanical wounding has been found to activate the phenylpropanoid pathway, primarily through the upregulation of phenylalanine ammonia-lyase (PAL) enzyme activity [29]. PAL is a critical enzyme in the biosynthesis of phenolic compounds, and its increased activity results in the accumulation of various phenolics as part of the plant's wound healing and defense response [30]. This wound-induced phenolic production has

been observed in numerous fruits and vegetables, including carrots, potatoes, and lettuce, demonstrating the widespread nature of this phenomenon in plant tissues [31].

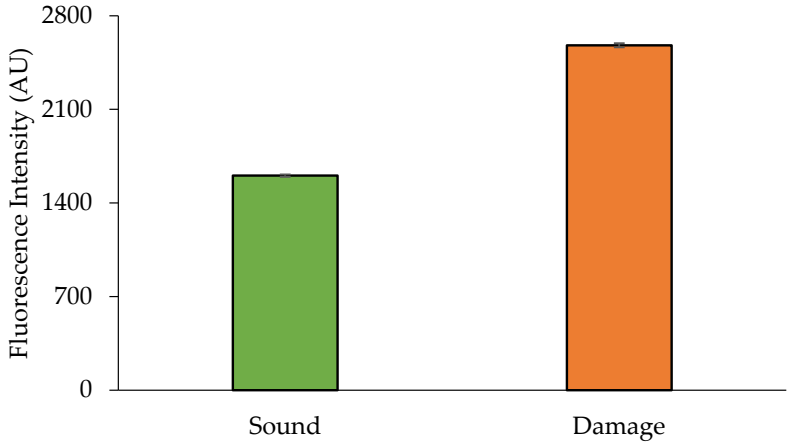

Figure 4. Bar plot showing the mean fluorescence intensity (arbitrary units, AU) at Ex/Em 380/689 nm for sound and damaged green bell pepper samples. Error bars represent the standard deviation (SD) of fluorescence intensity measurements, with n = 5 biological replicates per sample type.

3.2. Machine Learning Models in Classification

The classification performance of machine learning algorithms in distinguishing between damaged and sound green bell peppers using white digital images was unreliable, with an overall accuracy of less than 0.4 during both the calibration and prediction phases. Consequently, this approach will not be further discussed in this study. In contrast, fluorescence imaging combined with machine learning markedly improved in accurately classifying the peppers into two categories: damaged and sound. Table 2 summarizes five different algorithms' overall performance in classifying green bell peppers' freshness during the calibration phase.

Table 2. The overall accuracy during the calibration phase across the machine learning models for classifying green bell peppers into two categories (damaged and sound).

Model	AUC [1]	CA [2]	F1 [3]	Prec [4]	Recall [5]	MCC [6]
Logistic regression	0.91	0.87	0.87	0.88	0.87	0.83
Artificial neural network	0.92	0.88	0.88	0.88	0.88	0.83
Random forest	0.88	0.87	0.87	0.87	0.87	0.81
k-nearest neighbors	0.88	0.86	0.86	0.86	0.86	0.80
Support vector machine	0.88	0.87	0.87	0.87	0.87	0.81

[1] Area under the ROC (receiver operating characteristic) curve (AUC), [2] classification accuracy (CA), [3] the measure of the harmonic means of precision and recall (F1), [4] precision (Prec), [5] the times the model correctly identifies positive instances (Recall), [6] Matthews correlation coefficient (MCC).

For assessing the effectiveness of each model, various performance metrics were employed, including the area under the receiver operating characteristic curve (AUC), classification accuracy (CA), the F1 score (the harmonic mean of precision and recall), precision (Prec), recall (the proportion of correctly identified positive instances), and the Matthews correlation coefficient (MCC). As shown in the tables, all algorithms achieved a high overall CA exceeding 0.86, successfully differentiating between damaged and sound peppers.

For prediction evaluation, the performance of each model was analyzed to determine how well they predicted the quality of green bell peppers, categorized into two groups: damaged and sound. The overall performance metrics for each model are detailed in Table 3. LR, ANN, SVM, and RF achieved reliable calibration accuracy (CA = ~0.91). In

contrast, the kNN model showed a slightly lower accuracy of 0.9. The confusion matrices for these models are provided in Table 4.

Table 3. The overall accuracy in prediction across the machine learning models for classifying green bell peppers into the categories of damaged and sound.

Model	AUC [1]	CA [2]	F1 [3]	Prec [4]	Recall [5]	MCC [6]
Logistic regression	0.95	0.91	0.91	0.92	0.91	0.86
Artificial neural network	0.96	0.92	0.92	0.92	0.92	0.86
Random forest	0.92	0.91	0.91	0.91	0.91	0.84
k-nearest neighbors	0.92	0.90	0.90	0.90	0.90	0.83
Support vector machine	0.92	0.91	0.91	0.91	0.91	0.84

[1] Area under the ROC (receiver operating characteristic) curve (AUC), [2] classification accuracy (CA), [3] the measure of the harmonic means of precision and recall (F1), [4] precision (Prec), [5] the times the model correctly identifies positive instances (Recall), [6] Matthews correlation coefficient (MCC).

Table 4. The confusion matrix in prediction across the machine learning models to classify damaged and sound green bell peppers.

Logistic Regression	**Predicted Damaged**	**Predicted Sound**	**Total Actual**	**Accuracy (%)**
Actual damaged	TP = 82	FN = 8	90	91.11
Actual sound	FP = 8	TN = 82	90	91.11
Total predicted	90	90	180	91
Artificial Neural Network	**Predicted Damaged**	**Predicted Sound**	**Total Actual**	**Accuracy (%)**
Actual damaged	TP = 83	FN = 7	90	92.22
Actual sound	FP = 7	TN = 83	90	92.22
Total predicted	90	90	180	92
Random Forest	**Predicted Damaged**	**Predicted Sound**	**Total Actual**	**Accuracy (%)**
Actual damaged	TP = 82	FN = 8	90	91.11
Actual sound	FP = 8	TN = 82	90	91.11
Total predicted	90	90	180	91
k-Nearest Neighbors	**Predicted Damaged**	**Predicted Sound**	**Total Actual**	**Accuracy (%)**
Actual damaged	TP = 81	FN = 9	90	90
Actual sound	FP = 9	TN = 81	90	90
Total predicted	90	90	180	90
Support Vector Machine	**Predicted Damaged**	**Predicted Sound**	**Total Actual**	**Accuracy (%)**
Actual damaged	TP = 82	FN = 8	90	91.11
Actual sound	FP = 8	TN = 82	90	91.11
Total predicted	90	90	180	91

Compared to other studies on the classification of green bell peppers [32], the RF algorithm outperformed LR when classifying bell peppers using RGB images. LR has also been shown to be an effective mathematical tool for estimating marketability probabilities related to the quality deterioration of various pepper cultivars [33]. Additionally, a study by [34] reported that the combination of RGB imaging and ANN achieved perfect accuracy in determining the maturity of bell peppers. Ropelewska et al. [35] highlighted that the RF model exhibited outstanding performance, attaining an average discrimination accuracy of 99% for red bell peppers. Moreover, applying the RF algorithm has further enhanced classification accuracy for bell peppers [32].

Studies of other fruits [36] demonstrated that LR is a dependable approach for evaluating the marketability of tomatoes based on their microbial and chemical attributes across various agricultural practices. Earlier research has effectively utilized LR to examine the factors contributing to fruit disintegration in pears [37] and to analyze the influence of temperature on avocados' shelf life and quality [38]. Furthermore, LR models have been successfully applied to assess bruise formation in different tomato cultivars and to investigate how storage temperature affects the quality of produce [39].

ANN models have demonstrated significant potential in improving and assessing the quality of loquat fruit [40]. By utilizing color features, ANNs have been able to accurately estimate quality indices and accurately predict the freshness of bananas [41]. Bhargava and Bansal [42] found that their initial training dataset correctly classified 94.65% and 96.23% of images for golden delicious apples. However, in contrast to our results, SVM outperformed the kNN model in fruit classification, with SVM achieving an accuracy of 98.49% and kNN reaching 98.50% for plums [43]. Additionally, in a study evaluating the sensory quality of oranges as a measure of freshness, LR provided the most accurate predictions compared to kNN, SVM, and ANN models [26].

The models used in this study successfully differentiated between damaged and sound green peppers using fluorescence images. The adoption of digital cameras for evaluating agricultural products has gained traction due to their affordability and portability, especially compared to NIR spectrometers and other optical systems such as hyperspectral cameras. This research indicates that a digital camera can be an effective, non-invasive, rapid, and cost-efficient tool for evaluating the marketability and quality of green bell peppers. The results emphasize the potential of using fluorescence techniques in digital cameras for fruit quality prediction, showcasing how computer vision techniques and sophisticated learning models can automatically extract crucial information from images. This approach minimizes the necessity for manual intervention and improves accuracy.

This study has shown that fluorescence imaging combined with machine learning algorithms can successfully classify mechanical damage in green bell peppers. However, certain areas warrant further exploration and refinement to improve the methodology's robustness and adaptability. Future research could focus on expanding the scope of this technique to various agricultural products beyond green bell peppers. Different fruits and vegetables may present unique fluorescence characteristics, which could challenge and further validate the model's generalization capabilities. Additionally, the use of many other imaging methods, such as hyperspectral or multispectral imaging in tandem with fluorescence imaging, could enhance the detection of subtle, non-visible surface damage, potentially leading to higher classification accuracy. Testing alternative machine learning techniques, such as deep learning architectures or ensemble methods that incorporate additional environmental and contextual data (e.g., lighting variations in different growth stages), may further improve model robustness. Lastly, the 5×5 mm damage size may not represent all real-world damage types and severities, suggesting that the model's applicability could be limited in certain practical settings.

4. Conclusions

In this study, we showcased the effectiveness of machine learning algorithms in distinguishing between damaged and sound green bell peppers. The classification models based on logistic regression, k-nearest neighbors, random forest, and support vector machine reached an accuracy of >86% in calibration. They improved the performance in prediction, reaching an accuracy of >90%. These machine learning models serve as valuable tools for classification tasks within the agricultural industry. However, it is crucial to acknowledge that this research is grounded in a single harvest population from five different orchards; incorporating data from multiple harvests in different years will strengthen the model's robustness and generalizability.

Author Contributions: Conceptualization, D.F.; methodology, D.F., N.C., A.F.-N., M.H., A.R., L.R.; software, D.F., N.C., A.F.-N., M.H., A.R., L.R.; validation, D.F., N.C., A.F.-N., M.H., A.R., L.R.; formal analysis, D.F., N.C., A.F.-N., M.H., A.R., L.R.; investigation, D.F., M.H.; resources, D.F., N.C., A.F.-N., M.H., A.R., L.R.; data curation, D.F., N.C., A.F.-N., M.H., A.R., L.R.; writing—original draft preparation, D.F., N.C., A.F.-N., M.H., A.R., L.R.; writing—review and editing, D.F., N.C., A.F.-N., A.R.; visualization, D.F., N.C., A.F.-N., M.H., A.R., L.R.; supervision, D.F., A.R.; project administration, M.H., D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data in this study are confidential and available on request.

Acknowledgments: Noelia Castillejo is grateful for her contract, which has been cofinanced by Juan de la Cierva (JDC2022-049432-I) from MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR".

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

References

1. Tiamiyu, Q.O.; Adebayo, S.E.; Ibrahim, N. Recent advances on postharvest technologies of bell pepper: A review. *Heliyon* **2023**, *9*, e15302. [CrossRef] [PubMed]
2. The 97th Statistical Yearbook of Ministry of Agriculture, Forestry and Fisheries. Available online: https://www.maff.go.jp/e/data/stat/97th/index.html (accessed on 1 September 2024).
3. Hedayati, S.; Niakousari, M. Effect of Coatings of Silver Nanoparticles and Gum Arabic on Physicochemical and Microbial Properties of Green Bell Pepper (*Capsicum annuum*). *J. Food Process. Preserv.* **2015**, *39*, 2001–2007. [CrossRef]
4. Da, I.; Vieira, C.; Fatibello-Filho, O. Flow injection spectrophotometric determination of hydrogen peroxide using a crude extract of zucchini (*Cucurbita pepo*) as a source of peroxidase. *Analyst* **1998**, *123*, 1809–1812.
5. Lin, M.; Fawole, O.A.; Saeys, W.; Wu, D.; Wang, J.; Opara, U.L.; Nicolai, B.; Chen, K. Mechanical damages and packaging methods along the fresh fruit supply chain: A review. *Crit. Rev. Food Sci. Nutr.* **2023**, *63*, 10283–10302. [CrossRef] [PubMed]
6. Mei, M.; Li, J. An overview on optical nondestructive detection of bruises in fruit: Technology, method, application, challenge and trend. *Comput. Electron. Agric.* **2023**, *213*, 108195. [CrossRef]
7. Ali, M.M.; Hashim, N.; Bejo, S.K.; Jahari, M.; Shahabudin, N.A. Innovative nondestructive technologies for quality monitoring of pineapples: Recent advances and applications. *Trends Food Sci. Technol.* **2023**, *133*, 176–188.
8. Salvucci, G.; Pallottino, F.; De Laurentiis, L.; Del Frate, F.; Manganiello, R.; Tocci, F.; Vasta, S.; Figorilli, S.; Bassotti, B.; Violino, S.; et al. Fast olive quality assessment through RGB images and advanced convolutional neural network modeling. *Eur. Food Res. Technol.* **2022**, *248*, 1395–1405. [CrossRef]
9. Du, Z.; Zeng, X.; Li, X.; Ding, X.; Cao, J.; Jiang, W. Recent advances in imaging techniques for bruise detection in fruits and vegetables. *Trends Food Sci. Technol.* **2020**, *99*, 133–141. [CrossRef]
10. Momin, A.; Kondo, N.; Al Riza, D.F.; Ogawa, Y.; Obenland, D. A Methodological Review of Fluorescence Imaging for Quality Assessment of Agricultural Products. *Agriculture* **2023**, *13*, 1433. [CrossRef]
11. Momin, M.A.; Kondo, N.; Ogawa, Y.; Ido, K.; Ninomiya, K. Patterns of fluorescence associated with citrus peel defects. *Eng. Agric. Environ. Food* **2013**, *6*, 54–60. [CrossRef]
12. Fathi-Najafabadi, A.; Besada, C.; Gil, R.; Calatayud, M.A.; Salvador, A. Chlorophyll fluorescence imaging as a tool to evaluate calyx senescence during the ripening of persimmon fruit treated with gibberellic acid. *Postharvest Biol. Technol.* **2021**, *179*, 111582. [CrossRef]
13. Lu, Y.; Lu, R. Nondestructive defect detection of apples by spectroscopic and imaging technologies: A Review. *Trans. ASABE* **2017**, *60*, 1765–1790. [CrossRef]
14. Di Gioia, F.; Tzortzakis, N.; Rouphael, Y.; Kyriacou, M.C.; Sampaio, S.L.; Ferreira, I.C.F.R.; Petropoulos, S.A. Grown to be blue—Antioxidant properties and health effects of colored vegetables. Part II: Leafy, fruit, and other vegetables. *Antioxidants* **2020**, *9*, 97. [CrossRef] [PubMed]
15. Chávez-Mendoza, C.; Sanchez, E.; Muñoz-Marquez, E.; Sida-Arreola, J.P.; Flores-Cordova, M.A. Bioactive compounds and antioxidant activity in different grafted varieties of bell pepper. *Antioxidants* **2015**, *4*, 427–446. [CrossRef]
16. Sun, D.-W. *Hyperspectral Imaging for Food Quality Analysis and Control*; Elsevier: Amsterdam, The Netherlands, 2010.
17. Mohi-Alden, K.; Omid, M.; Firouz, M.S.; Nasiri, A. A machine vision-intelligent modelling based technique for in-line bell pepper sorting. *Inf. Process. Agric.* **2023**, *10*, 491–503. [CrossRef]
18. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [CrossRef]
19. Hilbe, J.M. *Logistic Regression Models*, 1st ed.; Chapman and Hall/CRC: New York, NY, USA, 2009.
20. Hastie, T.; Tibshirani, R.; Friedman, J. Model Assessment and Selection. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2008.
21. Moomkesh, S.; Mireei, S.A.; Sadeghi, M.; Nazeri, M. Early detection of freezing damage in sweet lemons using Vis/SWNIR spectroscopy. *Biosyst. Eng.* **2017**, *164*, 157–170. [CrossRef]
22. Michael Steinbach, P.-N.T. kNN: k-nearest neighbors. In *The Top Ten Algorithms in Data Mining*; Chapman and Hall/CRC: New York, NY, USA, 2009; pp. 165–176.
23. Fatchurrahman, D.; Castillejo, N.; Hilaili, M.; Nurwahyuningsih; Russo, L.; Kondo, N. Analysis of fluorescence changes in different sections of green bell pepper (*Capsicum annuum* L.) over storage periods. *Postharvest Biol. Technol.* **2024**, *217*, 113094. [CrossRef]
24. Jokanović, V.R. *Computer Vision and Internet of Things*; Chapman and Hall/CRC: New York, NY, USA, 2022. [CrossRef]
25. Wu, D.; Jiang, S.; Zhao, E.; Liu, Y.; Zhu, H.; Wang, W.; Wang, R. Detection of Camellia oleifera Fruit in Complex Scenes by Using YOLOv7 and Data Augmentation. *Appl. Sci.* **2022**, *12*, 11318. [CrossRef]

26. Al-Sammarraie, M.A.J.; Gierz, Ł.; Przybył, K.; Koszela, K.; Szychta, M.; Brzykcy, J.; Baranowska, H.M. Predicting Fruit's Sweetness Using Artificial Intelligence—Case Study: Orange. *Appl. Sci.* **2022**, *12*, 8233. [CrossRef]
27. Fatchurrahman, D.; Kuramoto, M.; Al Riza, D.F.; Ogawa, Y.; Suzuki, T.; Kondo, N. Fluorescence time series monitoring of different parts of green pepper (*Capsicum annuum* L.) under different storage temperatures. *Comput. Electron. Agric.* **2020**, *179*, 105850. [CrossRef]
28. Castillejo, N.; Martínez-Zamora, L.; Artés-Hernández, F. Postharvest UV radiation enhanced biosynthesis of flavonoids and carotenes in bell peppers. *Postharvest Biol. Technol.* **2022**, *184*, 111774. [CrossRef]
29. Hu, W.; Sarengaowa, W.; Guan, Y.; Feng, K. Biosynthesis of Phenolic Compounds and Antioxidant Activity in Fresh-Cut Fruits and Vegetables. *Front. Microbiol.* **2022**, *13*, 906069. [CrossRef] [PubMed]
30. Zapata, R.; Martínez-Zamora, L.; Cano-Lamadrid, M.; Artés-Hernández, F. Wounding Citrus Peel By-Products as Abiotic Stress to Induce the Synthesis of Phenolic Compounds? *Horticulturae* **2024**, *10*, 885. [CrossRef]
31. Reyes, L.F.; Villarreal, J.E.; Cisneros-Zevallos, L. The increase in antioxidant capacity after wounding depends on the type of fruit or vegetable tissue. *Food Chem.* **2007**, *101*, 1254–1262. [CrossRef]
32. Harel, B.; Parmet, Y.; Edan, Y. Maturity classification of sweet peppers using image datasets acquired in different times. *Comput. Ind.* **2020**, *121*, 103274. [CrossRef]
33. Díaz-Pérez, M.; Carreño-Ortega, Á.; Salinas-Andújar, J.A.; Callejón-Ferre, Á.J. Logistic regression to evaluate the marketability of pepper cultivars. *Agronomy* **2019**, *9*, 125. [CrossRef]
34. Villaseñor-Aguilar, M.J.; Bravo-Sánchez, M.G.; Padilla-Medina, J.A.; Vázquez-Vera, J.L.; Guevara-González, R.G.; García-Rodríguez, F.J.; Barranco-Gutiérrez, A.I. A maturity estimation of bell pepper (*Capsicum annuum* L.) by artificial vision system for quality control. *Appl. Sci.* **2020**, *10*, 5097. [CrossRef]
35. Ropelewska, E.; Sabanci, K.; Aslan, M.F. The Changes in Bell Pepper Flesh as a Result of Lacto-Fermentation Evaluated Using Image Features and Machine Learning. *Foods* **2022**, *11*, 2956. [CrossRef] [PubMed]
36. Melesse, S.; Sobratee, N.; Workneh, T. Application of logistic regression statistical technique to evaluate tomato quality subjected to different pre- and postharvest treatments. *Biol. Agric. Hortic.* **2016**, *32*, 277–287. [CrossRef]
37. Lammertyn, J.; Scheerlinck, N.; Jancsók, P.; Verlinden, B.E.; Nicolaï, B.M. A respiration-diffusion model for 'Conference' pears I: Model development and validation. *Postharvest Biol. Technol.* **2003**, *30*, 29–42. [CrossRef]
38. Dixon, J.; Dixon, J.; Pak, H.A.; Smith, D.B.; Elmsly, T.A.; Cutting, J.G.M. New Zealand Avocado Fruit Quality: The Impact of Storage Temperature and Maturity. In Proceedings of the V World Avocado Congress (Actas V Congreso Mundial del Aguacate), Granada-Málaga, Spain, 19–24 October 2003.
39. Díaz-Pérez, M.; Carreño-Ortega, Á.; Gómez-Galán, M.; Callejón-Ferre, Á.J. Marketability probability study of cherry tomato cultivars based on logistic regression models. *Agronomy* **2018**, *8*, 176. [CrossRef]
40. Huang, X.; Wang, H.; Qu, S.; Luo, W.; Gao, Z. Using artificial neural network in predicting the key fruit quality of loquat. *Food Sci. Nutr.* **2021**, *9*, 1780–1791. [CrossRef]
41. Cho, B.-H.; Koseki, S. Determination of banana quality indices during the ripening process at 1 different temperatures using smartphone images and an artificial neural 2 network. *Sci. Hortic.* **2021**, *288*, 110382. [CrossRef]
42. Bhargava, A.; Bansal, A. Automatic Detection and Grading of Multiple Fruits by Machine Learning. *Food Anal. Methods* **2020**, *3*, 751–761. [CrossRef]
43. Pourdarbani, R.; Sabzi, S.; Hernández-Hernández, M.; Hernández-Hernández, J.L.; García-Mateos, G.; Kalantari, D.; Molina-Martínez, J.M. Comparison of different classifiers and the majority voting rule for the detection of plum fruits in garden conditions. *Remote Sens.* **2019**, *11*, 2546. [CrossRef]