

Article

Development of Multimodal Fusion Technology for Tomato Maturity Assessment

Yang Liu ^{1,†}, Chaojie Wei ^{1,†}, Seung-Chul Yoon ², Xinzhi Ni ³, Wei Wang ^{1,*}, Yizhe Liu ¹, Daren Wang ¹, Xiaorong Wang ¹ and Xiaohuan Guo ¹

¹ Beijing Key Laboratory of Optimization Design for Modern Agricultural Equipment, College of Engineering, China Agricultural University, Beijing 100083, China

² Quality & Safety Assessment Research Unit, U. S. National Poultry Research Center, USDA-ARS, 950 College Station Rd., Athens, GA 30605, USA

³ Crop Genetics and Breeding Research Unit, United States Department of Agriculture Agricultural Research Service, 2747 Davis Road, Tifton, GA 31793, USA

* Correspondence: playerwxw@cau.edu.cn

† These authors contributed equally to this work.

Abstract: The maturity of fruits and vegetables such as tomatoes significantly impacts indicators of their quality, such as taste, nutritional value, and shelf life, making maturity determination vital in agricultural production and the food processing industry. Tomatoes mature from the inside out, leading to an uneven ripening process inside and outside, and these situations make it very challenging to judge their maturity with the help of a single modality. In this paper, we propose a deep learning-assisted multimodal data fusion technique combining color imaging, spectroscopy, and haptic sensing for the maturity assessment of tomatoes. The method uses feature fusion to integrate feature information from images, near-infrared spectra, and haptic modalities into a unified feature set and then classifies the maturity of tomatoes through deep learning. Each modality independently extracts features, capturing the tomatoes' exterior color from color images, internal and surface spectral features linked to chemical compositions in the visible and near-infrared spectra (350 nm to 1100 nm), and physical firmness using haptic sensing. By combining preprocessed and extracted features from multiple modalities, data fusion creates a comprehensive representation of information from all three modalities using an eigenvector in an eigenspace suitable for tomato maturity assessment. Then, a fully connected neural network is constructed to process these fused data. This neural network model achieves 99.4% accuracy in tomato maturity classification, surpassing single-modal methods (color imaging: 94.2%; spectroscopy: 87.8%; haptics: 87.2%). For internal and external maturity unevenness, the classification accuracy reaches 94.4%, demonstrating effective results. A comparative analysis of performance between multimodal fusion and single-modal methods validates the stability and applicability of the multimodal fusion technique. These findings demonstrate the key benefits of multimodal fusion in terms of improving the accuracy of tomato ripening classification and provide a strong theoretical and practical basis for applying multimodal fusion technology to classify the quality and maturity of other fruits and vegetables. Utilizing deep learning (a fully connected neural network) for processing multimodal data provides a new and efficient non-destructive approach for the massive classification of agricultural and food products.

Citation: Liu, Y.; Wei, C.; Yoon, S.-C.; Ni, X.; Wang, W.; Liu, Y.; Wang, D.; Wang, X.; Guo, X. Development of Multimodal Fusion Technology for Tomato Maturity Assessment. *Sensors* **2024**, *24*, 2467. <https://doi.org/10.3390/s24082467>

Academic Editor: Beniamino Gioli

Received: 7 March 2024

Revised: 2 April 2024

Accepted: 10 April 2024

Published: 11 April 2024



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

Keywords: multimodal fusion; tomato maturity; deep learning; non-destructive testing

1. Introduction

In recent years, market globalization has significantly increased the demand for high-quality products, underscoring the necessity of rapidly, non-destructively evaluating produce maturity, such as predicting the ripening stages and storage shelf lives of fruits and vegetables like tomatoes [1]. The accurate assessment of maturity also plays a key role in

supply chain management. Tomatoes, which are rich in vitamins A and C, boast bioactive components with antioxidant properties—carotenoids, flavonoids, and phenolic acids [2]. Furthermore, their antioxidant, anti-inflammatory, and antihypertensive properties have been linked to reducing the risk of cardiovascular diseases [3]. As tomatoes continue to ripen and undergo physiological changes post harvest [4], variability in their maturity post harvest can lead to higher spoilage risks in over-ripe tomatoes, resulting in significant losses during transportation and stores. Therefore, the objective and accurate classification of tomato maturity at harvest is crucial for grading and marketing, which could enhance the farm gate value of tomato production.

Traditional methods for assessing fruit and vegetable maturity, such as compression and puncture tests, have been widely used but are destructive and compromise the profitability of produce [5,6]. With technological advancements, a range of non-destructive techniques, including haptics [7,8], machine vision [9,10], visible and near-infrared spectroscopy (Vis/NIR) [11,12], hyperspectral imaging [13,14], electronic noses (e-noses) [15,16], and magnetic resonance imaging (MRI), now facilitate comprehensive maturity detection without damage [17,18]. These methods enable the evaluation of both internal and external fruit characteristics, offering insights into multiple quality parameters simultaneously. Specifically, non-destructive techniques like machine vision, Vis/NIR spectroscopy, and haptics have become pivotal in assessing tomato maturity and color grades. Machine vision, enhanced by deep learning, has shown remarkable success in classifying tomato maturity, with accuracies exceeding 90.7% [19–21]. However, while it is adept at identifying external maturity indicators, machine vision struggles with internal changes. Conversely, Vis/NIR spectroscopy has been effective in examining internal attributes, with various wavelengths revealing distinct ripening stages [22,23]. Additionally, the evolution of flexible sensor technology has propelled the use of tactile sensors in robotic applications, offering precise maturity assessments through firmness detection [24–27].

Assessing fruit maturity is a complex process encompassing color, internal quality, and firmness [28]. Current maturity detection often relies on a single non-destructive testing (NDT) method, yielding partial information that may lead to inaccurate maturity assessments. While imaging techniques can evaluate appearance, they often cannot adequately assess internal quality and firmness. Similarly, Vis/NIR spectroscopy offers insights into internal quality via spectral data, and tactile sensing technology captures fruit firmness through signal strength. However, each method's perspective does not fully address the multidimensional nature of fruit maturity. Multimodal fusion technology, which integrates data from multiple NDT methods, has emerged as a solution to overcoming these limitations, enhancing accuracy and compensating for the limitations of individual methods. Currently, multimodal fusion techniques have been used in a variety of fields, such as medical diagnosis [29], emotion recognition [30], education [31], industrial fault diagnosis [32], and autonomous driving [33]. A number of reports have applied multimodal fusion technology to agriculture-related research [34–38]. The application of multimodal fusion technology in agricultural technology, especially in the assessment of fruit and vegetable maturity, not only improves the accuracy and efficiency of the assessment but also provides a new perspective for agricultural production and quality assessment supply chain management.

The utilization of deep learning for determining the maturity of fruits and vegetables has garnered significant interest lately. Researchers are increasingly applying deep learning algorithms to decipher ripening data for maturity assessments. For instance, Suharjito et al. [39] enhanced the quality of oil palm fruits by integrating deep learning with machine vision for maturity detection and classification. Similarly, Raghavendra et al. [40] leveraged a combination of convolutional neural networks (CNNs) and multilayer perceptrons (MLPs) with data from RGB and hyperspectral imaging to accurately identify banana maturity, achieving a remarkable accuracy of 98.4%. Deep learning not only increases recognition accuracy but also expedites processing, facilitates the extraction of features from intricate datasets, and enables precise classification and prediction.

This study aimed to classify tomato maturity through a tri-modal fusion approach incorporating imaging, Vis/NIR spectroscopy, and haptic technologies. Its significant contributions include the following: (1) establishing a multimodal tomato dataset reflecting various maturity stages, using data comprising RGB images, transmission spectra, and haptic signals, categorized into immature, semi-mature, and mature stages; (2) analyzing disparities in ripening within and on the surface of tomatoes that lead to misjudgments when using single-modal technologies and addressing these through multimodal fusion techniques; and (3) designing a multimodal fusion classification network for accurate maturity estimation.

2. Materials and Methods

2.1. Experimental Design

Tomatoes without stems were obtained from a greenhouse farm in Tongzhou District, Beijing, China. Based on USDA standards [41] and previous studies [23,42], the color of the exocarp and cut surface along the equator of each tomato was obtained. The maturity categories in this study are defined as follows: immature (less than 10% red on the exocarp and cross-section), mature (more than 90% red on the exocarp and cross-section), and semi-mature (between 10 and 90% red on the exocarp or cross-section). In total, 214 tomatoes were classified into 79 immature, 60 semi-mature, and 75 mature tomatoes. In order to further study the multimodal technique, we also selected two negative samples and one positive sample for analysis. Negative sample 1 tomatoes had a red pericarp but a light green inner fruit cavity and flesh during the maturing process; negative sample 2 tomatoes had a red pericarp but a white inner cavity; and positive sample tomatoes had a uniform red color inside and outside for the best maturity. To mitigate environmental impact on prediction accuracy, the tomato samples were acclimated to a laboratory setting at 20 °C and 60% relative humidity for 12 h. To minimize error and ensure data reliability, each of the 214 tomato fruits was measured four times along the equatorial plane using each of the three devices, rotating 90° between acquisitions. This process generated 2568 data acquisitions encompassing image, spectral, and haptic modalities. Figure 1 shows a flowchart of the multimodal fusion tomato maturity prediction process.

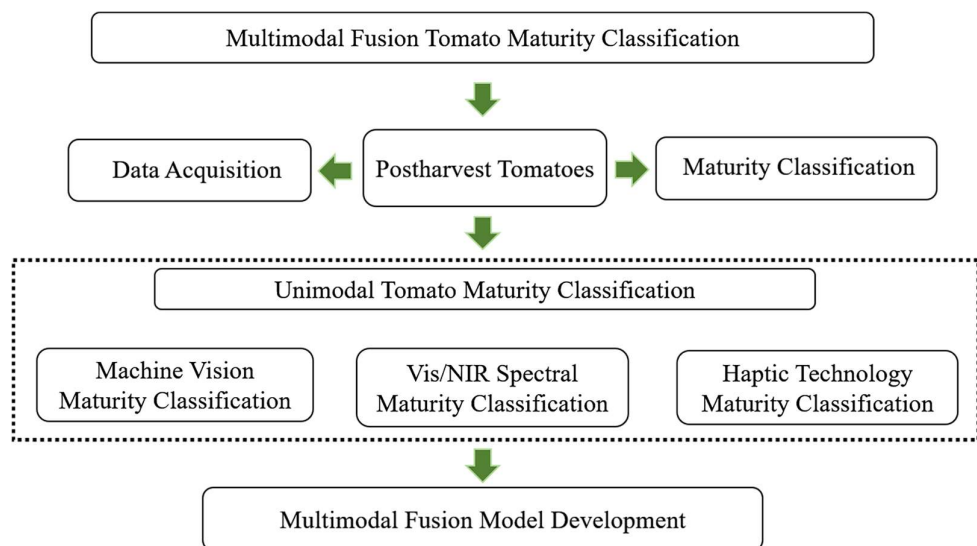


Figure 1. Flowchart of multimodal fusion tomato maturity prediction process.

2.2. Data Acquisition

2.2.1. Image Acquisition

A schematic diagram of the data acquisition device, as shown in Figure 2, shows the acquisition devices for image, spectral, and tactile information from left to right.

Color images were captured using a DaHeng MER-030-120UC color industrial camera. The tomatoes were placed on a conveyor platform and rolled along it. Each tomato fully entered the camera's field of view for image acquisition. It was rotated 90° for each acquisition and turned four times to capture a complete image of the tomato's surface. The surfaces of the light sources were equipped with diffuse reflectors to prevent any exposure on the tomato's outer surface. Each diffuse reflector was located between the light source and the sample and was made of frosted glass.

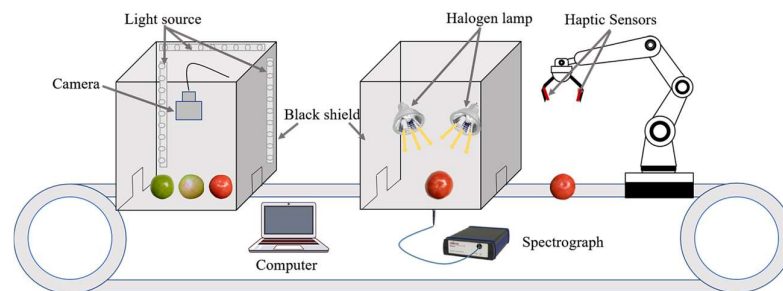


Figure 2. Diagram of data acquisition device.

2.2.2. Vis/NIR Spectral Information Acquisition

Spectral data were acquired using an AvaSpec-ULS2048XL-EVO spectrometer, covering a wavelength range of 350 nm to 1100 nm. Two 250 W halogen lamps, positioned 20 cm away from the sample, served as light sources in a transmission-based acquisition system. To guarantee the light's stability and uniformity, the lamps were preheated for 15 min prior to the experiments. The sample was positioned on a fruit cup, beneath which optical fibers were arranged to capture spectral information. Data acquisition was facilitated through an external trigger mechanism employing a photoelectric switch to gather spectral data. The integration time was set to 35 ms, with averaging performed five times to ensure accuracy and consistency in data collection.

2.2.3. Tactile Information Acquisition

The haptic information acquisition system utilized flexible thin-film pressure sensors (i-Motion, model IMS-C10A) attached to both sides of mechanical jaws to detect the sample's firmness. Data collection was conducted at a sampling frequency of 100 Hz for 15 s. The method entailed positioning a tomato on the experimental platform, maneuvering the robotic arm toward the sample at a set speed of 0.5 m/s, and then clamping the mechanical jaws to secure the tomato and collect data. To mitigate the impact of variable surface firmness on tomatoes, data were obtained from four equatorial points on each tomato. The overall pressure value for each tomato was calculated by averaging these four readings.

2.3. Data Preprocessing

Corresponding preprocessing was performed for the image, spectral, and tactile raw data, respectively, to reduce noise and improve the signal-to-noise ratio.

The chromatic distinction between the tomato and its background was pronounced. For image processing, we converted the image to grayscale and applied a fixed threshold of 50 to separate the tomato from the background. Subsequently, the pixel dimensions of the segmented image were standardized to 150 × 150.

Spectral data underwent preprocessing to counteract noise and bias from the spectrometer's dark current, including a black-and-white correction based on reference data obtained with the light source on (R_w) and off (R_d), using the formula

$$R = \frac{R_{raw} - R_d}{R_w - R_d} \quad (1)$$

The haptic data experienced instability during collection which was attributed to mechanical vibrations. We designed a Butterworth low-pass filter to eliminate the impact of environmental factors. We set the cutoff frequency of the filter to 10 Hz and the order to 4 for filtering.

2.4. Sample Quality Measurement

The firmness of the fruit was defined as the force (F) per unit area (A) using the equation $P = F/A$, and it was computed using the measured force value in newtons and the probe's contact area in square meters, where P was the firmness of the measured fruit and the unit was Pa. The firmness of the tomato fruit was measured by pushing the plunger tip (8 mm) into the opposite cut surface along the equatorial region using a hand-held penetrometer (GY-4, HANDPI, Beijing, China), and the value was expressed in Pa. Subsequently, the SSC of each tomato fruit was determined using traditional destructive methods. Each tomato was cut into small pieces, juice was then extracted from the entire tomato and filtered using double-layer gauze, and 1 mL of juice was dropped onto a fruit sugar refractometer (PAL-1, ATAGO, Tokyo, Japan).

2.5. A Deep Learning Framework for Multimodal Fusion

Multimodal fusion involves combining data from different modalities. Commonly used multimodal information fusion methods include feature-level fusion, decision-level fusion, and model-level fusion [43]. This study employed a multimodal feature fusion scheme. Following data collection, we performed data preprocessing and feature extraction for all modalities, including background removal for images, a black-and-white correction for spectral data, and filtering for tactile data. The features were then fused, and the fused features were input into a fully connected neural network for maturity prediction, as described next.

2.5.1. Feature Extraction

Visual Geometry Group 16 (VGG16), developed by Oxford University's Visual Geometry Group in 2014, is a convolutional neural network adept at processing image data [44]. VGG16 consists of five convolutional blocks and three fully connected layers. The first two convolutional blocks are composed of two convolutional layers and one pooling layer, and the last three convolutional blocks are composed of three convolutional layers and one pooling layer. The process of image feature extraction is shown schematically in Figure 3a. This study utilized VGG16 for tomato image feature extraction, as depicted in Figure 3a. By inputting pre-processed tomato images into the VGG16 model, we extracted a 1×8192 -dimensional feature vector from the output of the model's fully connected layer, representing the tomato image feature.

The primary benefit of CNN lies in its algorithmic strategy for directly extracting features from raw input data [45]. Its capabilities in learning and classification surpass those of conventional neural networks. Hence, a 1D-CNN model was developed for spectral feature extraction, as depicted in Figure 3b. The 1D-CNN model was composed of three convolutional layers; each layer was followed by batch normalization, a ReLU activation function, and a maximum pooling layer, and, finally, the classification results were output through two linear layers. The de-noised spectral data were fed into the 1D-CNN model, in which the convolution layer served the purpose of feature extraction, yielding a 1×10 -dimensional spectral feature vector.

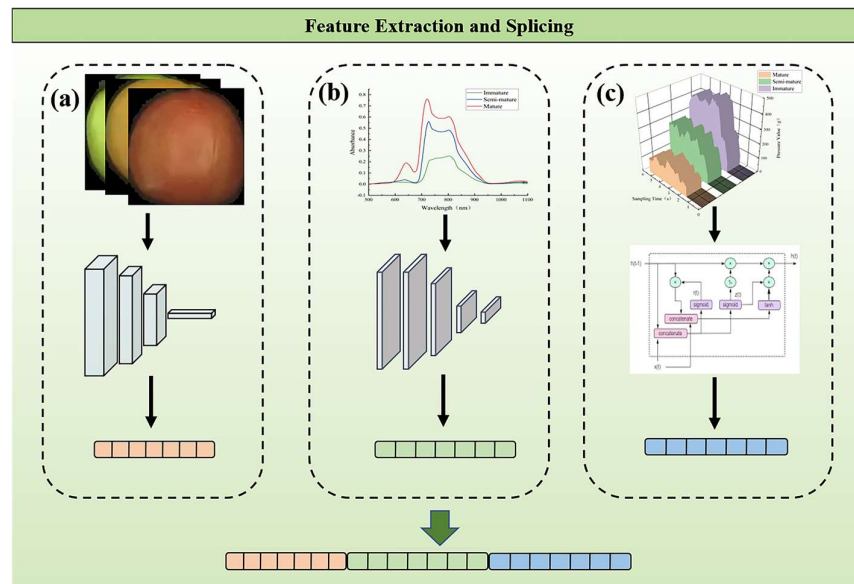


Figure 3. Diagram of feature extraction and fusion. (a) Image feature extraction; (b) Vis/NIR spectral feature extraction; (c) haptic feature extraction.

A Long Short-Term Memory (LSTM) network, a specialized variant of Recurrent Neural Network (RNN), incorporates long-term memory into an RNN by facilitating constant error backpropagation through its internal memory cells [46]. This attribute renders it exceptionally suitable for processing sequences with significant temporal structures, finding extensive application in tactile information processing [47,48]. An LSTM network was designed to decode one-dimensional tactile signals, as depicted in Figure 3c, and comprised three gates (input, forget, and output) and a unit state. This structure allowed the network to retain information over ambiguous time periods and to learn detailed temporal features. Our haptic dataset was input into the model, and the model fully learned the tactile feature information and finally extracted a 1×64 -dimensional tactile feature vector to represent the texture information of tomato.

2.5.2. Feature Fusion

Various feature fusion techniques exist, such as feature weighting [49] and feature mapping [50], yet these approaches often necessitate complex model adjustments and might lead to information loss. Feature weighting may fail to fully capture information across all modes, and feature mapping could compress information, omitting vital details. In multimodal feature fusion, early data fusion occurs at the input layer, known as feature stitching [37]. We utilized feature splicing to merge image, spectral, and haptic data in the fusion process. This method combined different modes into a single multimodal input, $[m, n, z]$, where m , n , and z denote the feature vectors for image, spectral, and haptic data, respectively. Thus, we created a comprehensive cross-modal feature vector with a dimensionality of 1 and a total length of 8266, equal to the sum of the three feature vectors. These spliced feature vectors were then input into the deep learning model for training, validation, and testing.

This approach not only augmented feature dimensionality but also preserved the integrity of each data type. By enabling the deep learning network to independently assess the relevance of each data type, we enhanced information richness and reduced error potential. This method offered a nuanced and thorough analysis, allowing for comprehensive insights into tomato maturity.

2.5.3. Multimodal Fusion Classification Networks

In this study, a fully connected deep neural network model was proposed for processing multimodal data, as shown in Figure 4. The model adopted a multi-layer fully connected architecture that incorporated a residual learning mechanism to enhance the model's learning ability and generalization performance when processing complex datasets. The model had a total of four fully connected layers, each of which contained 8266, 512, 512, and 256 neurons, respectively, and each fully connected layer was immediately followed by a batch normalization layer, a ReLU activation function, and a Dropout layer. The first two fully connected layers were used for feature extraction, and two fully connected layers were used for classification. In addition, a residual block was introduced between the second and third fully connected layers which contained two fully-connected layers, and the two connected layers were followed by the batch normalization and Dropout layers after each layer. The batch normalization layer was used to normalize the output of the layer to improve training stability and efficiency. The ReLU activation function was applied to introduce nonlinearities and enhance the expressive power of the model. The Dropout layer prevented overfitting by setting a reasonable scale size. The whole network realized the efficient processing of multimodal data through this layered structure and residual learning strategy.

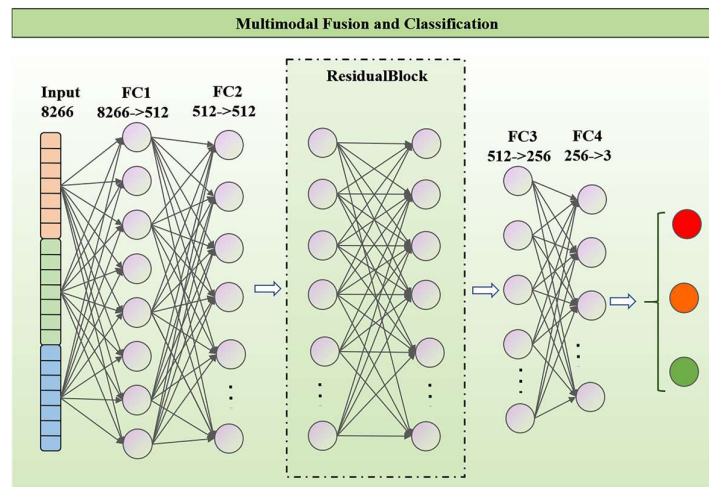


Figure 4. Multimodal fusion fully connected network structure.

2.6. Model Evaluation

To comprehensively evaluate the model performance, we chose three key metrics for the classification model, namely, accuracy, precision, and recall, to evaluate the model.

A TP (True Positive) result indicated that the actual situation was a positive situation and the prediction was positive, i.e., the prediction was correct, and the same result could be obtained for TN (True Negative), FP (False Positive), and FN (False Negative) results [51].

Accuracy reflected the proportion of samples that the model recognized correctly and was a fundamental measure of the overall performance of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

The precision rate, on the other hand, focused on the proportion of samples predicted by the model to be positively classified over all positively classified samples (TP + FP) and focused on assessing the classification accuracy of the model.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall measured the proportion of positively classified samples identified by the model among all positively classified samples ($TP + FN$), and it is critical for assessing the model's ability to capture positively classified samples.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

These three indicators together constitute a comprehensive and in-depth system of evaluating the model's performance which can reflect the performance of the model in multimodal data processing.

The experimental computer platform utilized Windows 10 Professional 64-bit on an Intel Core i7-7700HQ @ 2.80GHz quad-core processor. The maturity assessment model was developed using PyCharm IDE and the PyTorch framework, with training conducted on an NVIDIA GeForce GTX 1050 Ti 4GB GPU.

3. Results and Discussion

3.1. Analysis of Soluble Solids and Firmness of Tomatoes

The internal nutrients of tomatoes change during ripening in addition to changes in the exocarp and internal color. This study also investigated changes in the soluble solids content and hardness during the tomato ripening process, which can further be used for the establishment of SSC and hardness prediction models in future research. As shown in Figure 5, the soluble solids content of the tomatoes showed an ascending trend as they ripened, while the firmness of the tomatoes decreased gradually. The mean SSC values of tomatoes in different ripening periods were 4.15, 4.85, and 5.5 Brix, and the mean values of firmness were 2.82, 1.96, and 1.1 MPa. These findings indicate a clear negative correlation between the soluble solids content and fruit firmness during the tomato maturation process. However, these traditional methods of maturity assessment are destructive, posing challenges for rapid and non-destructive testing requirements in fruit and vegetable quality monitoring and inspection.

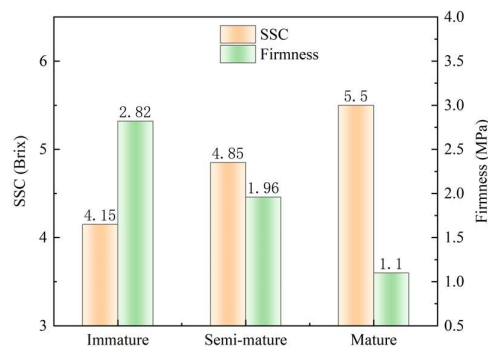


Figure 5. Soluble solids content and firmness values of tomatoes in different maturity stages.

3.2. Analysis of Original Data

3.2.1. Image Data Analysis

As illustrated in Figure 6a, significant color changes were observed in the tomatoes during ripening. The exocarp and mesocarp evolved from greenish-white to pink and finally to bright red, while the endocarp transitioned from green to white to red. Similarly, the funiculus color shifted from green to light yellow, culminating in red at full maturity. These color changes effectively illustrate the progression from immature to semi-mature and ultimately mature stages, highlighting the dynamic nature of the ripening process. Images, as the most commonly used means of determining tomato ripeness, are capable of capturing color changes in exterior appearance, and, as shown in Figure 6a, image processing techniques are able to calculate the percentage of red color on the surface of a

tomato and classify it correctly, thus accurately determining tomato ripeness by appearance. Due to a variety of factors, tomato maturity was not uniform, and there were cases in which the exterior was mature but the interior was not, making it impossible to identify the interior color and texture from the exterior image. As shown in Figure 6b, we analyzed three categories of tomatoes with a red appearance but different internal maturity conditions and determined by their appearance that the tomatoes had all reached maturity. Negative samples 1 and 2 were judged to be mature according to their appearance; the outer skin color ratio reached more than 95% but the interior was not mature, and the color ratio was about 70%. The color ratio of the positive sample reached more than 95%, showing the best mature state. Immature tomatoes contain lycopene, which can be harmful to the human body after consumption [52], so it is necessary to combine a variety of nondestructive testing techniques to determine the maturity of tomatoes comprehensively.

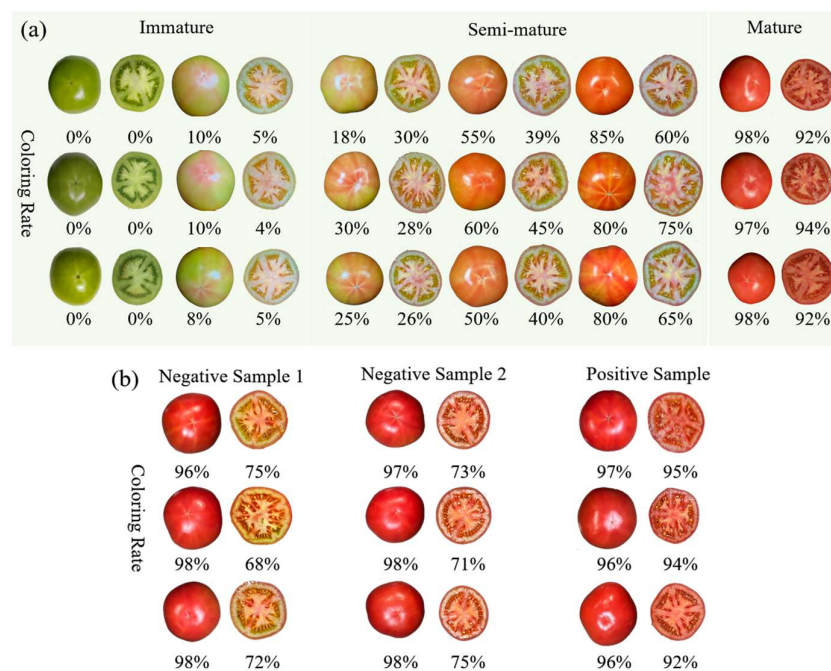


Figure 6. Images of tomatoes at different ripening stages; (a) images of tomatoes at three ripening stages: immature, semi-mature and mature; and (b) images of cross-sectioned unevenly mature tomatoes.

3.2.2. Analysis of Spectral Data

The ripening of the tomatoes was marked by significant changes in their internal nutrient composition, including a decrease in chlorophyll content and an increase in anthocyanin content [53]. These alterations are mirrored in the measured spectral absorbance curves presented in Figure 7a. The curves demonstrate consistent trends across different maturity stages, with mature tomatoes exhibiting higher spectral intensities between 600 and 950 nm. Notable absorption peaks were observed at 630 nm, 730 nm, 830 nm, and 1070 nm, corresponding to chlorophyll pigments, the O-H band's third overtone, the C-H band's fourth overtone, and the absorption region for carbohydrates [54]. Additionally, the 970 nm–1180 nm range was influenced by O-H bonds [22]. During tomato ripening, a decrease in chlorophyll content and an increase in lycopene combined to cause a change in the light transmission spectral curve of the tomato. These changes indicate a significant correlation between the optical properties of tomatoes and their natural growth processes. The spectra also can discriminate between unevenly internally and externally ripened tomatoes. The mean spectra with standard deviation values for the three categories of

tomatoes, namely negative sample 1, negative sample 2, and the positive sample, are depicted in the Figure 7b. The spectral trends of the three types of samples are basically consistent. There is a slight difference in spectral signals between negative sample 2 and the normal sample, but they have a significant difference compared to negative sample 1. For similar maturity, Vis/NIR spectroscopy could not completely differentiate maturity categories, so it was necessary to further rely on other NDT measurements. In daily life, touching tomatoes with hands to determine their texture is one of the important ways of judging ripeness in addition to vision. Therefore, haptic technology was introduced to solve the problem of tomato ripeness assessment.

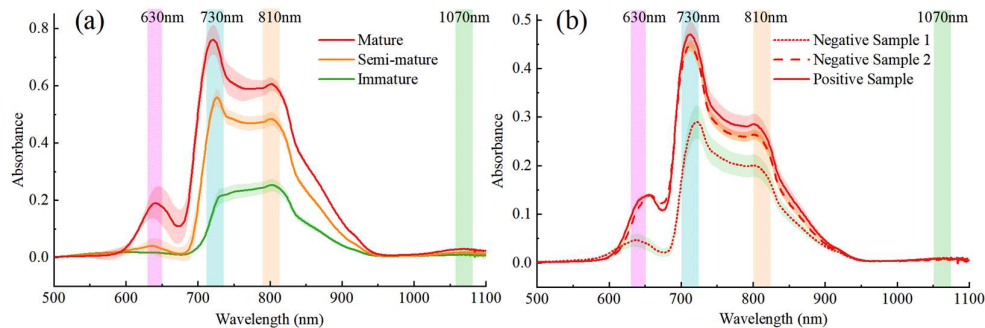


Figure 7. Spectral information of tomatoes at different maturity stages; (a) spectral information of tomatoes at three ripening stages: immature, semi-mature, and mature; (b) spectral information of cross-sectioned unevenly mature tomatoes.

3.2.3. Analysis of Haptic Data

Pronounced changes in tomato firmness throughout ripening were depicted using 3D color-mapped surfaces of tactile pressure signals (Figure 8a). Mature tomatoes exhibited pressure values ranging from 13 to 26 kPa, semi-mature tomatoes exhibited pressure values ranging from 26 to 46 kPa, and immature tomatoes exhibited pressure values ranging from 52 to 65 kPa. These variations were attributed to the enzymatic breakdown of pectin, leading to tissue softening [55]. Consequently, the marked differences in firmness at each ripening stage suggest a potential for non-destructive maturity assessments through haptic analysis, offering a straightforward method for evaluating tomato quality. In a case of heterogeneous internal and external ripening, the interior of negative sample 2 was off-white and the texture of the tomato was hard, so we utilized the characteristic of negative sample 2's hard texture to differentiate negative sample 2 from the positive sample by the haptic technique. As shown in Figure 8b, negative samples 1 and 2 were less mature and had a hard texture, but the mature positive sample had a softer texture. From the data, the pressure values of negative samples 1 and 2 reached approximately 29 kPa, and the pressure value of the positive sample reached about 13 kPa. Therefore, negative sample 2 and the positive sample could be distinguished through the difference in pressure values, thereby making up for the inability of the spectrum to distinguish samples from the two groups.

Although the internal ripening situation could not be identified through external color in the images, the combination of near-infrared spectroscopy and haptic sensing technology avoided the shortcomings of the images. The study results suggest that during the ripening process of tomatoes, the apparent color, internal structure, and firmness of tomatoes became inhomogeneous, resulting in inaccurate single-modal maturity classification, and the interference of other factors during the ripening process could not be ruled out.

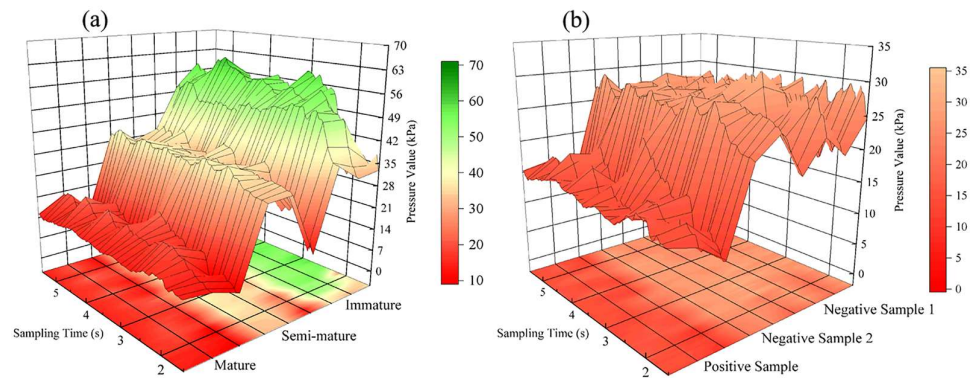


Figure 8. The haptic information of tomatoes at different maturity stages; (a) the haptic information of tomatoes at three maturity stages: immature, semi-mature, and mature; (b) the haptic information of cross-sectioned unevenly mature tomatoes.

3.3. Multimodal Fusion Maturity Classification Model

3.3.1. Unimodal Maturity Classification

Before performing multimodal fusion classification, we first built unimodal deep learning classification models for the three categories of tomato maturity: immature, semi-mature, and mature. We developed three deep learning models for this purpose: a VGG16 model for imaging, a one-dimensional CNN for spectral analysis, and an LSTM network for haptic data. The dataset was split as follows: 64% for training, 16% for validation, and 20% for testing. The dataset was divided into a training set, test set, and validation set at a ratio of 8:4:5. The outcomes of model training are detailed in Table 1.

Table 1. Tomato unimodal classification performance.

Models	Accuracy			Precision			Recall		
	Training Set	Validation Set	Test Set	Training Set	Validation Set	Test Set	Training Set	Validation Set	Test Set
Imagery	94.0%	93.4%	94.2%	94.5%	94.6%	94.2%	94.8%	94.7%	91.7%
Spectral	87.3%	84.7%	87.8%	89.7%	87.6%	89.9%	66.3%	65.2%	64.8%
Haptic	90.0%	88.7%	87.2%	90.4%	89.7%	89.9%	66.7%	66.7%	66.7%

Table 1 shows the accuracy, precision, and recall of the models built on images, spectra, and haptics, and the performance of the test set is further discussed. The three optimal image, spectral, and haptic unimodal classification models reached 99.2%, 87.8%, and 87.2% accuracy and 89.9%, 89.9%, and 99.2% precision. The recall rates of the three optimal unimodal classification models showed poor levels in general, especially for the spectral and haptic classification models that demonstrated only 64.8% and 66.7%, respectively. VGG16 demonstrated high accuracy in classifying tomato maturity, proving its effectiveness in distinguishing image variations associated with maturity levels. This underscored the potent utility of images in discerning tomato ripeness. However, images fell short in detecting the internal ripening stages of tomatoes, necessitating the integration of additional methods for a thorough analysis of tomato maturity. The spectral and haptic analyses showed lower accuracy values, particularly in recall rates. This could be attributed to the models' inability to adequately capture features essential for differentiating between categories, complicating the accurate identification of specific sample categories during prediction. Consequently, for spectral and tactile data, the deployment of more sophisticated models or enhanced feature engineering might be necessary to unearth more distinctive characteristics. For instance, employing more advanced deep learning frameworks or developing custom models tailored to specific tasks could have enhanced model accuracy.

3.3.2. Multimodal Fusion Maturity Classification

To evaluate the efficacy of the multimodal fusion technique, we conducted a comparative analysis between unimodal and multimodal fully connected neural network (FCNN) classification models across image (VGG16), spectral (1D-CNN), and haptic (LSTM) modalities, focusing on recall, precision, and accuracy metrics. Figure 9a showed that the FCNN model achieved the highest performance, with 99.4% recall, 99.3% precision, and 99.4% accuracy. In comparison, the VGG16 model showed 91.7% recall and over 94% for both precision and accuracy. The 1D-CNN and LSTM models demonstrated lower recall rates (66.7% and 64.8%, respectively) but maintained high precision at 89.9%, with accuracy rates of 87.2% and 87.8%. Overall, the FCNN model shows good comprehensive performance, and the recall, precision and accuracy in the test set of the FCNN model are significantly improved compared to the unimodal classification model, with an up to 34.6% improvement in recall and up to 9.4% and 12.2% improvements in precision and accuracy. The effectiveness and superiority of multimodal fusion methods in deep learning classification tasks are demonstrated.

To examine the multimodal fusion classification network from multiple perspectives and demonstrate the classification results more intuitively, we plotted a confusion matrix of the FCNN model's classification results in Figure 9b. Only one mature sample is misclassified as a semi-mature sample, which may have been caused by the closer proximity of the test sample to the semi-mature sample, but overall, the overall classification accuracy is still high. In addition, the loss values (Loss) during the model training process used to comprehensively assess the overall ability of the model were examined, as shown in Figure 9c. The curves of the loss function values of the model on the training and test sets showed a fast decreasing trend and then gradually stabilized. After approximately 30 instances of training time, the loss values of the model's training and test sets were stabilized at 0.03 and 0.09. The convergence of the model could be obtained through the Loss curve, which indicates that the model's training effect is relatively good and the model was able to correctly capture the laws in the training data. By comprehensively analyzing the confusion matrix and loss curves, we were able to understand the classification performance, training process, and generalization ability of each model more comprehensively, thus providing a more solid basis for model selection and adjustment.

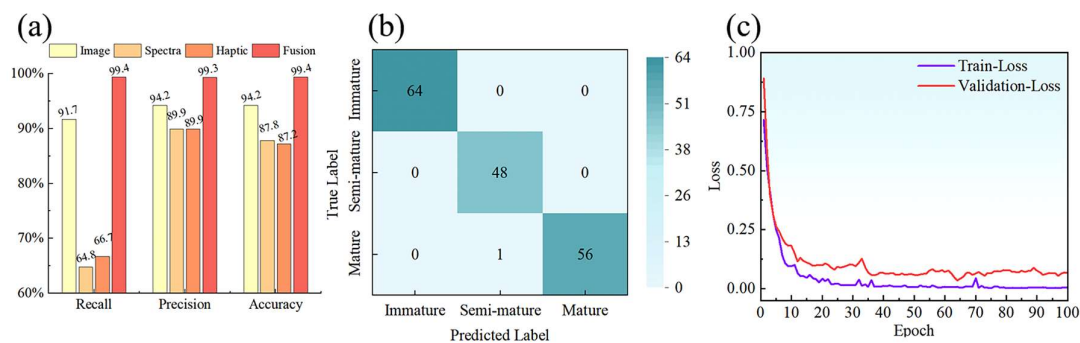


Figure 9. Comparison of model (a) evaluation metrics, (b) confusion matrix, and (c) Loss.

3.3.3. Independent Validation of Heterogeneous Samples of Internal and External Maturity

Beyond classifying tomatoes into immature, semi-mature, and mature categories based on apparent changes, the current study also extended to classifying tomatoes with uniform external redness but varied internal ripeness levels. We employed VGG16, 1D-CNN, and LSTM models to integrate the image, spectral, and haptic data into the fusion model for this purpose. A validation set of eighteen samples yielded an accuracy rate of 94.4%. The confusion matrix displayed in Figure 10a reveals a single instance in which a

negative sample 2 sample was incorrectly classified as a positive sample. The misclassified sample is shown in Figure 10b. We measured a firmness value of 1.3 MPa for this sample. This could be due to the low firmness value of this sample, which resulted in the misclassification of this sample due to the close proximity of the three features extracted from the image, spectral, and haptic data to the positive sample.

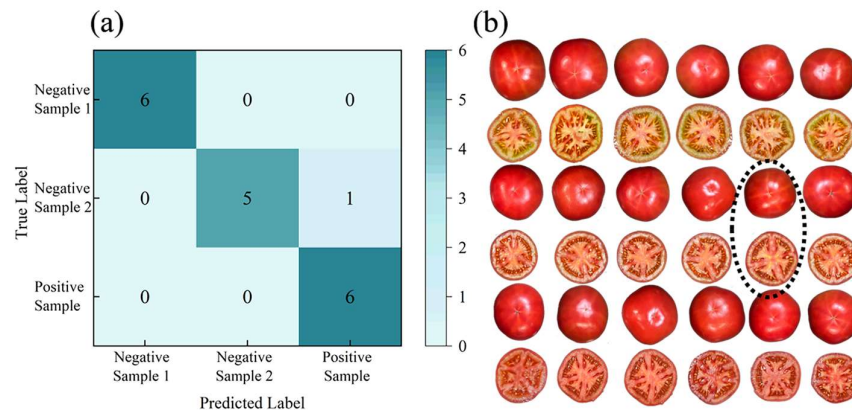


Figure 10. Validation results: (a) confusion matrix; (b) validation sample plot.

4. Conclusions

In this research, a multimodal fusion approach was proposed for classifying the ripeness of tomatoes. Image, Vis/NIR spectral, and haptic data were collected and analyzed. For a comparison with the multimodal fusion approach, three single-modal models were developed, achieving an optimal performance of 94.2%. Further, a multimodal fusion model utilizing a fully connected neural network was established which attained the best classification accuracy of 99.4%, with a 5.2% improvement. The multimodal fusion network outperformed the three traditional single-modal models by achieving a 99.4% accuracy, 99.3% precision, and 99.4% recall in classifying tomato maturity. In particular, the classification accuracy reached 94.4% in the case of inconsistent internal and external maturity. These findings prove that multimodal fusion technology can overcome the limitations of single-modal classifications and highlights its specific advantages. The multimodal fusion method identified different ripened tomatoes, improved food quality, developed a new way of conducting rapid and non-invasive tomato maturity evaluations, and laid a foundation for optimizing harvest time and maximizing the value of tomato products.

Despite the remarkable results of this research obtained, there are still some challenges in integrating this technology into existing online sorting systems, such as the real-time responsiveness of data processing and sorting, the adaptability of the fusion technology to multiple fruits and vegetables, and the cost-effectiveness of sorting systems. Future research requires the optimization of algorithms and hardware design, incorporating more NDT techniques to explore the effectiveness of the large-scale online sorting of tomato ripeness and other quality control parameters, such as pests or virus-infected tomatoes. The multimodal fusion method can be used as a reference for other types of fruits and vegetables in the future, such as kiwi, persimmon, and cucumber. If widely used, this method can improve the quality of fruit and vegetable sales according to the different maturity stages of fruits and vegetables in graded sales, bring significant economic benefits, and improve consumer satisfaction. Therefore, multimodal fusion technology will have a wide impact on the development of agriculture and food quality control technology.

Author Contributions: Conceptualization and writing—original draft preparation, Y.L. (Yang Liu) and C.W.; writing—review and editing, S.-C.Y., X.N., and W.W.; supervision, project

administration, and funding acquisition, W.W. methodology, X.W.; software, Y.L. (Yizhe Liu); validation, D.W. and X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 32272410) and the National Key Research and Development Program of China (No. 2022YFF0607900).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

- Gómez, A.H.; Wang, J.; Hu, G.; Pereira, A.G. Monitoring Storage Shelf Life of Tomato Using Electronic Nose Technique. *J. Food Eng.* **2008**, *85*, 625–631. <https://doi.org/10.1016/j.jfoodeng.2007.06.039>.
- Burton-Freeman, B.M.; Sesso, H.D. Whole Food versus Supplement: Comparing the Clinical Evidence of Tomato Intake and Lycopene Supplementation on Cardiovascular Risk Factors. *Adv. Nutr.* **2014**, *5*, 457–485. <https://doi.org/10.3945/an.114.005231>.
- Michaličková, D.; Belović, M.; Ilić, N.; Kotur-Stevuljević, J.; Slanař, O.; Šobajić, S. Comparison of Polyphenol-Enriched Tomato Juice and Standard Tomato Juice for Cardiovascular Benefits in Subjects with Stage 1 Hypertension: A Randomized Controlled Study. *Plant Foods Hum. Nutr.* **2019**, *74*, 122–127. <https://doi.org/10.1007/s11130-019-0714-5>.
- Seo, D.; Cho, B.-H.; Kim, K.-C. Development of Monitoring Robot System for Tomato Fruits in Hydroponic Greenhouses. *Agronomy* **2021**, *11*, 2211. <https://doi.org/10.3390/agronomy11112211>.
- Nguyen, L.T.; Tay, A.; Balasubramaniam, V.M.; Legan, J.D.; Turek, E.J.; Gupta, R. Evaluating the Impact of Thermal and Pressure Treatment in Preserving Textural Quality of Selected Foods. *LWT-Food Sci. Technol.* **2010**, *43*, 525–534. <https://doi.org/10.1016/j.lwt.2009.09.022>.
- Sirisomboon, P.; Tanaka, M.; Kojima, T. Evaluation of Tomato Textural Mechanical Properties. *J. Food Eng.* **2012**, *111*, 618–624. <https://doi.org/10.1016/j.jfoodeng.2012.03.007>.
- Zhang, J.; Wang, X.; Xia, J.; Xing, S.; Zhang, X. Flexible Sensing Enabled Intelligent Manipulator System (FSIMS) for Avocados (*Persea Americana* Mill) Ripeness Grading. *J. Clean. Prod.* **2022**, *363*, 132599. <https://doi.org/10.1016/j.jclepro.2022.132599>.
- Chen, Y.; Lin, J.; Du, X.; Fang, B.; Sun, F.; Li, S. Non-Destructive Fruit Firmness Evaluation Using Vision-Based Tactile Information. In Proceedings of the IEEE 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 2303–2309.
- Sabzi, S.; Nadimi, M.; Abbaspour-Gilandeh, Y.; Paliwal, J. Non-Destructive Estimation of Physicochemical Properties and Detection of Ripeness Level of Apples Using Machine Vision. *Int. J. Fruit Sci.* **2022**, *22*, 628–645. <https://doi.org/10.1080/15538362.2022.2092580>.
- Miraei Ashtiani, S.-H.; Javanmardi, S.; Jahanbanifard, M.; Martynenko, A.; Verbeek, F.J. Detection of Mulberry Ripeness Stages Using Deep Learning Models. *IEEE Access* **2021**, *9*, 100380–100394. <https://doi.org/10.1109/ACCESS.2021.3096550>.
- Jiang, X.; Zhu, M.; Yao, J.; Zhang, Y.; Liu, Y. Calibration of Near Infrared Spectroscopy of Apples with Different Fruit Sizes to Improve Soluble Solids Content Model Performance. *Foods* **2022**, *11*, 1923. <https://doi.org/10.3390/foods11131923>.
- Wedding, B.B.; Wright, C.; Grauf, S.; Gadek, P.; White, R.D. The Application of FT-NIRS for the Detection of Bruises and the Prediction of Rot Susceptibility of ‘Hass’ Avocado Fruit. *J. Sci. Food Agric.* **2019**, *99*, 1880–1887. <https://doi.org/10.1002/jsfa.9383>.
- Varga, L.A.; Makowski, J.; Zell, A. Measuring the Ripeness of Fruit with Hyperspectral Imaging and Deep Learning. In Proceedings of the IEEE 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8.
- Verma, P.K.; Pathak, P.; Kumar, B.; Himani, H.; Preety, P. Automatic Optical Imaging System for Mango Fruit Using Hyperspectral Camera and Deep Learning Algorithm. *IJIRITCC* **2023**, *11*, 112–117. <https://doi.org/10.17762/ijritcc.v11i5s.6635>.
- Tyagi, P.; Semwal, R.; Sharma, A.; Tiwary, U.S.; Varadwaj, P. E-Nose: A Low-Cost Fruit Ripeness Monitoring System. *J. Agric. Eng.* **2022**, *54*. <https://doi.org/10.4081/jae.2022.1389>.
- Chen, L.-Y.; Wu, C.-C.; Chou, T.-I.; Chiu, S.-W.; Tang, K.-T. Development of a Dual MOS Electronic Nose/Camera System for Improving Fruit Ripeness Classification. *Sensors* **2018**, *18*, 3256. <https://doi.org/10.3390/s18103256>.
- Pagés, G.; Deborde, C.; Lemaire-Chamley, M.; Moing, A.; Bonny, J.-M. MRSI vs CEST MRI to Understand Tomato Metabolism in Ripening Fruit: Is There a Better Contrast? *Anal. Bioanal. Chem.* **2021**, *413*, 1251–1257. <https://doi.org/10.1007/s00216-020-03101-w>.
- Kamal, T.; Cheng, S.; Khan, I.A.; Nawab, K.; Zhang, T.; Song, Y.; Wang, S.; Nadeem, M.; Riaz, M.; Khan, M.A.U.; et al. Potential Uses of LF-NMR and MRI in the Study of Water Dynamics and Quality Measurement of Fruits and Vegetables. *J. Food Process. Preserv.* **2019**, *43*, e14202. <https://doi.org/10.1111/jfpp.14202>.
- Kim, J.; Pyo, H.; Jang, I.; Kang, J.; Ju, B.; Ko, K. Tomato Harvesting Robotic System Based on Deep-ToMaToS: Deep Learning Network Using Transformation Loss for 6D Pose Estimation of Maturity Classified Tomatoes with Side-Stem. *Comput. Electron. Agric.* **2022**, *201*, 107300. <https://doi.org/10.1016/j.compag.2022.107300>.

20. Liu, L.; Li, Z.; Lan, Y.; Shi, Y.; Cui, Y. Design of a Tomato Classifier Based on Machine Vision. *PLoS ONE* **2019**, *14*, e0219803. <https://doi.org/10.1371/journal.pone.0219803>.
21. Kao, I.-H.; Hsu, Y.-W.; Yang, Y.-Z.; Chen, Y.-L.; Lai, Y.-H.; Perng, J.-W. Determination of Lycopersicon Maturity Using Convolutional Autoencoders. *Sci. Hortic.* **2019**, *256*, 108538. <https://doi.org/10.1016/j.scienta.2019.05.065>.
22. Alenazi, M.M.; Shafiq, M.; Alsadon, A.A.; Alhelal, I.M.; Alhamdan, A.M.; Solieman, T.H.I.; Ibrahim, A.A.; Shady, M.R.; Saad, M.A.O. Non-Destructive Assessment of Flesh Firmness and Dietary Antioxidants of Greenhouse-Grown Tomato (*Solanum Lycopersicum* L.) at Different Fruit Maturity Stages. *Saudi J. Biol. Sci.* **2020**, *27*, 2839–2846. <https://doi.org/10.1016/j.sjbs.2020.07.004>.
23. Huang, Y.; Dong, W.; Chen, Y.; Wang, X.; Luo, W.; Zhan, B.; Liu, X.; Zhang, H. Online Detection of Soluble Solids Content and Maturity of Tomatoes Using Vis/NIR Full Transmittance Spectra. *Chemom. Intell. Lab. Syst.* **2021**, *210*, 104243. <https://doi.org/10.1016/j.chemolab.2021.104243>.
24. Maharshi, V.; Sharma, S.; Prajesh, R.; Das, S.; Agarwal, A.; Mitra, B. A Novel Sensor for Fruit Ripeness Estimation Using Lithography Free Approach. *IEEE Sens. J.* **2022**, *22*, 22192–22199. <https://doi.org/10.1109/JSEN.2022.3210439>.
25. Azhari, S.; Setoguchi, T.; Sasaki, I.; Nakagawa, A.; Ikeda, K.; Azhari, A.; Hasan, I.H.; Hamidon, M.N.; Fukunaga, N.; Shibata, T.; et al. Toward Automated Tomato Harvesting System: Integration of Haptic Based Piezoresistive Nanocomposite and Machine Learning. *IEEE Sens. J.* **2021**, *21*, 27810–27817. <https://doi.org/10.1109/JSEN.2021.3124914>.
26. TermehYousefi, A.; Azhari, S.; Khajeh, A.; Hamidon, M.N.; Tanaka, H. Development of Haptic Based Piezoresistive Artificial Fingertip: Toward Efficient Tactile Sensing Systems for Humanoids. *Mater. Sci. Eng. C* **2017**, *77*, 1098–1103. <https://doi.org/10.1016/j.msec.2017.04.040>.
27. Parajuli, P.; Yoon, S.-C.; Zhuang, H.; Bowker, B. Characterizing the Spatial Distribution of Woody Breast Condition in Broiler Breast Fillet by Compression Force Measurement. *Food Meas.* **2024**, *18*, 1991–2003. <https://doi.org/10.1007/s11694-023-02330-8>.
28. Jena, A.; Bamola, A.; Mishra, S.; Jain, I.; Pathak, N.; Sharma, N.; Joshi, N.; Pandey, R.; Kaparwal, S.; Yadav, V.; et al. State-of-the-Art Non-Destructive Approaches for Maturity Index Determination in Fruits and Vegetables: Principles, Applications, and Future Directions. *Food Prod. Process. Nutr.* **2024**, *6*, 56.
29. Odusami, M.; Maskeliūnas, R.; Damaševičius, R. Pixel-Level Fusion Approach with Vision Transformer for Early Detection of Alzheimer's Disease. *Electronics* **2023**, *12*, 1218. <https://doi.org/10.3390/electronics12051218>.
30. Chen, L.; Wang, K.; Li, M.; Wu, M.; Pedrycz, W.; Hirota, K. K-Means Clustering-Based Kernel Canonical Correlation Analysis for Multimodal Emotion Recognition in Human–Robot Interaction. *IEEE Trans. Ind. Electron.* **2023**, *70*, 1016–1024. <https://doi.org/10.1109/TIE.2022.3150097>.
31. Chango, W.; Lara, J.A.; Cerezo, R.; Romero, C. A Review on Data Fusion in Multimodal Learning Analytics and Educational Data Mining. *WIREs Data Min. Knowl.* **2022**, *12*, e1458. <https://doi.org/10.1002/widm.1458>.
32. Qiu, S.; Cui, X.; Ping, Z.; Shan, N.; Li, Z.; Bao, X.; Xu, X. Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review. *Sensors* **2023**, *23*, 1305. <https://doi.org/10.3390/s23031305>.
33. Zhou, W.; Dong, S.; Lei, J.; Yu, L. MTANet: Multitask-Aware Network With Hierarchical Multimodal Fusion for RGB-T Urban Scene Understanding. *IEEE Trans. Intell. Veh.* **2023**, *8*, 48–58. <https://doi.org/10.1109/TIV.2022.3164899>.
34. Xia, F.; Lou, Z.; Sun, D.; Li, H.; Quan, L. Weed Resistance Assessment through Airborne Multimodal Data Fusion and Deep Learning: A Novel Approach towards Sustainable Agriculture. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *120*, 103352. <https://doi.org/10.1016/j.jag.2023.103352>.
35. Li, W.; Liu, Z.; Hu, Z. Effects of Nitrogen and Potassium Fertilizers on Potato Growth and Quality under Multimodal Sensor Data Fusion. *Mob. Inf. Syst.* **2022**, *2022*, 6726204. <https://doi.org/10.1155/2022/6726204>.
36. Lan, Y.; Guo, Y.; Chen, Q.; Lin, S.; Chen, Y.; Deng, X. Visual Question Answering Model for Fruit Tree Disease Decision-Making Based on Multimodal Deep Learning. *Front. Plant Sci.* **2023**, *13*, 1064399. <https://doi.org/10.3389/fpls.2022.1064399>.
37. Garillos-Manlriquez, C.A.; Chiang, J.Y. Multimodal Deep Learning and Visible-Light and Hyperspectral Imaging for Fruit Maturity Estimation. *Sensors* **2021**, *21*, 1288. <https://doi.org/10.3390/s21041288>.
38. Garillos-Manlriquez, C.A.; Chiang, J.Y. Multimodal Deep Learning via Late Fusion for Non-Destructive Papaya Fruit Maturity Classification. In Proceedings of the IEEE 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 10–12 November 2021; pp. 1–6.
39. Suhajito; Junior, F.A.; Koeswandy, Y.P.; Debi Nurhayati, P.W.; Asrol, M.; Marimin. Annotated Datasets of Oil Palm Fruit Bunch Piles for Ripeness Grading Using Deep Learning. *Sci. Data* **2023**, *10*, 72. <https://doi.org/10.1038/s41597-023-01958-x>.
40. Ganguli, S.; Selvan, P.T.; Nayak, M.M.; Chaudhury, S.; Espina, R.U.; Ofori, I. Deep Learning Based Dual Channel Banana Grading System Using Convolution Neural Network. *J. Food Qual.* **2022**, *2022*, 6050284. <https://doi.org/10.1155/2022/6050284>.
41. USDA. *U.S. Standards for Grades of Fresh Tomatoes*; United States Department of Agriculture; Agricultural Marketing Service: Washington, DC, USA, 1991.
42. Huang, Y.P.; Wang, D.Z.; Zhou, H.Y.; Yang, Y.T.; Chen, K.J. Ripeness Assessment of Tomato Fruit by Optical Absorption and Scattering Coefficient Spectra. *Spectrosc. Spectr. Anal.* **2020**, *40*, 3556–3561. [https://doi.org/10.3964/j.issn.1000-0593\(2020\)11-3556-0642](https://doi.org/10.3964/j.issn.1000-0593(2020)11-3556-0642).
43. Zhang, S.; Yang, Y.; Chen, C.; Zhang, X.; Leng, Q.; Zhao, X. Deep Learning-Based Multimodal Emotion Recognition from Audio, Visual, and Text Modalities: A Systematic Review of Recent Advancements and Future Prospects. *Expert Syst. Appl.* **2024**, *237*, 121692. <https://doi.org/10.1016/j.eswa.2023.121692>.
44. Jia, L.; Zhai, H.; Yuan, X.; Jiang, Y.; Ding, J. A Parallel Convolution and Decision Fusion-Based Flower Classification Method. *Mathematics* **2022**, *10*, 2767. <https://doi.org/10.3390/math10152767>.

45. Zhang, H.; Li, Y.; Zhang, Y.; Shen, Q. Spectral-Spatial Classification of Hyperspectral Imagery Using a Dual-Channel Convolutional Neural Network. *Remote Sens. Lett.* **2017**, *8*, 438–447. <https://doi.org/10.1080/2150704X.2017.1280200>.
46. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780, .
47. Pastor, F.; García-González, J.; Gandarias, J.M.; Medina, D.; Closas, P.; García-Cerezo, A.J.; Gómez-de-Gabriel, J.M. Bayesian and Neural Inference on LSTM-Based Object Recognition from Tactile and Kinesthetic Information. *IEEE Robot. Autom. Lett.* **2021**, *6*, 231–238. <https://doi.org/10.1109/LRA.2020.3038377>.
48. Bottcher, W.; Machado, P.; Lama, N.; McGinnity, T.M. Object Recognition for Robotics from Tactile Time Series Data Utilising Different Neural Network Architectures. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021.
49. Zhang, J.; Wu, X.; Huang, C. AdaMoW: Multimodal Sentiment Analysis Based on Adaptive Modality-Specific Weight Fusion Network. *IEEE Access* **2023**, *11*, 48410–48420. <https://doi.org/10.1109/ACCESS.2023.3276932>.
50. Zhang, W.; Mi, L.; Thompson, P.M.; Wang, Y. A Geometric Framework for Feature Mappings in Multimodal Fusion of Brain Image Data. In *Information Processing in Medical Imaging*; Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 11492, pp. 617–630, ISBN 978-3-030-20350-4.
51. Liu, B.; Ge, R.; Zhu, Y.; Zhang, B.; Zhang, X.; Bao, Y. IDAF: Iterative Dual-Scale Attentional Fusion Network for Automatic Modulation Recognition. *Sensors* **2023**, *23*, 8134. <https://doi.org/10.3390/s23198134>.
52. Nakayasu, M.; Umemoto, N.; Akiyama, R.; Ohyama, K.; Lee, H.J.; Miyachi, H.; Watanabe, B.; Muranaka, T.; Saito, K.; Sugimoto, Y.; et al. Characterization of C-26 Aminotransferase, Indispensable for Steroidal Glycoalkaloid Biosynthesis. *Plant J.* **2021**, *108*, 81–92. <https://doi.org/10.1111/tpj.15426>.
53. Qin, J.; Lu, R. Measurement of the Optical Properties of Fruits and Vegetables Using Spatially Resolved Hyperspectral Diffuse Reflectance Imaging Technique. *Postharvest Biol. Technol.* **2008**, *49*, 355–365. <https://doi.org/10.1016/j.postharvbio.2008.03.010>.
54. Williams, P.; Norris, K. *Near-Infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists, Inc.: St. Paul, MO, USA, 1987; ISBN 0-913250-49-X.
55. Van Dijk, C.; Boeriu, C.; Stolle-Smits, T.; Tijskens, L.M.M. The Firmness of Stored Tomatoes (*Cv. Tradiro*). 2. Kinetic and Near Infrared Models to Describe Pectin Degrading Enzymes and Firmness Loss. *J. Food Eng.* **2006**, *77*, 585–593. <https://doi.org/10.1016/j.jfoodeng.2005.07.017>.

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