

# Fruit-HSNet: A Machine Learning Approach for Hyperspectral Image-Based Fruit Ripeness Prediction

Ahmed Baha Ben Jmaa<sup>1</sup> <sup>a</sup>, Faten Chaieb<sup>1</sup> <sup>b</sup> and Anna Fabijańska<sup>2</sup> <sup>c</sup>

<sup>1</sup>*Efrei Research Lab, Paris Panthéon-Assas University, Paris, France*

<sup>2</sup>*Institute of Applied Computer Science, Lodz University of Technology, Łódź, Poland*

**Keywords:** Fruit Ripeness Prediction, Hyperspectral Image, DeepHS Fruit Dataset, Smart Agriculture.

**Abstract:** Fruit ripeness prediction (FRP) is a classification-based agricultural computer vision task that has attracted much attention, thanks to its wide-ranging advantages in agriculture field for both pre-harvest and post-harvest management. Accurate and timely FRP can be achieved using machine/deep learning-based hyperspectral image classification techniques. However, challenges including the limited availability of labeled data and the lack of robust methods generalizable to various hyperspectral cameras and fruit types can compromise the effectiveness of hyperspectral image-based FRP. Addressing these challenges, this paper introduces Fruit-HSNet, a machine learning architecture specifically designed for hyperspectral classification of fruit ripeness. Fruit-HSNet incorporates a spatio-spectral feature extraction module based on Fourier Transform and central pixel spectral signature followed by learnable feature fusion and a classifier optimized for ripeness classification. The proposed architecture was evaluated using the DeepHS Fruit dataset, the largest publicly available labeled real-world hyperspectral dataset for predicting fruit ripeness, which includes five different types of fruits—avocado, kiwi, mango, kaki, and papaya—captured with three distinct hyperspectral cameras at various stages of ripeness. Experimental results highlight that Fruit-HSNet substantially outperforms existing deep learning methods, from baseline to state-of-the-art models, with improvements of 12%, achieving a new state-of-the-art overall accuracy of 70.73%.

## 1 INTRODUCTION

In the field of smart agriculture, agricultural computer vision is attracting increasing attention for various applications, from irrigation management to automated classification of agricultural products, enabling automated and simplified agricultural tasks (Ghazal et al., 2024; Luo et al., 2023; Lu and Young, 2020). Fruit Ripeness Prediction (FRP) is an agricultural computer vision task that involves classifying fruits to their degree of ripeness, offering several advantages for both pre-harvest and post-harvest management, including minimizing losses, improving quality, and economizing resources (Rizzo et al., 2023).

Traditionally, FRP has been performed using methods such as visual observation and chemical analysis of the fruit. However, these techniques are subjective, labor-intensive, and costly, involving

a significant margin of error while consuming human and material resources. The emergence of machine/deep learning and imaging technologies, including hyperspectral imaging, has enabled the development of new FRP methods by leveraging the power of learning algorithms to learn hidden patterns. These methods offer advantages over traditional methods, such as the ability to make accurate and timely predictions (Rizzo et al., 2023; Ram et al., 2024).

Hyperspectral imaging (HSI), in particular, unlike conventional imaging techniques, offers the advantage of capturing spatial and spectral information across a wide range of the electromagnetic spectrum, providing details not visible to humans. Formally, a hyperspectral image  $\mathbf{H} \in \mathbb{R}^{M \times N \times B}$  is defined as a three-dimensional data cube with two spatial dimensions,  $M$  and  $N$ , representing spatial information, and one spectral dimension,  $B$ , representing spectral information (i.e., wavelength), encapsulating the reflectance properties of the materials present in the image at different wavelengths. The intensity value of each pixel at spatial coordinates  $(x, y)$  and wavelength

<sup>a</sup>  <https://orcid.org/0009-0006-0333-2643>

<sup>b</sup>  <https://orcid.org/0000-0002-2968-2426>

<sup>c</sup>  <https://orcid.org/0000-0002-0249-7247>

$\lambda$  corresponding to a specific spectral band can be described as  $\mathbf{H}(x, y, \lambda) = r_\lambda$ , where  $r_\lambda$  denotes the spectral response or reflectance at that wavelength. The entire spectral response  $\mathbf{H}(\mathbf{x}, \mathbf{y}, :) = \mathbf{r}$  for a pixel at the spatial coordinates  $(x, y)$  represents the spectral reflectance curve of the object at that location, encompassing its full spectral profile. These spectral reflectance curves are essential for distinguishing materials based on their unique spectral properties, often related to their chemical composition and structure (Ahmad et al., 2022).

Hyperspectral image classification (HIC) has been widely studied in the literature, and various methods have been proposed (Kumar et al., 2024; Ahmad et al., 2024; Ahmad et al., 2022). These range from traditional machine learning methods such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and dimensionality reduction techniques, to deep learning methods based on convolution and attention techniques. However, the development of these methods is application-specific aware, which limits their generalizability. This makes the adaptation of HIC to new applications a significant challenge. Indeed, while state-of-the-art methods for HIC have shown impressive results in certain applications, they fail to maintain comparable performance across different applications (Frank et al., 2023).

In this context, several works and datasets have been proposed for hyperspectral image-based FRP (Zhu et al., 2017; Pinto Barrera et al., 2019; Varga et al., 2021; Varga et al., 2023a; Frank et al., 2023; Rizzo et al., 2023). Principally, the DeepHS Fruit dataset (Varga et al., 2021; Varga et al., 2023a) and the DeepHS-Net family of architectures (Varga et al., 2021; Varga et al., 2023a; Varga et al., 2023b) represent the state-of-the-art. The DeepHS Fruit dataset is the largest commonly available real-world hyperspectral dataset for FRP, distinguished by its variety in the number of fruits, types of hyperspectral cameras used, and stages of maturity. The DeepHS-Net family of architectures, a convolution-based deep learning methods, is specifically designed for hyperspectral classification of fruit ripeness. This family includes two principal convolutional neural network (CNN) architectures: (1) *DeepHS-Net* (Varga et al., 2021), which uses depthwise separable 2D convolutions, and (2) *DeepHS-Hybrid-Net* (Varga et al., 2023a), which combines 2D and 3D depthwise separable convolutions. Additional variants derived from these two architectures incorporate HyveConv (Hyperspectral Visual Embedding Convolution), a wavelength-aware 2D convolution (Varga et al., 2023b). HyveConv employs a continuous representation of convolution kernels, sampling these

kernels based on the wavelengths of the inputs. This design makes the convolution independent of the camera type used and efficiently reduces the number of parameters.

Despite the performance demonstrated by DeepHS-Net architectures in FRP, significant challenges remain unresolved. The lack of robust methods that can generalize across different hyperspectral cameras and fruit types, with the limited size of datasets, compromises the effectiveness of hyperspectral image-based FRP. In response to these limitations, this paper introduces Fruit-HSNet, an architecture specifically designed for FRP from hyperspectral images. The main objective is to ensure consistent and accurate classification across different hyperspectral cameras, fruit types, and stages of ripeness.

**Contributions.** Our key contributions are summarized as follows: (1) We propose Fruit-HSNet, a new architecture specifically designed for fruit hyperspectral image classification to identify different stages of fruit ripeness by leveraging spatio-spectral descriptors, which include Fourier Transform-based features, central pixel spectral signatures, and learnable feature fusion. (2) We conducted comprehensive evaluations on the DeepHS Fruit dataset, the largest publicly available labeled hyperspectral dataset for fruit maturity prediction, which includes five different types of fruits—avocado, kiwi, mango, kaki, and papaya—captured using three distinct hyperspectral cameras. (3) We demonstrate that Fruit-HSNet achieved a new state-of-the-art overall accuracy of 70.73% on the DeepHS Fruit benchmark dataset, which is a 12% improvement over the previous best-reported results, specifically in the challenging categories of avocados and kiwis, which are critical due to their ripening processes.

**Paper Organization.** In the following, Section 2 details our methodology, including an exploratory analysis of the DeepHS Fruit dataset and the introduction of the Fruit-HSNet architecture. Subsequent sections evaluate the model’s performance, analyze the results, and discuss conclusions along with future research directions.

## 2 METHODOLOGY

### 2.1 DeepHS Fruit Dataset

DeepHS Fruit dataset (Varga et al., 2021; Varga et al., 2023a) is the largest publicly available real-world hyperspectral dataset labeled for fruit ripeness predic-

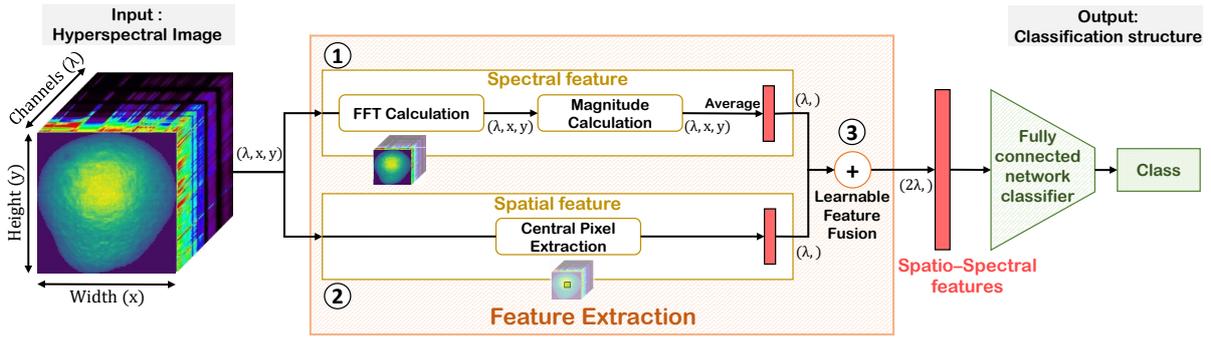


Figure 1: Illustration of Fruit-HSNet architecture.

tion. This dataset includes hyperspectral images of five different fruit types, captured by three distinct hyperspectral cameras, and categorized according to various ripeness level.

**Dataset Composition.** The DeepHS Fruit dataset consists of 30 configurations, where each configuration, denoted as  $config_i$ , corresponds to the dataset for a specific fruit  $fruit_i$  within the category  $category_i$ , captured with the camera  $camera_i$ . The fruits ( $fruit_i$ ) included are avocado, kiwi, mango, kaki, and papaya. Fruit ripeness in this dataset is classified into three distinct categories ( $category_i$ ): ripeness, firmness, and sweetness. The cameras ( $camera_i$ ) used are the Specim FX 10, Corning microHSI 410, and Innospec Redeye, with the following exceptions:

- No sweetness category for avocados.
- No records captured with Innospec Redeye camera for mango, kaki, and papaya.
- No records captured with Corning microHSI camera for kiwi.

**Class Labels.** For each  $config_i$ , three classes are per category, defined as follows:

- **Ripeness:** unripe, ripe, overripe.
- **Firmness:** too firm, perfect, too soft.
- **Sweetness:** not sweet, sweet, overly sweet.

**Data Collection.** In addition to the hyperspectral image label, metadata are available including type of fruit, orientation (front or back), capturing camera, and wavelengths of the recorded spectra. The cameras vary in their spectral band capture:

- *Camera 1: Specim FX 10* captures 224 spectral bands with a wavelength range of 400 to 1000 nm.
- *Camera 2: Corning microHSI 410* captures 249 spectral bands with a wavelength range of 920 to 1730 nm.
- *Camera 3: Innospec Redeye* captures 252 spectral bands, also spanning 920 to 1730 nm.

**Dataset Distribution.** DeepHS Fruit dataset comprises a total of 2706 labeled hyperspectral images, distributed among the fruits as follows: 461 images for Avocado, 568 images for Kiwi, 336 images for Mango, 336 images for Kaki, and 252 images for Papaya.

## 2.2 Fruit-HSNet: Proposed Method

This section introduces the architecture of Fruit-HSNet and its working principle for fruit ripeness classification. Let  $\mathbf{H} \in \mathbb{R}^{M \times N \times B}$  represent a hyperspectral image, the input of the Fruit-HSNet architecture, where  $B$  is the number of spectral channels, and  $M$  and  $N$  are the height and width of the image, respectively. Fruit-HSNet extracts both spectral and spatial features from  $\mathbf{H}$ , which are informative and discriminative, via a feature extraction module and then classifies them via a classification module based on a fully connected neural network.

Feature extraction is performed through a dual-branch approach: (1) Spectral Feature Extraction, (2) Spatial Feature Extraction, followed by (3) Learnable Feature Fusion.

**(1) Spectral Feature Extraction Module.** In this branch, a Fourier Transform (FT) is applied to  $\mathbf{H}$  to transform the spatial information  $\mathbf{H}(:, :, \lambda)$  of each spectral channel  $\lambda$  into the frequency domain:

$$\mathbf{F}_\lambda = \mathbf{FT}(\mathbf{H}(:, :, \lambda)), \quad 1 \leq \lambda \leq B \quad (1)$$

After applying FT across the spatial dimensions, the magnitudes of  $\mathbf{F}_\lambda$  are computed and averaged over each channel  $\lambda$  to form  $\mathbf{m}$ . This frequency transformation makes the periodic patterns of textures and structural changes in the fruit skin more discernible. Indeed, changes in the fruit during ripening affect its spectral signature. By transforming features into the frequency domain, Fruit-HSNet can effectively capture patterns associated with various stages of fruit ripeness that might be less evident in the spatial domain.

**(2) Spatial Feature Extraction Module.** In this branch, the central pixel spectral signature,  $\mathbf{s} = [s_\lambda, 1 \leq \lambda \leq B]$ , is extracted, where

$$s_\lambda = \mathbf{H} \left( \frac{M}{2}, \frac{N}{2}, \lambda \right), \quad 1 \leq \lambda \leq B \quad (2)$$

This allows focusing on potentially the most chemically informative region of the fruit, which is generally indicative of its overall ripeness.

**(3) Learnable Feature Fusion.** Features output from the two branches are adaptively weighted by learnable parameters  $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^B$ . The weighted features are then concatenated to form a feature vector  $\mathbf{f}$  that combines both spectral and spatial information.

$$\mathbf{f} = \mathbf{w}_1 \mathbf{m} + \mathbf{w}_2 \mathbf{s} \quad (3)$$

The introduction of learnable weights for each feature type allows Fruit-HSNet to adaptively prioritize which type of feature (spectral or spatial) is more informative based on their relevance to fruit ripeness. The adaptability provided by these weights enables the model to be applied effectively across different types of fruits, varying stages of ripeness, and different camera types, where the importance of spectral versus spatial information may differ.

## 3 EXPERIMENTS AND RESULTS

### 3.1 Experimental Setup

Experiments were conducted using 30 datasets from the DeepHS Fruit dataset (Varga et al., 2021; Varga et al., 2023a) as described in Section 2.1, adhering to the standard data splitting and preprocessing procedures outlined in (Frank et al., 2023) to ensure result comparability.

Fruit-HSNet was trained on each dataset with the cross-entropy loss function, using a batch size of 16. Optimization involved the use of the Adam algorithm, initiated with a learning rate of 0.001 alongside a weight decay factor of  $1 \times 10^{-4}$ . Additionally, a learning rate scheduler reduced the rate by a factor of 0.7 every 10 epochs to fine-tune the training process. Furthermore, a variable number of epochs is determined experimentally for each configuration to avoid overfitting.

Fruit-HSNet’s fusion feature module uses learnable weights, initialized using a normal distribution. In the classification module, the fully connected neural network comprises two linear layers with dimensions  $[2\lambda, 512, 256]$ , where  $\lambda \in \{224, 249, 252\}$ . Following these layers, batch normalization is applied,

Table 1: Comparative performance of Fruit-HSNet and state-of-the-art methods across all fruits, hyperspectral cameras, and stages of ripeness on DeepHS Fruit dataset (Varga et al., 2021; Varga et al., 2023a). \*Results were obtained from (Frank et al., 2023).

Method	Accuracy
<i>Convolution-based Methods</i>	
2D CNN (spatial) (Paoletti et al., 2019)	44.85 %
ResNet-152 (He et al., 2016)	47.00 %
HybridSN (Roy et al., 2020)	48.74 %
ResNet-18 (He et al., 2016)	49.05 %
SpectralNET (Chakraborty and Trehan, 2021)	49.25 %
2D CNN (spectral) (Frank et al., 2023)	49.27 %
1D CNN (Paoletti et al., 2019)	51.30 %
Gabor CNN (Ghamisi et al., 2018)	52.57 %
EMP CNN (Ghamisi et al., 2018)	52.76 %
2D CNN (Paoletti et al., 2019)	54.42 %
3D CNN (Paoletti et al., 2019)	56.06 %
<i>Attention/Transformer-based Methods</i>	
SpectralFormer (Hong et al., 2022)	41.71 %
Attention-based CNN (Lorenzo et al., 2020)	44.88 %
HiT (Yang et al., 2022)	48.16 %
<i>DeepHS-Net Family</i>	
DeepHS-Hybrid-Net (Varga et al., 2023a)	55.01 %
DeepHS-Net+HyveConv (Varga et al., 2023b)	57.57 %
DeepHS-Net (Varga et al., 2021)	58.28 %
<b>Fruit-HSNet (Ours)</b>	<b>70.73 %</b>

along with ReLU (Rectified Linear Unit) activation functions and dropout layers, implemented with a dropout rate of 0.4.

### 3.2 Comparison with State-of-the-Art

This section presents a comparative performance evaluation of Fruit-HSNet for hyperspectral classification of fruit ripeness. The evaluation covers specific aspects of performance, starting with a general evaluation across all fruits, hyperspectral cameras, and stages of ripeness in Section 3.2.1. It is followed by detailed evaluations focusing on fruit-specific and camera-specific variations in Sections 3.2.2 and 3.2.3, respectively. The section concludes with an in-depth performance analysis for two critical case studies: avocados and kiwis.

#### 3.2.1 Global Performance Evaluation

In this part of the evaluation, we compared Fruit-HSNet across all fruits, hyperspectral cameras, and stages of ripeness against baseline and existing state-of-the-art methods. Baseline methods encompass deep learning models for hyperspectral image classification, ranging from convolutional to attention mechanisms, including 2D and 3D CNN variants, and adapted transformer architectures. The state-of-the-art models are represented by the DeepHS-Net family

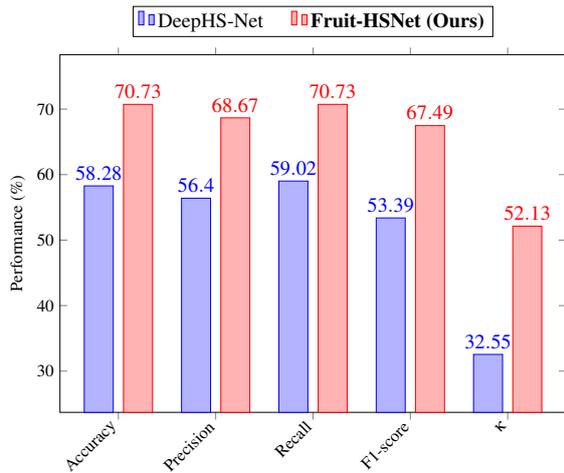


Figure 2: Performance metrics of Fruit-HSNet across all fruits, hyperspectral cameras, and stages of ripeness.

of convolution-based methods.

In Table 1, we report the overall classification accuracy for each method. This accuracy metric is calculated as the average across 30 diverse datasets, each representing a unique combination of fruit type, camera type, and ripeness category, thus providing a robust measure of model generalizability and effectiveness. Subsequently, the DeepHS-Net method is used for detailed benchmarking as it is the best-performing competitor. Furthermore, to provide a detailed analysis of performance, various classification metrics, including accuracy, precision, recall, F1-score, and Cohen’s Kappa ( $\kappa$ ), were reported in Figure 2 to compare Fruit-HSNet against the DeepHS-Net method.

### 3.2.2 Fruit-Specific Performance Evaluation

In this section, performance was analyzed for each fruit type. An overview of the classification performance was presented in Table 2 and Figure 3a, followed by a detailed analysis for each fruit in each category (ripeness, fruitiness, sweetness) in Table 3.

### 3.2.3 Camera-Specific Performance Evaluation

To assess the robustness of Fruit-HSNet across various hyperspectral cameras, we evaluated its performance by presenting the behavior of classification performance in a global manner in Table 4 and Figure 3b, and in detailed form for each fruit ripeness category in Table 5.

### 3.2.4 Detailed Performance Evaluation for Avocado and Kiwi

As avocados and kiwis have a delicate ripeness cycle, this section details the performance of these fruits

Table 2: Fruit-specific performance comparison of Fruit-HSNet.

		DeepHS-Net	Fruit-HSNet (Ours)
<b>Avocado</b>	Overall Accuracy	77.62%	85.19% ( $\uparrow$ 7.57%)
	Average Accuracy	77.62%	82.91% ( $\uparrow$ 5.29%)
	F1-score	76.22%	84.26% ( $\uparrow$ 8.04%)
	$\kappa$	66.03%	76.52% ( $\uparrow$ 10.49%)
<b>Kiwi</b>	Overall Accuracy	60.11%	71.23% ( $\uparrow$ 11.12%)
	Average Accuracy	60.11%	63.69% ( $\uparrow$ 3.58%)
	F1-score	58.02%	68.21% ( $\uparrow$ 10.19%)
	$\kappa$	36.36%	52.73% ( $\uparrow$ 16.37%)
<b>Mango</b>	Overall Accuracy	42.59%	65.28% ( $\uparrow$ 22.69%)
	Average Accuracy	42.59%	60.83% ( $\uparrow$ 18.24%)
	F1-score	34.80%	63.73% ( $\uparrow$ 28.93%)
	$\kappa$	3.51%	44.86% ( $\uparrow$ 41.35%)
<b>Kaki</b>	Overall Accuracy	51.85%	59.72% ( $\uparrow$ 7.87%)
	Average Accuracy	44.87%	52.13% ( $\uparrow$ 7.26%)
	F1-score	42.03%	51.08% ( $\uparrow$ 9.05%)
	$\kappa$	21.37%	33.83% ( $\uparrow$ 12.46%)
<b>Papaya</b>	Overall Accuracy	62.96%	72.23% ( $\uparrow$ 9.27%)
	Average Accuracy	55.64%	66.30% ( $\uparrow$ 10.66%)
	F1-score	55.90%	70.17% ( $\uparrow$ 14.27%)
	$\kappa$	35.51%	52.69% ( $\uparrow$ 17.18%)

Table 3: Detailed fruit-specific performance comparison of Fruit-HSNet across ripeness, firmness, and sweetness.

		DeepHS-Net	Fruit-HSNet (Ours)
<b>Ripeness</b>	Avocado	77.16%	87.04% ( $\uparrow$ 9.88%)
	Kiwi	57.87%	79.86% ( $\uparrow$ 21.99%)
	Mango	41.67%	54.17% ( $\uparrow$ 12.50%)
	Kaki	45.84%	50.00% ( $\uparrow$ 4.17%)
	Papaya	51.85%	77.78% ( $\uparrow$ 25.93%)
<b>Firmness</b>	Avocado	78.09%	83.33% ( $\uparrow$ 5.25%)
	Kiwi	63.61%	72.47% ( $\uparrow$ 8.86%)
	Mango	43.06%	70.83% ( $\uparrow$ 27.77%)
	Kaki	63.89%	66.67% ( $\uparrow$ 2.77%)
	Papaya	70.37%	77.78% ( $\uparrow$ 7.41%)
<b>Sweetness</b>	Kiwi	58.86%	61.35% ( $\uparrow$ 2.49%)
	Mango	43.06%	70.84% ( $\uparrow$ 27.78%)
	Kaki	45.83%	62.50% ( $\uparrow$ 16.67%)
	Papaya	66.67%	61.12% ( $\downarrow$ 5.55%)

by camera type and by ripeness category. A detailed analysis of the two categories, ripeness and firmness, is presented in Tables 6 and 7, including standard classification metrics (accuracy, F1-score, and  $\kappa$ ).

## 3.3 Ablation Study

In this section, we investigate the influence of key architectural components of Fruit-HSNet on the model’s overall performance.

**Spectral Feature Extraction Module Ablation.** To determine the best discriminative spectral feature extraction module for FRP-based hyperspectral image, we compared the performance of Fruit-HSNet using Fourier Transform versus Wavelet Transform in table 8.

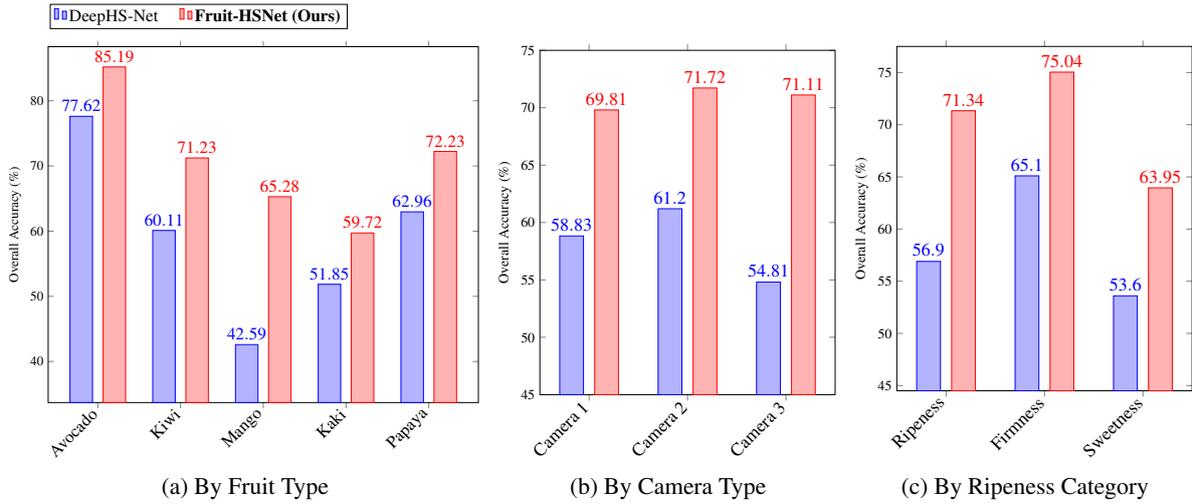


Figure 3: Performance Evaluation of Fruit-HSNet for hyperspectral classification of fruit ripeness.

Table 4: Camera-specific performance comparison of Fruit-HSNet.

		DeepHS-Net	Fruit-HSNet (Ours)
Camera 1	Overall Accuracy	58.83%	69.81% (↑ 10.98%)
	Average Accuracy	52.79%	63.98% (↑ 11.19%)
	F1-score	53.12%	67.12% (↑ 14%)
	$\kappa$	31.42%	51.63% (↑ 20.21%)
Camera 2	Overall Accuracy	61.20%	71.72% (↑ 10.52%)
	Average Accuracy	55.01%	65.50% (↑ 10.49%)
	F1-score	54.92%	67.67% (↑ 12.75%)
	$\kappa$	35.19%	52.09% (↑ 16.9%)
Camera 3	Overall Accuracy	54.81%	71.11% (↑ 16.3%)
	Average Accuracy	52.03%	67.78% (↑ 15.75%)
	F1-score	50.82%	68.15% (↑ 17.33%)
	$\kappa$	29.90%	53.60% (↑ 23.70%)

Table 5: Detailed camera-specific performance comparison of Fruit-HSNet across ripeness, firmness, and sweetness.

		DeepHS-Net	Fruit-HSNet (Ours)
Ripeness	Camera 1	53.70%	61.94% (↑ 8.24%)
	Camera 2	61.57%	74.30% (↑ 12.73%)
	Camera 3	55.55%	88.89% (↑ 33.33%)
Firmness	Camera 1	64.70%	76.76% (↑ 12.06%)
	Camera 2	69.44%	77.08% (↑ 7.64%)
	Camera 3	57.40%	66.67% (↑ 9.26%)
Sweetness	Camera 1	57.90%	70.95% (↑ 13.05%)
	Camera 2	49.69%	61.11% (↑ 11.42%)
	Camera 3	48.15%	44.44% (↓ 3.71%)

**Spatial Feature Extraction Module Ablation.** We studied three different methods for extracting spatial features: (1) central pixel across all spectral bands (the spectral signature of the central pixel), (2) mean of all pixels across all spectral bands (the average spectral signature), (3) variance of all pixels across all spectral bands (the variance in spectral signatures across all pixels) (See Table 8).

**Feature Fusion Ablation.** In this part, we evaluate the performance of concatenating spatial and

Table 6: Detailed performance evaluation for avocado and kiwi (Ripeness categorie).

		DeepHS-Net	Fruit-HSNet (Ours)	
Avocado	Camera 1	Accuracy	83.33%	83.33% (↑ 0%)
		F1-score	82.94%	83.20% (↑ 0.26%)
		$\kappa$	75.00%	75.00% (↑ 0%)
	Camera 2	Accuracy	88.89%	88.89% (↑ 0%)
		F1-score	88.57%	88.57% (↑ 0%)
		$\kappa$	83.33%	83.33% (↑ 0%)
	Camera 3	Accuracy	59.26%	88.89% (↑ 29.63%)
		F1-score	52.17%	88.57% (↑ 36.4%)
		$\kappa$	38.89%	83.33% (↑ 44.44%)
Kiwi	Camera 1	Accuracy	63.89%	70.83% (↑ 6.94%)
		F1-score	64.55%	71.11% (↑ 6.56%)
		$\kappa$	45.83%	56.25% (↑ 10.42%)
	Camera 3	Accuracy	51.85%	88.89% (↑ 37.04%)
		F1-score	47.96%	88.57% (↑ 40.88%)
		$\kappa$	27.78%	83.33% (↑ 55.55%)

Table 7: Detailed performance evaluation for avocado and kiwi (firmness categorie).

		DeepHS-Net	Fruit-HSNet (Ours)	
Avocado	Camera 1	Accuracy	75.00%	83.33% (↑ 8.33%)
		F1-score	78.52%	83.33% (↑ 4.81%)
		$\kappa$	62.90%	70.37% (↑ 7.47%)
	Camera 2	Accuracy	96.30%	100% (↑ 3.7%)
		F1-score	96.70%	100% (↑ 3.3%)
		$\kappa$	94.00%	100% (↑ 6%)
	Camera 3	Accuracy	62.96%	66.67% (↑ 3.71%)
		F1-score	58.43%	61.90% (↑ 3.47%)
		$\kappa$	42.05%	47.06% (↑ 5.01%)
Kiwi	Camera 1	Accuracy	75.36%	78.26% (↑ 2.9%)
		F1-score	75.56%	74.58% (↓ 0.98%)
		$\kappa$	58.35%	60.88% (↑ 2.53%)
	Camera 3	Accuracy	51.85%	66.67% (↑ 14.82%)
		F1-score	50.47%	64.13% (↑ 13.66%)
		$\kappa$	27.78%	50.00% (↑ 22.22%)

spectral features compared to using either spatial or spectral features alone. This aims to demonstrate the added value of integrating spectro-spatial features in enhancing the classification accuracy of Fruit-

Table 8: Feature choices impact on Fruit-HSNet performance. \*calculated across all spectral bands.

Model Variant	Accuracy	Drop in Performance
<b>Fruit-HSNet (Ours)</b>	<b>70.73 %</b>	—
Fruit-HSNet with Wavelet Transform	52.74 %	↓ 17.99%
Fruit-HSNet with average of all pixels*	62.98 %	↓ 7.75%
Fruit-HSNet with variance of all pixels*	59.92 %	↓ 10.81%

Table 9: Impact of introducing feature fusion module and learnable mechanism on Fruit-HSNet performance. \*FE denotes Feature Extraction.

Model Variant	Accuracy	Drop in Performance
<b>Fruit-HSNet (Ours)</b>	<b>70.73 %</b>	—
Fruit-HSNet w/o spectral FE* module	48.60 %	↓ 22.13 %
Fruit-HSNet w/o spatial FE* module	58.40 %	↓ 12.33 %
Fruit-HSNet w/o learnable feature fusion	60.96 %	↓ 9.77 %

HSNet (See Table 9).

**Ablation of Learnable Mechanisms in Feature Fusion.** In this part, we compare the efficiency of concatenating learned features versus a simple concatenation (without learning) of features. The learned fusion aims to intelligently combine features in a way that maximizes the relevant information from each feature extraction block (See Table 9).

## 4 FINDINGS AND ANALYSIS

**How effective is Fruit-HSNet in hyperspectral image-based fruit ripeness prediction, and how does it compare to state-of-the-art methods?** The results presented in Table 1 and Figure 2 demonstrate the effectiveness of Fruit-HSNet for hyperspectral image-based fruit ripeness prediction. Fruit-HSNet produces very promising results despite being trained on a small dataset. The success of our method is due to two factors: the use of (1) informative and discriminative spatial and spectral features for fruit ripeness classification, and (2) a learnable feature fusion mechanism that naturally applies attention to the spatio-spectral features, effectively capturing the most relevant features for the candidate fruit, camera, and/or ripeness stage. We detail and discuss our evaluations below.

Fruit-HSNet clearly outperforms other methods, ranging from baseline to state-of-the-art, with an accuracy of 70.73%. Previous models, which are considered state-of-the-art for this dataset, such as DeepHS-Net, achieve accuracy and F1-score performances of 58.28% and 53.39%, respectively, indicating a significant improvement by Fruit-HSNet: an increase of 12.45% in accuracy and 14.1% in F1-score. This improvement can be considered promis-

ing since the results on this dataset have saturated around the 50% range, as presented in 1.

Attention/transformer-based methods generally show inferior performance compared to convolution-based methods, suggesting that attention-based architectures do not capture the spatial and spectral features of hyperspectral images as effectively for this specific application. This reaffirms the observation that state-of-the-art methods for hyperspectral image classification, which have shown impressive results in certain applications, fail to maintain comparable performance for different applications, emphasizing the need to adapt attention mechanisms for the specificity of this application.

Analyzing the performance based on the Kappa metric further confirms the superior performance of Fruit-HSNet. While DeepHS-Net shows reasonable accuracy, its Kappa metric is relatively low at 32.55%. Fruit-HSNet, on the other hand, achieves a Kappa of 52.13%, representing an improvement of 19.58%, thereby enhancing the reliability and consistency of predictions in scenarios with imbalanced class distributions, demonstrating the robustness of Fruit-HSNet to variations in input data.

**What is the performance of Fruit-HSNet specific to each type of fruit?** So far, we have assessed the performance across all fruits, hyperspectral cameras, and stages of ripeness. We are now analyzing the performance of Fruit-HSNet by fruit type. Table 2 and Figure 3a show that Fruit-HSNet consistently outperforms DeepHS-Net across all fruit types (Avocado, Kiwi, Mango, Kaki, Papaya) in terms of all metrics including Overall Accuracy, Average Accuracy, F1-score, and Cohen’s kappa coefficient. The improvement margins in Fruit-HSNet over DeepHS-Net are significant.

According to the F1-score, which combines precision and recall, Fruit-HSNet shows an improvement

in predicting avocado ripeness: an increase of 8.04%. For the Kiwi fruit, Fruit-HSNet demonstrates improvements across all metrics as well, with a significant increase in the kappa metric by 16.37%, indicating a more reliable model performance. For Mango, the least predictable fruit in terms of performance for DeepHS-Net, with an overall accuracy of 42.59% and a kappa of 3.51%, we note spectacular improvements with Fruit-HSNet, showing an improvement of 22.69% and 41.35% respectively, indicating increased sensitivity of Fruit-HSNet to the spectral characteristics of this particular fruit, and validating the hypothesis of the generalizability of our method across various fruit types. Both Kaki and Papaya show significant improvements, particularly in the F1 and kappa scores, indicating a better balance between precision and recall.

In now analyzing the performance of each fruit by stage of ripeness, Table 3 shows that Fruit-HSNet is not only globally more accurate but also better at capturing specific quality attributes of each fruit: For ripeness, we note significant improvements for all fruits, with substantial enhancements for Kiwi and Mango. This indicates a better feature extraction capability of Fruit-HSNet to discern spectral signatures related to ripeness. Regarding firmness, all fruits show performance improvement. In particular, the performance of Kiwi and Mango has significantly increased, due to the increased sensitivity of Fruit-HSNet to repetitive textural attributes detectable by the spatial feature extraction module based on Fourier transform. Sweetness, a more subtle and complex attribute to capture spectrally, also shows improvements, especially in Mango and Kaki.

**How does the performance of Fruit-HSNet vary with different hyperspectral cameras?** The performance of Fruit-HSNet was evaluated using three different hyperspectral cameras, each with unique spectral sensitivities. The results, detailed in Tables 4 and 5, reveal significant variations in model efficiency depending on the camera used, highlighting the crucial impact of imaging hardware on the task of hyperspectral image-based fruit ripeness prediction.

Camera 1 (Specim FX 10) operates in the visible to near-infrared (VNIR) range. With this camera, Fruit-HSNet achieved an overall accuracy of 69.81%, an improvement of 10.98% over DeepHS-Net. The average accuracy increased by 11.19% to 63.98%, and the F1-score increased by 14% to 67.12%. The Kappa coefficient improved by 20.21% to 51.63%. These improvements indicate that the spectral information in the VNIR range is effectively used by Fruit-HSNet, enhancing the discrimination of fruit quality attributes such as surface color and certain visible

chemical compounds.

Cameras 2 (Corning microHSI 410) and 3 (Innospec Redeye) both operate in the short-wave infrared (SWIR) range, covering wavelengths from 920 to 1730 nm. Despite similar spectral ranges and number of bands, subtle differences in sensor technology and spectral sensitivity may explain the performance variations.

With Camera 2, Fruit-HSNet achieved an overall accuracy of 71.72%, an improvement of 10.52%, and an average accuracy of 65.50%, increased by 10.49%. The F1-score increased by 12.75% to 67.67%, and the Kappa coefficient improved by 16.9% to 52.09%. The SWIR range captured by Camera 2 is sensitive to the internal qualities of fruits, such as moisture content and structural properties, which are crucial for assessing attributes like firmness and internal composition.

Camera 3 offered the most significant improvements. Fruit-HSNet achieved an overall accuracy of 71.11%, representing the highest increase of 16.3%. The average accuracy improved by 15.75% to 67.78%, and the F1-score increased by 17.33% to 68.15%. The Kappa coefficient markedly increased by 23.7% to 53.60%. Notably, for the ripeness criterion, Camera 3 achieved an exceptional accuracy of 88.89%, an improvement of 33.33%. This suggests that the sensor characteristics of Camera 3 are particularly effective at capturing spectral features associated with fruit maturation processes, such as changes in water absorption bands and alterations in chemical composition.

However, performance variations across different quality criteria underscore the influence of camera characteristics on the task of FRP. For firmness, Cameras 1 and 2 achieved higher accuracies (76.76% and 77.08%, respectively) compared to 66.67% for Camera 3. This implies that the spectral features related to fruit properties that influence firmness are better captured by Cameras 1 and 2. Regarding sweetness, Camera 1 obtained the highest accuracy at 70.95%, showing an improvement of 13.05%. Camera 2 follows with an accuracy of 61.11%, while Camera 3 showed a decrease in performance to 44.44%, indicating a drop of 3.71%. This decrease cannot be decisive regarding Camera 3, as the sweetness with Camera 3 contains only the kiwi fruit.

This confirms that Fruit-HSNet is generalizable across all types of cameras, as it achieves high results.

**How robust is Fruit-HSNet for predicting the ripeness of avocados and kiwis?** In this section, we focus on analyzing the performance of avocados and kiwis given the particular and delicate nature of their ripening process. Tables 6 and 7 show that Fruit-HSNet accurately predicts the ripeness of avo-

cados and kiwis across different hyperspectral cameras. Compared to DeepHS-Net, Fruit-HSNet demonstrates substantial improvements.

For avocados, Fruit-HSNet showed consistent and superior performance across three cameras: (*Camera 1*) Fruit-HSNet improved over DeepHS-Net in terms of F1-score, achieving a similarly high accuracy of 83.33%. (*Camera 2*) Like DeepHS-Net, Fruit-HSNet demonstrated excellent performance, with an accuracy of 88.89%, an F1-score of 88.57%, and a kappa statistic of 83.33%. (*Camera 3*) Fruit-HSNet significantly outperformed DeepHS-Net, with a substantial increase of 29.63% in accuracy, 36.4% in F1-score, and 44.44% in kappa statistic.

For kiwis, the performance of Fruit-HSNet was also superior with two cameras: (*Camera 1*) An increase of 6.94% in accuracy and 6.56% in F1-score. (*Camera 3*) An increase of 37.04% in accuracy and 40.88% in F1-score.

#### **How does the choice of spectral feature extraction module affect the performance of Fruit-HSNet?**

As shown in Table 8, the choice between the Fourier Transform and the Wavelet Transform for spectral feature extraction significantly influences the model's accuracy. The Fourier Transform based feature extraction bloc demonstrated a superior overall accuracy, with an improvement of 17.99% compared to the Wavelet Transform. This suggests that the periodic patterns of textures and structural changes in the fruit skin extracted by the Fourier Transform are more discernible than those captured by the Wavelet Transform, which focuses on local frequency and time information.

**How does the choice of spatial feature extraction module influence the performance of Fruit-HSNet?** Table 8 presents a comparison of different methods for extracting spatial features. Extracting the spectral signature from the central pixel leads to the highest accuracy, which is an improvement of 7.75% and 10.81% over using the mean and variance of pixels per spectral band, respectively. These results underline that the most chemically informative region of the fruit is the center, which is generally indicative of its overall ripeness.

**What is the significance of employing spatio-spectral features in improving Fruit-HSNet performance?** The integration of spatial and spectral features is further validated by Table 9, where the combination of spatio-spectral features surpasses spatial or spectral features with improvements of 22.13% and 12.33%, respectively. This validates the importance of a spatio-spectral approach in the classification of hyperspectral images, as discussed in (Kumar

et al., 2024; Ahmad et al., 2024; Frank et al., 2023; Ahmad et al., 2022).

#### **What impact do learnable mechanisms in feature fusion have on the performance of Fruit-HSNet?**

Table 9 explores the effect of feature fusion with and without learning. Incorporating learning in the feature fusion process led to an improvement of 9.77%. This improvement emphasizes that the introduction of learnable weights for each feature type allows Fruit-HSNet to adaptively prioritize which type of feature (spectral or spatial) is more informative based on their relevance to fruit ripeness.

## **5 CONCLUSION AND FUTURE WORK**

In this paper, we introduce Fruit-HSNet, a novel machine learning architecture specifically designed for hyperspectral image-based fruit ripeness prediction. Fruit-HSNet features a small-simple architecture that integrates spatio-spectral feature extraction based on Fourier transform and the central pixel's spectral signature, followed by learnable feature fusion and a deep fully connected neural network. The experiments conducted on the DeepHS Fruit dataset demonstrated that Fruit-HSNet outperforms existing baselines and state-of-the-art methods across five fruits and three hyperspectral cameras, achieving a new state-of-the-art overall accuracy of 70.73%.

Future work involves continuous improvement of fruit ripeness prediction results through the integration of an attention mechanism to select the best features. Additionally, for real-world applications, future work will focus on integrating Fruit-HSNet into IoT devices and mobile platforms to facilitate real-time ripeness prediction.

## **ACKNOWLEDGEMENTS**

This work was supported by the Agence Universitaire de la Francophonie (AUF) through the IntenSciF program as part of the BIO-Serr (Intelligent Toolbox for Greenhouse Establishment and Monitoring Assistance) Project.

## **REFERENCES**

Ahmad, M., Distifano, S., Khan, A. M., Mazzara, M., Li, C., Yao, J., Li, H., Aryal, J., Vivone, G., and Hong, D. (2024). A comprehensive survey for hyperspectral

- image classification: The evolution from conventional to transformers. *arXiv abs/2404.14955*.
- Ahmad, M., Shabbir, S., Roy, S. K., Hong, D., Wu, X., Yao, J., Khan, A. M., Mazzara, M., Distefano, S., and Chanussot, J. (2022). Hyperspectral image classification—traditional to deep models: A survey for future prospects. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:968–999.
- Chakraborty, T. and Trehan, U. (2021). Spectralnet: Exploring spatial-spectral waveletcnn for hyperspectral image classification. *ArXiv*, abs/2104.00341.
- Frank, H., Varga, L. A., and Zell, A. (2023). Hyperspectral benchmark: Bridging the gap between hsi applications through comprehensive dataset and pretraining. *arXiv preprint arXiv:2309.11122*.
- Ghamisi, P., Maggiori, E., Li, S., Souza, R., Tarablaka, Y., Moser, G., De Giorgi, A., Fang, L., Chen, Y., Chi, M., Serpico, S. B., and Benediktsson, J. A. (2018). New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, markov random fields, segmentation, sparse representation, and deep learning. *IEEE Geoscience and Remote Sensing Magazine*, 6(3):10–43.
- Ghazal, S., Munir, A., and Qureshi, W. S. (2024). Computer vision in smart agriculture and precision farming: Techniques and applications. *Artificial Intelligence in Agriculture*, 13:64–83.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society.
- Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., and Chanussot, J. (2022). Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote. Sens.*, 60:1–15.
- Kumar, V., Singh, R. S., Rambabu, M., and Dua, Y. (2024). Deep learning for hyperspectral image classification: A survey. *Computer Science Review*, 53:100658.
- Lorenzo, P. R., Tulczyjew, L., Marcinkiewicz, M., and Nalepa, J. (2020). Hyperspectral band selection using attention-based convolutional neural networks. *IEEE Access*, 8:42384–42403.
- Lu, Y. and Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture*, 178:105760.
- Luo, J., Li, B., and Leung, C. (2023). A survey of computer vision technologies in urban and controlled-environment agriculture. *ACM Computing Surveys*, 56(5):1–39.
- Paoletti, M. E., Haut, J. M., Plaza, J., and Plaza, A. J. (2019). Deep learning classifiers for hyperspectral imaging: A review. *Isprs Journal of Photogrammetry and Remote Sensing*, 158:279–317.
- Pinto Barrera, J., Rueda-Chacón, H., Arguello, H., and De, A. (2019). Classification of hass avocado (*persea americana* mill) in terms of its ripening via hyperspectral images. *TecnoLógicas*, 22:109–128.
- Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K., and Sun, X. (2024). A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*, 222:109037.
- Rizzo, M., Marcuzzo, M., Zangari, A., Gasparetto, A., and Albarelli, A. (2023). Fruit ripeness classification: A survey. *Artificial Intelligence in Agriculture*, 7:44–57.
- Roy, S. K., Krishna, G., Dubey, S. R., and Chaudhuri, B. B. (2020). Hybridsn: Exploring 3-d-2-d CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote. Sens. Lett.*, 17(2):277–281.
- Varga, L. A., Frank, H., and Zell, A. (2023a). Self-supervised pretraining for hyperspectral classification of fruit ripeness. In *6th International Conference on Optical Characterization of Materials, OCM 2023*, pages 97–108.
- Varga, L. A., Makowski, J., and Zell, A. (2021). Measuring the ripeness of fruit with hyperspectral imaging and deep learning. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Varga, L. A., Messmer, M., Benbarka, N., and Zell, A. (2023b). Wavelength-aware 2d convolutions for hyperspectral imaging. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3777–3786.
- Yang, X., Cao, W., Lu, Y., and Zhou, Y. (2022). Hyperspectral image transformer classification networks. *IEEE Trans. Geosci. Remote. Sens.*, 60:1–15.
- Zhu, H., Chu, B., Fan, Y., Tao, X., Yin, W., and He, Y. (2017). Hyperspectral imaging for predicting the internal quality of kiwifruits based on variable selection algorithms and chemometric models. *Scientific Reports*, 7.