Smart Agricultural Technology 9 (2024) 100644

Contents lists available at ScienceDirect

Smart Agricultural Technology

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

Non-invasive detection of pesticide residues in freshly harvested olives using hyperspectral imaging technology

Diego Manuel Martínez Gila [a,b,*], David Bonillo Martínez [a], Silvia Satorres Martínez [a,b], Pablo Cano Marchal [a,b], Javier Gámez García [a,b]

[a] Robotics, Automation and Computer Vision Group, Electronic and Automation Engineering Department, University of Jaén, Campus Las Lagunillas s/n, Jaén, 23071, Spain

[b] Institute for Olive Orchards and Olive Oils, University of Jaén, Campus Las Lagunillas s/n, Jaén, 23071, Spain

ARTICLE INFO

Keywords:
Freshly harvested olives
Computer vision
Image processing
Hyperspectral imaging
Pesticide detection
In-line process monitoring

ABSTRACT

Pesticides play a crucial role in boosting the overall yield and productivity of agricultural produce by controlling pests, insects, and various plant diseases. However, excessive use of pesticides has led to contamination of food products and water bodies, as well as disruption of ecological and environmental systems. Global health authorities have set limits for pesticide residues in individual food items to ensure the availability of safe foods in the supply chain and to assist farmers in developing optimal agronomic practices for crop production. In Spain, specifically regarding olive cultivation, the Ministry of Agriculture, Fisheries, and Food establishes a safety period that farmers must observe from the application of the pesticide until the fruit is harvested. This period ensures that the batch of olives will comply with the maximum residue level allowed. This article proposes a methodology based on hyperspectral imaging to detect whether the olives have been sprayed with pesticide products and, if so, when the spraying occurred. The proposed methodology operates at the pixel level, where each pixel of the hyperspectral image is an instance. The pesticides evaluated were Diflufenican, Oxyfluorfen, Deltamethrin, λ-Cyhalothrin, and Tebuconazole. The results are promising and the success rates achieved over 80% accuracy for most pesticides in controlled laboratory conditions, with individual performance varying according to each pesticide's chemical properties and stability on the olive surface. While the results are promising, the scalability of this approach for larger and more diverse batches of olives requires validation under field conditions, where variations in environmental factors, olive variety, and ripeness may impact the detection accuracy. Furthermore, the study highlights key wavelengths around 750 nm and 550 nm as effective discriminators, suggesting potential for cost-effective, simplified imaging systems. Although hyperspectral imaging shows potential as an accessible, in-line monitoring solution for cooperative use, further analysis of implementation costs is recommended to confirm its feasibility on an industrial scale.

1. Introduction

According to the Food and Agriculture Organization of the United Nations (FAO), around three million tons of pesticides are used annually in the agricultural sector. Without the usage of pesticides the production loss on fruits would be 78%, 54% over vegetables and 32% considering cereals [33], as well as averting fruit degradation post-harvest due to pathogen infections [23]. Within the olive sector, the consequences of decreased crop yields should be added to the expected outcomes of climate change, with reductions in rainfed olive production ranging from 17% to 20% for southern Iberian Peninsula regions [10]. Furthermore, these new climatic conditions will alter the ways in which pests attack the olive groves, in a less predictable manner due to the complexity of interactions within the ecosystem and the diversity of olive crop varieties [15].

Within the olive sector, the use of pesticides varies depending on factors such as geographical location, climatic conditions, specific pests and agricultural practices in the region [29]. Even though the use of pesticides is beneficial for increasing crop yields, their use presents numerous disadvantages for life in general. At the review published by de Souza et al. [28] it is widely proved that pesticides, even though applied to soil or trees, reach surface water bodies and groundwater

* Corresponding author.
E-mail address: dmgila@ujaen.es (D.M. Martínez Gila).

https://doi.org/10.1016/j.atech.2024.100644
Received 17 September 2024; Received in revised form 30 October 2024; Accepted 7 November 2024
Available online 12 November 2024

Table 1
Required volumes of each product to prepare successive dilutions in 500 ml. Note: Dilutions #2 and #3 are prepared from concentration #1, and dilutions #4 and #5 are prepared from concentration #3.

Product	#1 1% Dilution (ml in 500 ml)	#2 0.1% Dilution (ml of #1 in 500 ml)	#3 0.01% Dilution (ml of #1 in 500 ml)	#4 0.0001% Dilution (ml of #3 in 500 ml)	#5 0.00001% Dilution (ml of #3 in 500 ml)
Diflufenican	10.00	50.00	5.00	5.00	0.50
Oxyfluorfen	20.83	50.00	5.00	5.00	0.50
Deltamethrin	200.00	50.00	5.00	5.00	0.50
λ-Cyhalothrin	333.33	50.00	5.00	5.00	0.50
Tebuconazole	20.00	50.00	5.00	5.00	0.50

through surface runoff, percolation, evaporation and precipitation. The widespread impact of pesticides extends to microflora, microfauna, and plants, integrating these chemicals into the food chain and posing risks to overall ecosystem [30] and human health [6,26,32,5]. Given these environmental and health risks, strict regulations have been implemented to limit pesticide residues in food products, including olives. Recently, concerns have also emerged regarding the presence of mineral-origin hydrocarbons (MOH) in olive oil, a potential byproduct of pesticide contamination. Thus, regulating pesticide residues in harvested olives is critical to ensure consumer safety and environmental protection.

Considering the harmful effects of pesticides on consumers' health regulations must be established to control the presence of pesticides in the harvested olive batches [11]. In this regard, the European Union and the Codex Alimentarius Commission of FAO have established maximum residue limit (MRL) for each pesticide in olive oil to ensure consumer safety [21,16]. These MRLs represent the highest level of pesticide residue legally allowed in food products. In Spain, the Ministry of Agriculture, Fisheries, and Food defines the concept of the safety period, which is the minimum time that must pass between the application of the pesticide in the olive grove and the harvesting of the fruit. After this period, they ensure compliance with the MRL in the oil, but to our knowledge, there is no tool that can be applied to the fruit to quickly verify at the entrance of the cooperative if the safety period has been met.

Pesticide detection techniques can be divided into two types, invasive and non-invasive. The former are considered off-line techniques, meaning they require extracting the sample from the process and transporting it to an analytical laboratory where it is prepared and processed using specific instrumentation. Examples of these techniques include gas chromatography (GC) [7] and liquid chromatography (LC) [14]. The latter do not require sample processing and, in principle, could be adapted for use directly in the production line (in-line).

Among the non-invasive techniques, taking as a reference Sindhu and Manickavasagan [27], we can name Raman spectroscopy, fast in detection times but difficult to operate in real time due to the need to take samples by contact with the surface of the fruit to be inspected [34,17,25]; photoacoustic spectroscopy, which is difficult to operate at industry level [19]; electronic nose, complex to operate for obtaining repeatable results due to the complexity of headspace formation [22,12]; fluorescent spectroscopy, with high selectivity and sensitivity but again difficult to introduce in an in-line process. Also it can be affected by the emission properties of other elements present [31]. Given these limitations, hyperspectral imaging in the near-infrared spectrum is proposed in this study as it offers a rapid, non-contact, and cost-effective alternative suitable for real-time, in-line monitoring of pesticide residues. This approach not only enables direct detection of pesticide residues on freshly harvested olives but also addresses the need for a scalable, industrially viable solution that current techniques fail to meet.

The aim of this work is to assess the feasibility of using hyperspectral images for the detection of different pesticides commonly used in olive groves and also to identify the most useful wavelengths for this issue. Those pesticides are Diflufenican, Oxyfluorfen, Deltamethrin, λ-Cyhalothrin and Tebuconazole. Diflufenican and Oxyfluorfen are two herbicidal compounds. They are commonly used in agriculture to control weeds and unwanted plants in crops. Both interfere with plant photosynthesis. On the other hand, Deltamethrin and λ-Cyhalothrin are both synthetic pyrethroid insecticides belonging to the pyrethroid family. They are widely used in agriculture, as well as in crop and garden protection to control insects, including mosquitoes, flies, fleas, ticks, and other pests. Finally, Tebuconazole is a fungicide belonging to the group of triazoles, which is a class of chemical compounds used to control pathogenic fungi in plants. This fungicide is widely used in agriculture to protect various crops, such as cereals, fruits, vegetables and ornamental crops, against fungal diseases.

The following Section 2 will present the materials and methods employed, detailing the experimental setup used, the set of olive samples prepared and analyzed, and the methodology employed for classification according to MRL. Subsequently, Section 3 will detail the results obtained after training and validating different classification algorithms. Finally, Section 4 shows the conclusions of this research work.

2. Materials and methods

2.1. Experimental samples

The olives selected for the experimentation were harvested from a Picual variety olive grove located in the province of Jaén (Spain). To ensure that the olives were initially free of pesticides, an organic olive grove was chosen. The olive harvesting method was by vibration, and a number of olives were selected, collectively weighing a total of 20 kg.

To maximize real-world applicability, pesticides were used at concentrations commonly found in agricultural practices. These concentrations, specified as percentage weight/volume (p/v), were as follows: Diflufenican 50% p/v, Oxyfluorfen 24% p/v, Deltamethrin 2.5% p/v, λ-Cyhalothrin 1.5% p/v and Tebuconazole 25% p/v. These solutes were diluted with 500 ml of distilled water as indicated in Table 1. A total of 25 dilutions were created (5 for each pesticide).

To prepare samples of contaminated olives, the initially collected 20 kg of olives were divided into 26 subgroups of approximately 650 grams each. One of these subgroups served as the control sample, while each of the remaining 25 subgroups was treated with one of the 25 pesticide dilutions prepared. For each contaminated sample, the olives were spread out evenly on a plastic base, and the corresponding dilution was applied using a spray nozzle to ensure uniform coverage. This spraying process ensured that all olives within each sample were thoroughly coated with the pesticide solution. From each treated sample, 10 olives were randomly selected to be inspected using the experimental setup configured by the authors' research group, while the remaining olives were sent to an accredited external oil analysis laboratory. In this external laboratory, a standardized destructive multiresidue detection method based on Gas Chromatography coupled with Mass Spectrometry (GC-MS/MS) was applied to certify the contamination level of each sample. This methodology is detailed in report number 2012-M6 from the European Union Reference Laboratory for pesticide residues in Fruits and Vegetables (EURL-FV). The results obtained from the accredited laboratory are used to train and validate the classification models documented

Table 2
Results obtained from the analyses using GC-MS/MS. All values are expressed in mg/Kg. The symbol "-" indicates that no analyses were performed for that concentration due to logistical issues; however, this does not affect the reliability of the results, as these dilutions would yield concentrations close to or below the MRL. To compensate for these missing samples, uncontaminated control samples were used. The limits of detection (LOD) for the method is 0.01 mg/Kg. The EU MRL column indicates the maximum residue limit authorized by the European Union. Samples above this limit are marked with a double asterisk (**), while those below it are marked with a single asterisk (*).

	EU MRL	#1	#2	#3	#4	#5
Diflufenican	0.60	31.80**	3.42**	0.46*	0.01*	<0.01*
Oxyfluorfen	1.00	72.30**	11.50**	1.12**	-	<0.01*
Deltamethrin	0.60	37.20**	19.10**	0.77**	-	<0.01*
λ-cyhalothrin	0.50	84.90**	11.40**	0.95**	0.01*	<0.01*
Tebuconazole	0.50	116.00**	19.20**	1.29**	<0.01*	<0.01*

Fig. 1. The images in the figures show the hyperspectral image acquisition system used in the experimentation. They depict the hyperspectral camera, halogen spotlights, the servomotor, and a sample olive.

in Subsection 2.4. Table 2 shows the results obtained using the referenced methodology.

2.2. Experimental setup and image acquisition

The hyperspectral camera used for the experimentation was the PIKA L model from the company Resonon. It is a linear camera sensitive in the spectral range between 400 and 1000 nm (VNIR), which is relevant for pesticide detection due to its capacity to capture both visible and near-infrared light. This range has been shown to correlate with pesticide residues [20]. The camera offers a spatial resolution of 900 pixels per line and a spectral resolution of 300 channels per line. The lighting system was comprised of two 100 W halogen spotlights (Fig. 1).

The camera lens was positioned 20 centimeters away from the olive, and the acquisition methodology was pushbroom type, meaning that the camera captures images line-by-line as the object moves relative to the sensor. In this setup, the olive was mounted on the axis of an Arduino-controlled servo motor, which rotated the fruit while the camera captured each line of the image sequentially. The image acquisition rate was 30 fps, and 750 images were acquired per olive to ensure spectral information is captured from the entire surface envelope of the fruit. The inspection time for each olive was 25 seconds, resulting in a 3D data cube (hypercube) of dimensions 300x900x750. In this cube, the 300 channels represent spectral information across different wavelengths, the 900 spatial pixels capture the width of each line scanned across the olive, and the 750 frames correspond to the sequential lines acquired as the olive rotates. This structure enables us to capture comprehensive spectral and spatial data across the entire surface of the olive.

2.3. Image processing and feature extraction

Once acquired, each hyperspectral image was standardized according to Eq. (1).

$$I_s^{std}(w,n,m) = \frac{I_s^{raw}(w,n,m) - B(w,n,m)}{W(w,n,m) - B(w,n,m)} \quad (1)$$

where $I_s^{std}(w,n,m)$ is the corrected hyperspectral image, $I_s^{raw}(w,n,m)$ is the original hyperspectral image, $B(w,n,m)$ is s the black image obtained by closing the iris of the camera (approximately 0% reflectance), and $W(w,n,m)$ is the white image obtained from a white sheet (approximately 99% reflectance). In addition, w is the number of wavelength (from 1 to 300), n and m represent the number of pixels in the image (n ranges from 0 to 900, and m ranges from 0 to 750), and s is the number of sample (Fig. 2).

From each hyperspectral image, a square region of interest of size 10x10 pixels was selected, with its center coinciding with the central point of the image. Additionally, it was verified that the selected region was free of glares or bruises. From these regions, the feature matrices $X_S^{P,C}(px,w)$ were formed where each cell contains the absorbance value of the olive for wavelength w and pixel px. Each row of the matrix will henceforth be referred to as an instance. The subscript S indicates the sample number ($S \in \{1,2,...,10\}$), P the pesticide with which the sample has been contaminated ($P \in \{difl, oxyf, delt, cyha, tebu, ctrl\}$), and C its concentration ($C \in \{\#1,\#2,\#3,\#4,\#5\}$) according to Table 1.

2.4. Classification algorithms

To assess the feasibility of using hyperspectral technology for pesticide detection, each pesticide was considered as an independent case study, and in each case, different classification algorithms were evaluated using the feature matrices constructed in the previous section as the data source.

The classification algorithms evaluated were Decision Tree (DT) [2], Linear Discriminant (LD) [8], Bayesian Classifier (BC) [24], Support Vector Machine with Quadratic (SVM-Q) and Cubic (SVM-C) Kernels [3], and K-Nearest Neighbors (KNN) [4]. These algorithms were cho-

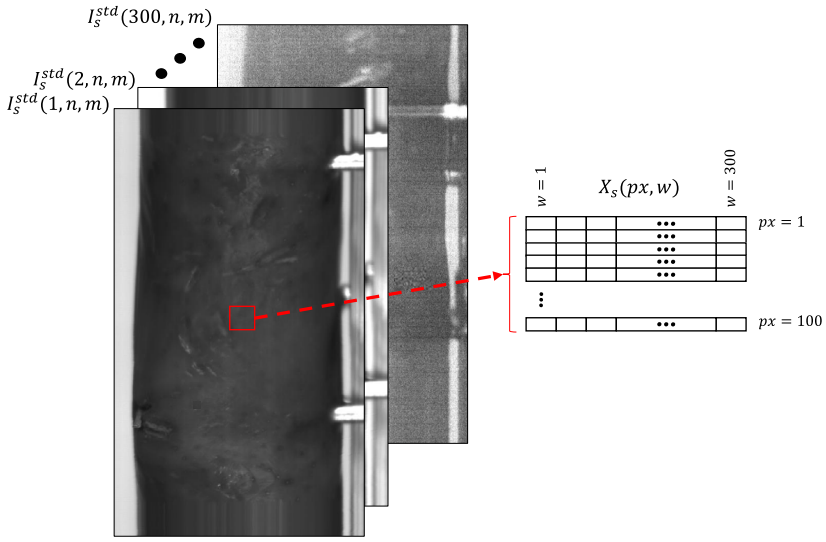


Fig. 2. Procedure for extracting the feature matrix X_s from the hyperspectral image I_s^{std}.

sen for their effectiveness in previous spectral data classification studies and their ability to handle high-dimensional data typical of hyperspectral imaging [27]. The selection includes both linear and non-linear classifiers, providing a comprehensive approach to assess which model best suits pesticide residue detection. All algorithms are included in the Statistics and Machine Learning Toolbox v11.7 of Matlab R2020a.

Each classification algorithm was fine-tuned (model training) using 50% of the olive samples and then used (model validation) to predict whether the remaining 50% were contaminated with the pesticide under study, ensuring a balanced distribution of classes in both the training and validation sets. As a result of this validation process, a confusion matrix was obtained for each classification model. In these matrices, True Positives (TP) represent samples correctly predicted as contaminated, meaning the predicted class is above the maximum residue limit (MLR) and the actual class is also above the MLR. True Negatives (TN) refer to samples accurately classified as non-contaminated, with both the predicted and actual classes below the MLR. False Positives (FP) are samples incorrectly identified as contaminated; in these cases, the predicted class is above the MLR, but the actual class is below the MLR. False Negatives (FN) are samples incorrectly classified as non-contaminated, where the predicted class is below the MLR while the actual class is above the MLR. This setup allowed us to calculate evaluation metrics such as precision (Eq. (2)), recall (Eq. (3)), accuracy (Eq. (4)), F1-score (Eq. (5)) and the area under the curve (AUC) for each classifier, providing a comprehensive assessment of each model's performance.

- Precision: Measures the proportion of correctly predicted positive observations to the total predicted positive observations. A high precision (closer to 100%) indicates that the model has a low rate of false positives, meaning it is effective at predicting only the true positive cases.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \tag{2}$$

- Recall: Measures the proportion of correctly predicted positive observations to all actual positive observations. A high recall (closer to 100%) suggests that the model successfully captures most true positive cases, indicating a low rate of false negatives.

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \tag{3}$$

- Accuracy: Measures the proportion of correctly predicted observations (both positive and negative) to the total observations. Accuracy closer to 100% implies that the model is making very few errors overall, effectively predicting both positive and negative cases.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{4}$$

- F1-Score: The harmonic mean of Precision and Recall, providing a balanced metric between the two. An F1-score closer to 100% indicates that the model balances precision and recall effectively, making it suitable for cases where both false positives and false negatives need to be minimized.

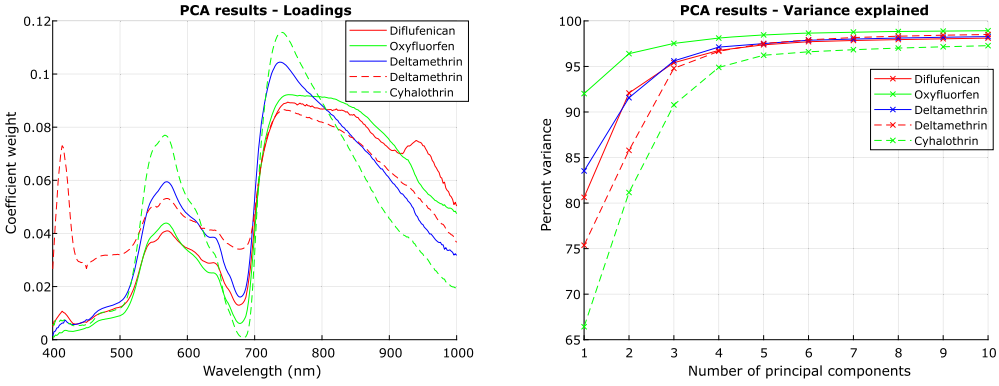
$$F1\text{-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \tag{5}$$

- Area Under the Curve (AUC): It represents the area under the Receiver Operating Characteristic (ROC) curve, calculated based on sensitivity and specificity at various thresholds. AUC values closer to 100% mean the model is highly capable of distinguishing between positive and negative classes across different thresholds.

3. Results and discussion

All algorithms solve a binary classification problem, where one class was assigned to olives that showed a residue level above the standard, and a different class to olives with residue levels below the standard, including uncontaminated control samples. In this manner, the instances assigned to each class are listed in Table 3. To achieve balanced classes in terms of the number of instances, control samples were used where necessary, specifically for the cases of oxyfluorfen and deltamethrin. These control samples were added to replace the missing values indicated in Table 2, ensuring equal representation of contaminated and non-contaminated classes. This approach allowed us to address potential class imbalance and enhance the reliability of the classification model.

To analyze if there is class clustering, all instances used in the training phase were transformed into the principal component features space (PCA). PCA was applied to reduce the dimensionality of the hyperspectral data, which is essential given its high-dimensional nature. This transformation helps retain the most relevant features capturing maximum variance, enhancing computational efficiency without compromising critical information for classification. Fig. 4 shows the distribution of points in the feature space of the first three components, ordered by explained variance. These three components were selected because they explained more than 90% of the variance in all cases, capturing the majority of relevant information needed for sample analysis (Fig. 3b). Observing the distribution of points according to the first and second



(a) (b)

Fig. 3. (a) Loading vector of the first principal component for the set of instances used in training the classification models; (b) Variance explained by the first ten principal components.

Table 3
Labeling of instances according to the results obtained using GC-MS/MS.

	Training phase		Test phase	
	$> MRL$	$< MRL$	$> MRL$	$< MRL$
Diflufenican	$X^{difl,\#1}_{1:5}$ $X^{difl,\#2}_{1:5}$	$X^{difl,\#3}_{1:5}$ $X^{difl,\#4}_{1:5}$ $X^{difl,\#5}_{1:5}$	$X^{difl,\#1}_{6:10}$ $X^{difl,\#2}_{6:10}$	$X^{difl,\#3}_{6:10}$ $X^{difl,\#4}_{6:10}$ $X^{difl,\#5}_{6:10}$
Oxyfluorfen	$X^{oxyf,\#1}_{1:5}$ $X^{oxyf,\#2}_{1:5}$ $X^{oxyf,\#3}_{1:5}$	$X^{oxyf,\#5}_{1:5}$ $X^{ctrl}_{1:5}$	$X^{oxyf,\#1}_{6:10}$ $X^{oxyf,\#2}_{6:10}$ $X^{oxyf,\#3}_{6:10}$	$X^{oxyf,\#5}_{6:10}$ $X^{ctrl}_{6:10}$
Deltamethrin	$X^{delt,\#1}_{1:5}$ $X^{delt,\#2}_{1:5}$ $X^{delt,\#3}_{1:5}$	$X^{delt,\#5}_{1:5}$ $X^{ctrl}_{1:5}$	$X^{delt,\#1}_{6:10}$ $X^{delt,\#2}_{6:10}$ $X^{delt,\#3}_{6:10}$	$X^{delt,\#5}_{6:10}$ $X^{ctrl}_{6:10}$
λ-Cyhalothrin	$X^{cyha,\#1}_{1:5}$ $X^{cyha,\#2}_{1:5}$ $X^{cyha,\#3}_{1:5}$	$X^{cyha,\#3}_{1:5}$ $X^{cyha,\#4}_{1:5}$	$X^{cyha,\#1}_{6:10}$ $X^{cyha,\#2}_{6:10}$ $X^{cyha,\#3}_{6:10}$	$X^{cyha,\#3}_{6:10}$ $X^{cyha,\#4}_{6:10}$
Tebuconazole	$X^{tebu,\#1}_{1:5}$ $X^{tebu,\#2}_{1:5}$ $X^{tebu,\#3}_{1:5}$	$X^{tebu,\#3}_{1:5}$ $X^{tebu,\#4}_{1:5}$	$X^{tebu,\#1}_{6:10}$ $X^{tebu,\#2}_{6:10}$ $X^{tebu,\#3}_{6:10}$	$X^{tebu,\#3}_{6:10}$ $X^{tebu,\#4}_{6:10}$

principal components (PC1 and PC2, respectively), it can be seen that for diflufenican (4a) and deltamethrin (4c), the classes exhibited better clustering compared to the other cases. Specifically, in the case of diflufenican, it is evident that the histograms are clearly skewed according to PC1.

Fig. 3a shows the weight of PC1 on each wavelength. It can be observed that, in all cases, the variability of the point cloud according to PC1 can be explained using wavelengths above 700 nm. Additionally, in all cases, PC1 exhibits a clear peak around 750 nm and a second peak near 550 nm. This result could be useful for reducing the cost of the setup by using visible and infrared cameras, each with a band-pass filter. A similar case can be observed in the work of Li et al. [18], where the authors detected a peak around 700 nm in the loading vector of PC1 for deltamethrin. Both studies utilize spectral data for pesticide detection; however, Li et al. [18] applied a data fusion strategy combining UV-vis and NIR spectra with an extreme learning machine (ELM) algorithm to detect deltamethrin in formulations, emphasizing the use of multiple spectral regions for improved sensitivity. In contrast, our study focuses on hyperspectral imaging within a single spectral range (VNIR) applied directly to olive samples, which offers a non-invasive approach for monitoring pesticide residues on fruit. This difference highlights the unique applicability of our method to in-field testing, while Li et al. [18] approach is well-suited for laboratory analysis of pesticide formulations.

The results of the training phase can be observed in Table 4. In this phase, the models were constructed using 75% of the instances for training and validated with the remaining 25%. The accuracy obtained by all the models was very high, being equal to or close to 100%. This is because the instances used for model construction and validation correspond to pixels from the hyperspectral image of the same olives.

When the constructed and validated models from the training phase were applied to instances extracted from new olives (not considered in the training phase), the results obtained were those shown in Table 5. In this table, it can be observed that the model with the highest accuracy was the one for Diflufenican, with an accuracy rate of 95.03%, and it also exhibited high precision (95.72%) and recall (91.68%), resulting in an F1-score of 93.65%. These strong metrics suggest that the model is effective in minimizing both false positives and false negatives. This performance can be attributed to the high concentration and stability of Diflufenican on the olives Food and Authority [9]. However, these results should be compared with other olive varieties and different ripeness indices for broader validation.

For the herbicide Oxyfluorfen, the results were less promising, with an accuracy rate of 65.58% achieved with the KNN algorithm. The precision and recall for Oxyfluorfen were lower (68.73% and 78.23%, respectively), reflected in a modest F1-score of 73.17%. This performance, characterized by a high number of false positives, may be influenced by the photolytic degradation of Oxyfluorfen under halogen light used during image acquisition [1].

Among the insecticides, λ-Cyhalothrin achieved strong results with an accuracy of 90.11%, precision of 91.60%, recall of 91.96%, and an F1-score of 91.78%, indicating that the SVM-Q model is effective in detecting the presence or absence of this pesticide. Comparable results were obtained by He et al. [13] for garlic chive leaves using hyperspectral technology. Conversely, the model for Deltamethrin achieved lower performance, with an accuracy of 73.79%, precision of 75.04%, and recall of 84.37%, resulting in an F1-score of 79.43%. The high false positive rate for Deltamethrin likely contributed to the model's reduced precision.

Finally, the fungicide Tebuconazole achieved good results, with an accuracy of 87.25%, precision of 92.10%, recall of 86.14%, and an F1-score of 89.02%. This high performance can be attributed to the prolonged persistence of Tebuconazole on the olive surface, as shown in Table 2 with the highest residue concentrations detected by GC-MS/MS.

4. Conclusions

This study demonstrates the feasibility of using hyperspectral imaging for pesticide detection in freshly harvested olives. The proposed

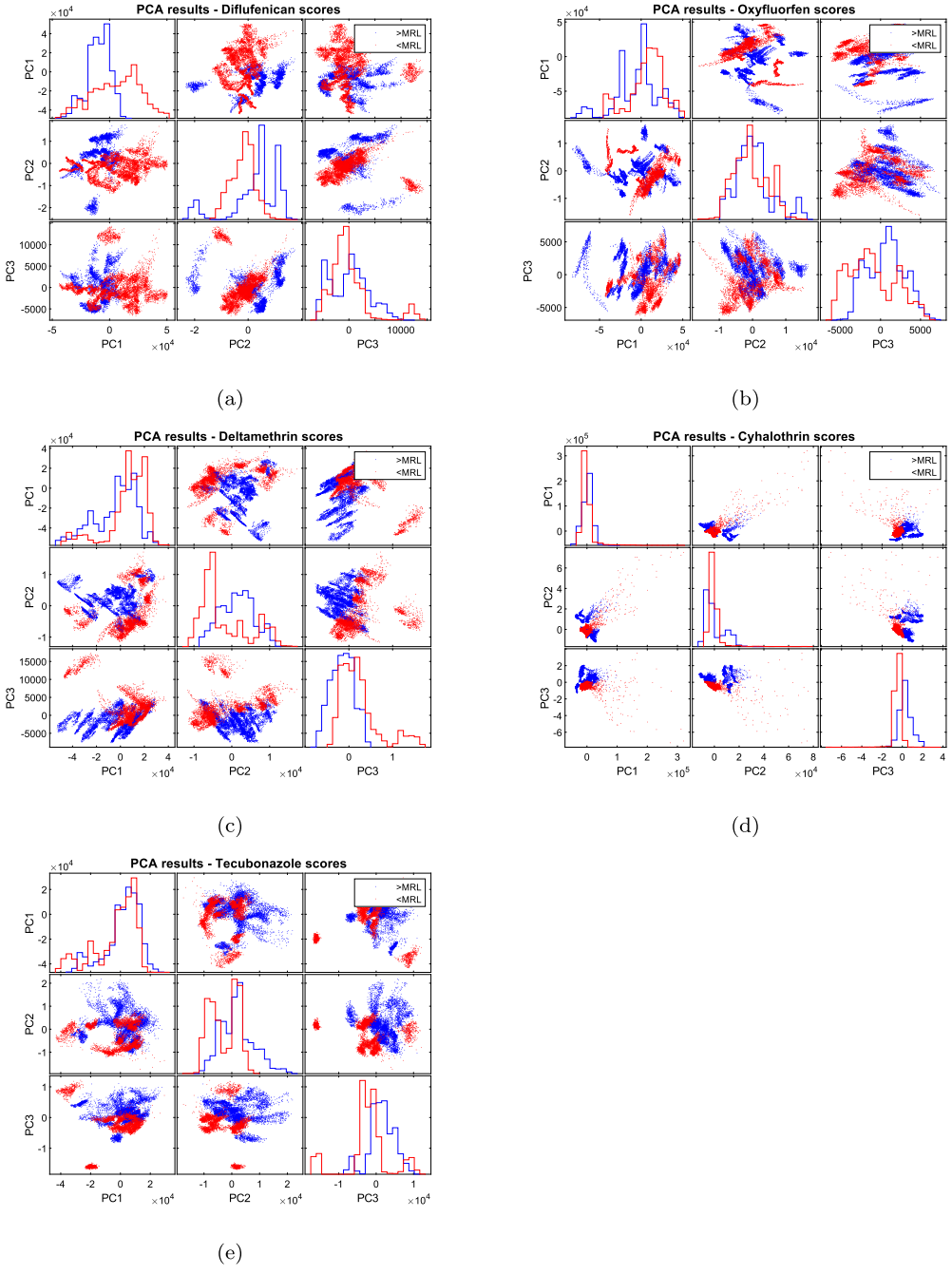


Fig. 4. Results of the PCA analysis using the training data matrices of the classification models for predicting contaminations. Subfigures show the scatter plot in the feature space of the first three principal components.

Table 4
Results for the training phase of the binary classification models, including precision, recall, accuracy, F1-score, and AUC-ROC metrics in percentages.

Pesticide	Best model	True class (<MLR)		True class (>MLR)		P	R	A	F1	AUC
		Predicted as <MLR	Predicted as >MLR	Predicted as <MLR	Predicted as >MLR					
Diflufenican	LD	1654	0	1	1101	100.00%	99.90%	100.00%	99.95%	100.00%
Oxyfluorfen	KNN	1102	0	1	1653	100.00%	99.90%	100.00%	99.97%	100.00%
Deltamethrin	SVM-C	1102	0	1	1653	100.00%	99.90%	100.00%	99.97%	100.00%
λ-Cyhalothrin	SVM-Q	1100	3	1	1652	99.82%	99.90%	99.90%	99.88%	100.00%
Tebuconazole	SVM-Q	1103	0	1	1652	100.00%	99.96%	99.96%	99.97%	100.00%

Abbreviations: LD: Linear Discriminant; KNN: K-Nearest Neighbor; SVM-C: Support Vector Machine Cubic Kernel; SVM-Q: Support Vector Machine Quadratic Kernel; P: Precision; R: Recall; A: Accuracy; F1: F1-score; AUC: Area Under the Curve.

Table 5
Results for the test phase of the binary classification models, including precision, recall, accuracy, F1-score, and AUC-ROC metrics in percentages.

Pesticide	Best model	True class (<MLR)		True class (>MLR)		P	R	A	F1	AUC
		Predicted as <MLR	Predicted as >MLR	Predicted as <MLR	Predicted as >MLR					
Diflufenican	LD	6434	181	367	4043	95.03%	95.00%	95.03%	95.02%	99.19%
Oxyfluorfen	KNN	2055	2355	1440	5175	65.58%	64.90%	65.58%	65.24%	62.41%
Deltamethrin	SVM-C	2554	1856	1034	5581	73.79%	73.70%	73.79%	73.74%	80.56%
λ-Cyhalothrin	SVM-Q	3852	558	532	6083	90.11%	90.00%	90.11%	90.05%	96.29%
Tebuconazole	SVM-Q	3921	489	917	5698	87.25%	87.10%	87.25%	87.17%	96.20%

Abbreviations: LD: Linear Discriminant; KNN: K-Nearest Neighbor; SVM-C: Support Vector Machine Cubic Kernel; SVM-Q: Support Vector Machine Quadratic Kernel; P: Precision; R: Recall; A: Accuracy; F1: F1-score; AUC: Area Under the Curve.

acquisition technology, covering the visible and near-infrared spectrum, offers low installation costs and rapid response times, making it suitable for in-line applications, such as at the critical moment when farmers unload olives at the cooperative yard. Implementing quality control at this point in the process allows cooperatives to immediately assess contamination levels, facilitating efficient decision-making and ensuring that only compliant batches proceed to production. This approach enhances operational efficiency by reducing the time, costs, and labor associated with traditional testing methods.

Key wavelengths around 750 nm and 550 nm were identified as effective for discriminating pesticide residues. The classification models achieved the best results in detecting levels above the regulatory limit for Diflufenican and λ-Cyhalothrin, with accuracy rates of 95.03% and 90.11%, respectively. Additionally, metrics such as precision, recall, F1-score, and AUC-ROC provide a comprehensive evaluation, confirming the models' discriminative power for these pesticides.

However, this study presents several limitations. The current experimental setup involves applying pesticide solutions directly onto harvested olives, which ensures uniform concentration and distribution on the fruit surface. When pesticides are applied directly to trees, the distribution and concentration on the olives may vary due to environmental factors such as rain, wind, and sunlight, which can degrade or disperse the chemicals unevenly. Further experimentation under these realistic conditions would provide a more accurate validation of the technique's effectiveness, accounting for natural variations in pesticide residue levels.

Additionally, class imbalance in pesticide detection can impact model reliability, especially in smaller datasets. While effective at the pixel level, future work should address potential biases by employing larger and more diverse sample sets that include multiple olive varieties and ripeness stages. Environmental factors, such as lighting and temperature fluctuations, may also impact the robustness of hyperspectral imaging. Further experimentation under variable lighting conditions and ambient temperatures is recommended to ensure consistent performance in real-world conditions.

The broader interdisciplinary potential of hyperspectral technology also warrants discussion. Beyond pesticide detection in olives, hyperspectral imaging holds promise for food safety applications, enabling rapid, non-invasive quality control for various food products. Furthermore, this technology could extend to sectors like medical diagnostics, where hyperspectral imaging is being explored for early disease detection and tissue analysis, demonstrating its adaptability and potential impact across fields.

Overall, while the models show high efficacy, additional research is needed to validate these findings under field conditions by applying treatments directly to trees and evaluating the technique's applicability to heterogeneous olive batches. Addressing these challenges will contribute to a robust and scalable solution, enabling widespread adoption of hyperspectral imaging in agriculture and beyond.

CRediT authorship contribution statement

Diego Manuel Martínez Gila: Writing – original draft, Methodology, Formal analysis, Conceptualization. **David Bonillo Martínez:** Software, Methodology, Investigation, Data curation. **Silvia Satorres Martínez:** Supervision, Resources, Conceptualization. **Pablo Cano Marchal:** Investigation, Formal analysis, Conceptualization. **Javier Gámez García:** Writing – original draft, Visualization, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was partially funded by the Spanish Ministry of Science and Innovation under the project ESPECTROLIVE with reference AEI-010500-2023-232 and by the projects of the national plan with references PDC2022-133995-I00 and PID2023-150832OB-I00.

Data availability

Data will be made available on request.

References

[1] M. Anastassiadou, G. Bernasconi, A. Brancato, L. Carrasco Cabrera, L. Ferreira, L. Greco, S. Jarrah, A. Kazocina, R. Leuschner, J.O. Magrans, I. Miron, S. Nave, R. Pedersen, H. Reich, A. Rojas, A. Sacchi, M. Santos, A. Theobald, B. Vagenende, A. Verani, Review of the existing maximum residue levels for oxyfluorfen according to Article 12 of Regulation (EC) No 396/2005, EFSA J. 18 (2020) e06269, https://doi.org/10.2903/J.EFSA.2020.6269, https://onlinelibrary.wiley.com/doi/full/10.2903/j.efsa.2020.6269, https://onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.6269.

[2] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and regression trees, in: Classification and Regression Trees, 2017, pp. 1–358, https://www.taylorfrancis.com/books/mono/10.1201/9781315139470/classification-regression-trees-leo-breiman-jerome-friedman-olshen-charles-stone.

[3] C. Cortes, V. Vapnik, L. Saitta, Support-vector networks, Mach. Learn. 20 (20:3 20) (1995) 273–297, https://doi.org/10.1007/BF00994018, https://link.springer.com/article/10.1007/BF00994018.

[4] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1967) 21–27, https://doi.org/10.1109/TIT.1967.1053964.

[5] Y. El-Nahhal, I. El-Nahhal, Cardiotoxicity of some pesticides and their amelioration, Environ. Sci. Pollut. Res. 28 (2021) 44726–44754, https://doi.org/10.1007/s11356-021-14999-9, https://link.springer.com/10.1007/s11356-021-14999-9.

[6] E. Elahi, C. Weijun, H. Zhang, M. Nazeer, Agricultural intensification and damages to human health in relation to agrochemicals: application of artificial intelligence, Land Use Policy 83 (2019) 461–474, https://doi.org/10.1016/j.landusepol.2019.02.023, https://linkinghub.elsevier.com/retrieve/pii/S0264837718308548.

[7] C. Ferrer, M.J. Gómez, J.F. García-Reyes, I. Ferrer, E.M. Thurman, A.R. Fernández-Alba, Determination of pesticide residues in olives and olive oil by matrix solid-phase dispersion followed by gas chromatography/mass spectrometry and liquid chromatography/tandem mass spectrometry, J. Chromatogr. A 1069 (2005) 183–194, https://doi.org/10.1016/j.chroma.2005.02.015, https://linkinghub.elsevier.com/retrieve/pii/S0021967305002578.

[8] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (1936) 179–188, https://doi.org/10.1111/J.1469-1809.1936.TB02137.X, https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-1809.1936.tb02137.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1936.tb02137.x.

[9] E. Food, S. Authority, Reasoned opinion on the modification of the existing MRLs for diflufenican in olives for oil production, EFSA J. 10 (2012) 2649, https://doi.org/10.2903/J.EFSA.2012.2649, https://onlinelibrary.wiley.com/doi/full/10.2903/j.efsa.2012.2649, https://onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2012.2649, https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2012.2649.

[10] H. Fraga, J.G. Pinto, F. Viola, J.A. Santos, Climate change projections for olive yields in the Mediterranean Basin, Int. J. Climatol. 40 (2020) 769–781, https://doi.org/10.1002/joc.6237, https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.6237.

[11] J.F. García-Reyes, C. Ferrer, M.J. Gómez-Ramos, A.R. Fernández-Alba, J.F. García-Reyes, A. Molina-Díaz, Determination of pesticide residues in olive oil and olives, TrAC, Trends Anal. Chem. 26 (2007) 239–251, https://doi.org/10.1016/j.trac.2007.01.004, https://linkinghub.elsevier.com/retrieve/pii/S0165993607000052.

[12] D.M.M. Gila, J.G. García, A. Bellincontro, F. Mencarelli, J.G. Ortega, Fast tool based on electronic nose to predict olive fruit quality after harvest, Postharvest Biol. Technol. 160 (2020) 111058, https://doi.org/10.1016/j.postharvbio.2019.111058.

[13] W. He, H. He, F. Wang, S. Wang, R. Lyu, Non-destructive detection and recognition of pesticide residues on garlic chive (Allium tuberosum) leaves based on short wave infrared hyperspectral imaging and one-dimensional convolutional neural network, J. Food Meas. Charact. 15 (2021) 4497–4507, https://doi.org/10.1007/S11694-021-01012-7/TABLES/4, https://link.springer.com/article/10.1007/s11694-021-01012-7.

[14] M. Hu, D. Yang, D.J. Huber, Y. Jiang, M. Li, Z. Gao, Z. Zhang, Reduction of postharvest anthracnose and enhancement of disease resistance in ripening mango fruit by nitric oxide treatment, Postharvest Biol. Technol. 97 (2014) 115–122, https://doi.org/10.1016/j.postharvbio.2014.06.013, https://linkinghub.elsevier.com/retrieve/pii/S0925521414001835.

[15] E. Lantero, B. Matallanas, C. Callejas, Current status of the main olive pests: useful integrated pest management strategies and genetic tools, Appl. Sci. 13 (2023) 12078, https://doi.org/10.3390/app132112078, https://www.mdpi.com/2076-3417/13/21/12078.

[16] C. Lentza-Rizos, E.J. Avramides, Pesticide residues in olive oil, in: Reviews of Environmental Contamination and Toxicology: Continuation of Residue Reviews, 1995, pp. 111–134, http://link.springer.com/10.1007/978-1-4612-2530-0_4.

[17] H. Li, P. Merkl, J. Sommertune, T. Thersleff, G.A. Sotiriou, Sers hotspot engineering by aerosol self-assembly of plasmonic ag nanoaggregates with tunable interparticle distance, Adv. Sci. 9 (2022) 2201133, https://doi.org/10.1002/advs.202201133.

[18] Q. Li, Y. Huang, J. Zhang, S. Min, A fast determination of insecticide deltamethrin by spectral data fusion of UV–vis and NIR based on extreme learning machine, Spectrochim. Acta, Part A, Mol. Biomol. Spectrosc. 247 (2021) 119119, https://doi.org/10.1016/J.SAA.2020.119119.

[19] L. Liu, Y. Wang, C. Gao, H. Huan, B. Zhao, L. Yan, Photoacoustic spectroscopy as a non-destructive tool for quantification of pesticide residue in apple cuticle, Int. J. Thermophys. 36 (2015) 868–872, https://doi.org/10.1007/s10765-014-1705-2.

[20] Y. Lu, X. Li, W. Li, T. Shen, Z. He, M. Zhang, H. Zhang, Y. Sun, F. Liu, Detection of chlorpyrifos and carbendazim residues in the cabbage using visible/near-infrared spectroscopy combined with chemometrics, Spectrochim. Acta, Part A, Mol. Biomol. Spectrosc. 257 (2021) 119759, https://doi.org/10.1016/J.SAA.2021.119759.

[21] F. Luchetti, Importance and future of olive oil in the world market—an introduction to olive oil, European J. Lipid Sci. Technol. 104 (2002) 559–563.

[22] D.M. Martínez Gila, C. Sanmartin, J. Navarro Soto, F. Mencarelli, J. Gómez Ortega, J. Gámez García, Classification of olive fruits and oils based on their fatty acid ethyl esters content using electronic nose technology, J. Food Meas. Charact. 15 (2021) 5427–5438, https://doi.org/10.1007/s11694-021-01103-5.

[23] D. Min, F. Li, M. Ali, X. Zhang, Y. Liu, Application of methyl jasmonate to control disease of postharvest fruit and vegetables: a meta-analysis, Postharvest Biol. Technol. 208 (2024) 112667, https://doi.org/10.1016/j.postharvbio.2023.112667, https://linkinghub.elsevier.com/retrieve/pii/S0925521423004283.

[24] T.M. Mitchell, Learning classifiers based on Bayes rule, Mach. Learn. 1 (2010) 1–17, https://doi.org/10.1093/bioinformatics/btq112, http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf.

[25] J. Qin, M.S. Kim, K. Chao, S. Dhakal, B.K. Cho, S. Lohumi, C. Mo, Y. Peng, M. Huang, Advances in Raman spectroscopy and imaging techniques for quality and safety inspection of horticultural products, Postharvest Biol. Technol. 149 (2019) 101–117, https://doi.org/10.1016/j.postharvbio.2018.11.004.

[26] A. Serrano-Medina, A. Ugalde-Lizárraga, M. Bojorquez-Cuevas, J. Garnica-Ruiz, M. González-Corral, A. García-Ledezma, G. Pineda-García, J. Cornejo-Bravo, Neuropsychiatric disorders in farmers associated with organophosphorus pesticide exposure in a rural village of Northwest México, Int. J. Environ. Res. Public Health 16 (2019) 689, https://doi.org/10.3390/ijerph16050689, http://www.mdpi.com/1660-4601/16/5/689.

[27] S. Sindhu, A. Manickavasagan, Nondestructive testing methods for pesticide residue in food commodities: a review, Compr. Rev. Food Sci. Food Saf. 22 (2023) 1226–1256, https://doi.org/10.1111/1541-4337.13109, https://ift.onlinelibrary.wiley.com/doi/10.1111/1541-4337.13109.

[28] R.M. de Souza, D. Seibert, H.B. Quesada, F. de Jesus Bassetti, M.R. Fagundes-Klen, R. Bergamasco, Occurrence, impacts and general aspects of pesticides in surface water: a review, Process Saf. Environ. Prot. 135 (2020) 22–37, https://doi.org/10.1016/j.psep.2019.12.035, https://linkinghub.elsevier.com/retrieve/pii/S0957582019318683.

[29] I.N. Therios, Olives, vol. 18, CABI, 2009.

[30] M. Tudi, H. Daniel Ruan, L. Wang, J. Lyu, R. Sadler, D. Connell, C. Chu, D.T. Phung, Agriculture development, pesticide application and its impact on the environment, Int. J. Environ. Res. Public Health 18 (2021) 1112, https://doi.org/10.3390/ijerph18031112, https://www.mdpi.com/1660-4601/18/3/1112.

[31] J. Wang, S. Wang, N. Liu, F. Shang, A detection method of two carbamate pesticides residues on tomatoes utilizing excitation-emission matrix fluorescence technique, Microchem. J. 164 (2021) 105920, https://doi.org/10.1016/j.microc.2021.105920.

[32] K.J. Yang, J. Lee, H.L. Park, Organophosphate pesticide exposure and breast cancer risk: a rapid review of human, animal, and cell-based studies, Int. J. Environ. Res. Public Health 17 (2020) 5030, https://doi.org/10.3390/ijerph17145030, https://www.mdpi.com/1660-4601/17/14/5030.

[33] W. Zhang, F. Jiang, J. Ou, Global pesticide consumption and pollution: with China as a focus, Proc. Int. Acad. Ecol. Environ. Sci. 1 (2011) 125.

[34] X. Zhang, L. Chen, X. Fang, Y. Shang, H. Gu, W. Jia, G. Yang, Y. Gu, L. Qu, Rapid and non-invasive surface-enhanced Raman spectroscopy (sers) detection of chlorpyrifos in fruits using disposable paper-based substrates charged with gold nanoparticle/halloysite nanotube composites, Mikrochim. Acta 189 (2022) 197, https://doi.org/10.1007/s00604-022-05261-1.