FRAUNHOFER INSTITUTE (GERMANY)

Flora Zidane,
Jérôme Lanteri, Julien Marot,
Laurent Brochier,
Nadine Joachimowicz,
Hélène Roussel, and
Claire Migliaccio

Nondestructive Control of Fruit Quality via Millimeter Waves and Classification Techniques

Investigations in the automated health monitoring of fruits.

Digital Object Identifier 10.1109/MAP.2020.3003222
Date of current version: 20 August 2020

Fast and efficient nondestructive evaluation (NDE) methods for food control is still an ongoing field of research. We have recently proposed to combine W-band imaging with nonlinear support vector machine (SVM) classifiers to sort out healthy from damaged fruits for a single variety of fruit. We have tested it on apples and peaches separately

 1045-9243/20©2020IEEE

with a mean accuracy of 96%. We have also shown the limitation of a biclass SVM since it has failed to sort healthy from damaged fruits when the set of fruits was composed of a mix of apples and peaches. In this article, we continue to explore the capability of SVM associated with millimeter-wave (mm-wave) low-terahertz (THz) measurements. First, we tackle the problem of classifying a mix of fruits with a multiclass SVM using the Digital Binary Tree architecture. With this method, the error rate does not exceed 2%. Secondly, we move from the W- to D-band (low-THz). The main reason is the increase of the lateral resolution and the possibility to have more compact systems in the view of an industrial deployment. We start our D-band investigations with range measurements to estimate the average permittivity of the apple in this frequency bandwidth. We have found a drastic decrease compared to the microwave region. It is consistent with the behavior of the water, which is one of the main components of the apple. Then we trained the SVM with the D-band database and finally performed the classification on unknown samples and obtained an accuracy of 100%.

> **Optically based solutions associated with machine learning algorithms were previously developed to assess the fruits' quality.**

INTRODUCTION

In present markets, fruits may come from distant places after being stored in refrigerators. Hence, they may suffer from dehydration, nutrient contamination, and so forth. Automated health monitoring of fruits is, therefore, of great interest. There are many destructive evaluation techniques that can predict their condition when fruits are sold, such as penetration tests [1], [2], ultrasound exploration that is used to determine the concentration of a particular component [3], and polar organic solvents [4]. All of these methods are mechanical, expensive and sometimes ineffective. Moving to NDE methods has been a significant progress. Among them, near-infrared [5], [6], X-ray imaging [7]–[9], acoustic pulse [10], [11], and nuclear magnetic resonance [12] are efficient. But they are limited by their cost and complexity.

To tackle this problem, optically based solutions associated with machine learning algorithms (MLAs) were previously developed to assess the fruits' quality. They exhibit a low computational load while providing elevated classification performances. Among MLAs, *k*-nearest neighbor [13], SVMs [14]–[16], and neural networks [17], [18] yielded valuable results on various fruits, such as apples, cherries, strawberries, and kiwifruits.

However, optical methods remain limited because they cannot detect damages below the skin. Indeed, an optical wave is purely evanescent inside the fruit. On the contrary, and despite a worse resolution, microwaves have a much better penetration depth. In 1973, Nelson wrote an exhaustive article on the permittivity of fruits and vegetables and their dependence on their maturity and variety [19], demonstrating the potential of microwaves for NDE control of agricultural products. The increasing interest for food NDE with microwave imaging is due, at least, to three reasons.

1) Microwave imaging is very sensitive to the water content of the tested items.
2) Microwaves penetrate dielectric materials and the resulting electromagnetic field signature is characteristic of the overall volume, including the internal part of the tested item.
3) The expanding technology of microwaves makes it possible to obtain fast and low-cost equipment, which is safe for operators.

Consequently, various industrial applications using microwave sensors have emerged including NDE of material moisture content [20]–[22], water solution concentrations [23], continuous process monitoring for biogas plants [24], vegetable oils [25], and plastic inclusions in hazelnut/cocoa cream [26]. The drawback of microwaves is that their intrinsic poor resolution turns the image interpretation into a difficult task. Therefore, an automatic image exploitation is necessary. Machine learning methods like SVM have already been applied successfully in microwave applications, such as microwave device and antenna modeling by SVMs [27]–[29], resolution of inverse scattering [30], solving of electromagnetic problems [31], modeling of the electromagnetic response of complex shaped reflector elements [32], or for resolving the medical problem as the breast cancer detection by using mm-wave imaging and SVMs [33], [34]. Other successful applications include optimizing the extraction of abnormal features from microwave breast tomography data [35] to diagnose hepatic malignancies by dielectric measurements in vivo using microwaves [36], retrieving the characteristics of a defect with eddy current (EC) testing in real time [37], or testing and evaluating (nondestructive testing/NDE) of faulty conductive tubes from EC measurements [38].

In our previous work, we have developed an automatic nondestructive method to separate healthy apples (respectively, peaches) from damaged apples (respectively, peaches) by using mm-wave imaging and nonlinear SVM [39]. We obtained a mean accuracy of 94% for apple sorting and 100% for peach sorting but we failed in trying to sort "healthy" peaches and apples from "damaged" peaches and apples by using biclass SVM. The aim of this article is to show the results obtained from the classification of different healthy fruits from damaged fruits, i.e., from a database containing images of peaches and apples for detecting the damaged fruit. Next, we present the results of sorting apples with different varieties of apples in low-THz frequencies. We chose to move from the W-band to the D-band because recent studies have shown the potential of THz imaging. In the following, we report a nonexhaustive section of these studies. First, we can cite the monitoring of grapevine water status by measuring the reflectivity at the trunk in the THz band [40], the development of a quantitative method to calculate the water content of thin materials by THz imaging [41], and the detection of defects due to the difference in moisture content in the thin surface layers of tomatoes [42]. The experimental results in [42] have shown the feasibility of the detection of internal damages. In [43], the authors estimate the water content in plants leaves.

Second, and as concerns more precisely fruit quality assessment, we report the detection of the chemical characteristics and signatures of inner compositions for quality control in fruits with THz imaging [44]. In [45], THz imaging is associated with a machine learning driven approach for fruit classification. The limitation of k-nearest neighbors and the D-tree was demonstrated, and the interest of SVM, used with a Gaussian radial basis function (RBF) kernel, was emphasized.

WORKING METHODOLOGY

WORKFLOW

The complete workflow consists of the five steps summarized as follows and described in Figure 1.

- **Step 1: Taking mm-wave measurements**. The measurement setup is shown in Figure 2. Here we follow the same process as in [39] where we explain in detail the choices of the measurement parameters. In particular, we discuss the importance of having frequency diversity as well as incidences and views diversity by the means of using different measurement patches. One of the critical points is the choice of the intermediate frequency (IF) that in turn influences the noise floor level and the measurement time. We have chosen an IF of 100 Hz as the best compromise. With these settings, the noise floor is -70 dBm for both D- and W-bands.

  Each fruit is measured with a 3D spherical scanner according to the system described in [46]. We perform scattering measurements in monostatic configuration under horizontal polarization frequency per frequency. We measure each fruit at three frequencies for increasing the number of measured data. We measure patches of $10° \times 10°$ with a scan step of 0.2°, which consist in 51×51 measurements points. The measurement of one patch lasts 130 s. The total scan area covered by the measured patches is $30° \times 30°$ centered above the fruit.
- **Step 2: Computing mm-wave images**. Each measured patch is processed with a 2D fast Fourier transform to compute a 2D red, green, blue (RGB) image at each frequency.
- **Step 3: Segmentation**. We transform the 2D RGB images into binarized ones (see an example of a binary image in Figure 1) by using Otsu's method [47]. The key point to transform an RGB image into a binary one is the choice of the threshold that "decides" whether the pixel will be transferred into a "0" or a "1." In our setup, the RGB images are very different from each other because they are strongly dependent on the measurement frequency and the patch, i.e., scan angles, hence the interest of a dynamic threshold. The Otsu method is an optimization method enabling the computation of the threshold dynamically and automatically. The threshold depends on the intensity of each image and thus perfectly matches our requirements. We resize the images into 51×51 pixels with the down-scaling function.
- **Step 4: Selecting training and test data sets**. We split the set of binary images into training and test data sets. We find the optimal parameters of the SVM classifier with the training data set whereas we use the test data set to compute the classifier's accuracy.

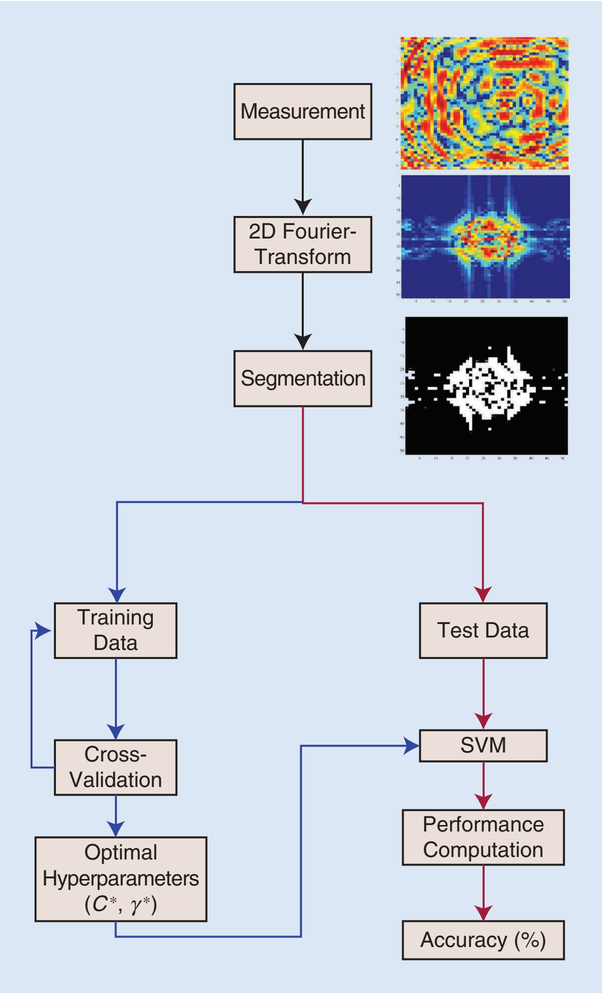

FIGURE 1. The flowchart of the proposed classification process.

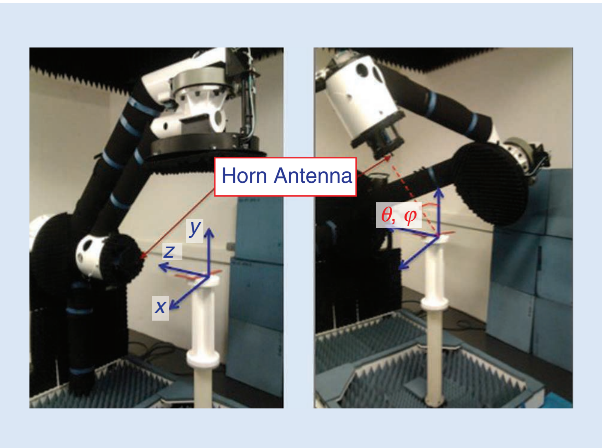

FIGURE 2. The measurement setup.

- **Step 5: Computing SVM optimal hyperparameters and accuracy**. Here we investigate and discuss the performance of biclass and multiclass SVM for the sorting of fruits. Both methods are briefly explained in the sections "Biclass SVM" and "Multiclass SVM," respectively.

THz imaging is associated with a machine learning driven approach for fruit classification.

BICLASS SVM

An SVM is an MLA [48], which consists in finding an optimal separating hyperplane between positive and negative classes. The SVM algorithm aims to maximize a margin around the hyperplane [49], [50]. We find the optimal hyperplane by solving the following problem:

$$\max_{\alpha}\left[\sum_{i=1}^{n}\alpha_i - \sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \,.\, K(x_i, x_j)\right] \tag{1}$$

$$\alpha_i \geq 0 \qquad \sum_{i=1}^{n}\alpha_i y_i = 0. \qquad i = 1 \ldots n, \tag{2}$$

where

- n is the number of samples
- x_i is the ith sample (usually a vector) of the training data S. It corresponds to the vectorized version of a segmented image obtained at a given frequency for given measurement patch
- $S = (x_i, y_i) \mid x_i \in R^n, \quad y_i \in \{-1, 1\}_{i=1}^{n}$
- y_i is the label of the sample of training data x_i
- α is a Lagrangian variable.

If $\alpha_i \neq 0$, x_i is called a *support vector*. The kernel function K is used to measure the similarity between two samples. There are different kernel functions, such as the linear, polynomial, or Gaussian ones [51]. Here we use the Gaussian RBF defined by (3):

$$K(x_i, x_j) = \exp(-\gamma \,\|\langle x_i, x_j\rangle\|). \quad \gamma > 0. \tag{3}$$

The hyperparameters of the SVM are C and γ. C is a regulation parameter, which adjusts the width of the margin to reach the highest possible accuracy. γ, also called *scale*, defines the variance of data. Throughout the article, we apply the threefold cross-validation with grid search for each training data set, both to find the optimal values of C^* and γ^* and to avoid overfitting.

MULTICLASS SVM

The SVM classifier was originally just binary, but a large set of applications requires the classification of heterogeneous populations, which encloses several types of individuals, i.e., several classes. The extension from biclass to multiclass SVM is still a subject of research [52] with many approaches currently available. We have selected a version of SVM utilizing a binary decision tree (BDT) of SVM (SVM-BDT) because it is very fast, provided that we know how to prioritize the different classification stages.

The SVM-BDT method combines the BDT architecture and the SVM method to solve a multiclass problem. In the training phase, the SVM-BDT method needs $(N-1)$ SVMs for training N classes but the test phase only needs $(\log_2 N)$ SVMs to classify the samples. This greatly reduces the classification time of an unknown sample. The general workflow of an SVM-BDT classifier is shown in Figure 3. The test phase always starts with the binary node of the root; the unknown sample is then assigned to one of the two possible classes and then transferred to the subtree corresponding to the assigned group. This is repeated recursively down the tree until the sample reaches a leaf node that represents the class it has been assigned to. There are many ways to separate classes into two groups. This step is crucial; the groups must be well chosen for the good performance of SVM-BDT [53].

DATA SETS

Measurements of fruits were conducted between September 2018 and October 2019 on 11 apples and three peaches. Each fruit is measured at three frequencies over the same scan area as defined in the section "Workflow," hence we generate 303 images per fruit. We started our investigation in the W-band, but we recently pursued in the low-THz region (D-band) to take advantage of a better spatial resolution and a better compactness due to the antenna size reduction. Due to the long time frame of the measurements and the decay of the fruits, it is not possible to measure the same fruits at each measurement campaign. We also had to change the type of fruit according to the seasons, which explains why we have tested apples and peaches. We have conducted three campaigns, two in the W-band and the last one in the D-band. Table 1 summarizes the main features of the three measurement campaigns.

- **First campaign: September 2018**. We measure three pairs of what are called *Golden* apples in the W-band. The apples are sorted by size, i.e., large, medium, and small. A pair of apples consists of one healthy and one damaged apple from the same size.
- **Second campaign: July 2019**. We measure three peaches in the W-band. Two of them are damaged (peaches 1 and 2) and the other one is healthy.
- **Third campaign: October 2019**. We measure a mix of what are called *Gala* along with Golden apples in the D-band. We wanted to test both the possibility to discriminate damaged fruits of different apple's varieties and the performance of a low-THz frequency band for this application. The total number of apples is five, among which three are damaged Gala (1, 2, and 3), one is a damaged Golden (4), and the last one is a healthy Golden (5).

RESULTS OF FIRST AND SECOND MEASUREMENT CAMPAIGNS

Conducting the first and second measurement campaigns, we have obtained the following results:

- A significant difference in the dielectric constant between healthy and damaged apples was measured at low frequency. This is our starting point to use a classifier based on scattered field measurements.

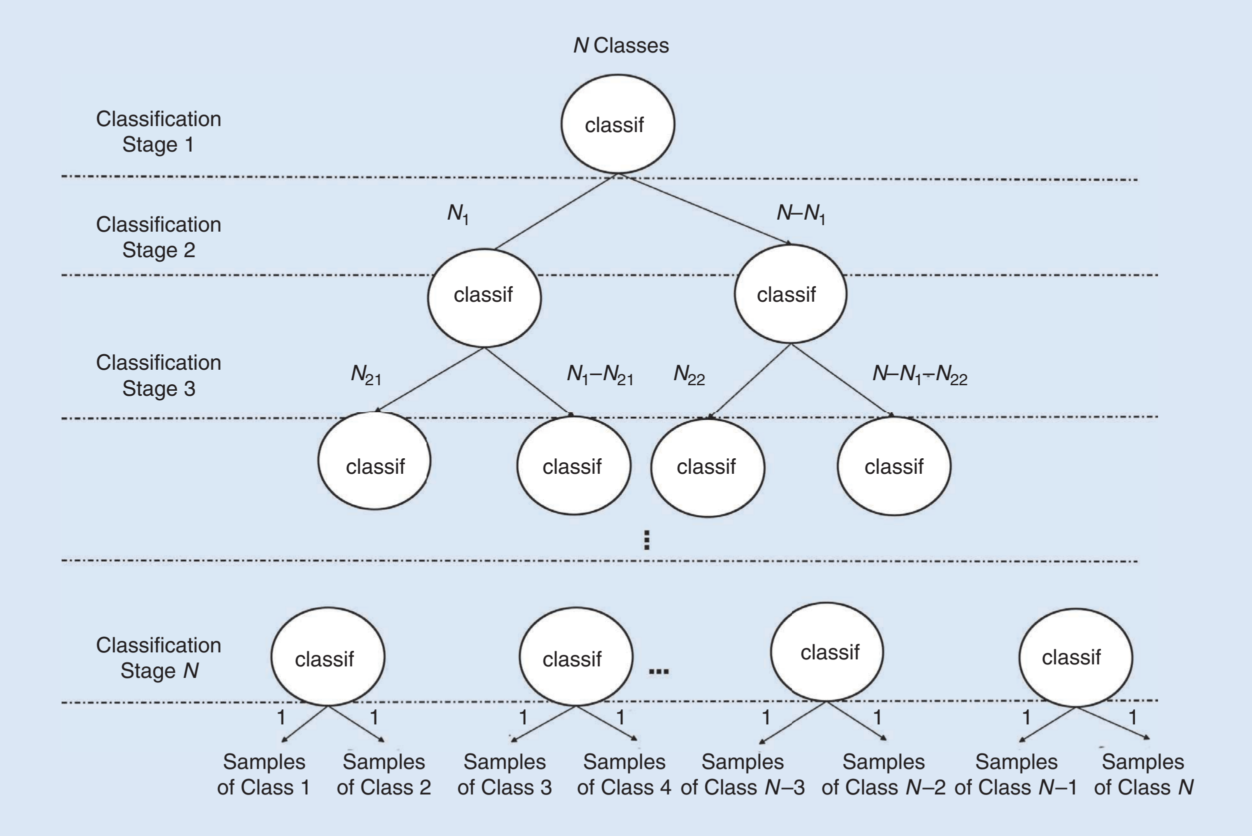

FIGURE 3. The workflow of the multiclass SVM. Classif = SVM classifier.

- The nonlinear SVM classifier is efficient to sort damaged apples from healthy ones. The same is true for peaches.
- The optimal parameters C and γ depend on the fruit. We are reminded here of these values, since we will use them in the following. For apples: $(C^*, \gamma^*) = (125, 0.01)$. For peaches: $(C^*, \gamma^*) = (100, 0.01)$.
- The results are sensitive to the choice of the measurement frequencies. New optimal values (C^*, γ^*) must be calculated when the frequency changes or when we remove a measurement frequency.
- It is possible to reduce the measurement time by a ratio of 16 since the classification still works with patches of 13×13 points. A switched antenna array of $13*13$ elements is feasible (even though challenging) in the W- and D-bands and will enable real-time measurements, which is up to now the main limitation toward a practical implementation.
- The classification fails if we try to sort healthy from damaged fruits when we train the SVM with a mix of images from apples and peaches.

In the following section, we try to solve the problem of the classification of a mix of damaged fruits by using a multiclass SVM.

TABLE 1. PARAMETERS OF THE THREE MEASUREMENT CAMPAIGNS.

Parameter	First Campaign	2D Campaign	3D Campaign
Frequency (GHz)	92-94-96	92-94-96	122-124-126
Number of images	1,818	909	1,515
Training data set	*L* and *M*	1, 3	4, 5
Test data set	*S*	2	1, 2, 3
Training data set	—	—	1, 5
Test data set	—	—	2, 3, 4

L = large-size apples; *M* = medium-size apples; *S* = small-size apples.

CLASSIFICATION OF HEALTHY AND DAMAGED FRUITS WITH MULTICLASS SVM

As we know that the value of dielectric permittivity ϵ_r varies according to the variety and the state of the fruit [54], it is important to know whether the variety of fruits prevails over the damages or vice versa. This can be easily checked by applying a biclass SVM to a mix of fruits while defining two classes only—healthy and damaged fruits. We test this hypothesis by using the data of the first and second measurement campaigns. The partition between the training data set and the test data set is as follows.

- **Training:** Four apples and two peaches. (Apples: the pair of large-size apples and the pair of medium-size apples. Peaches: 1 and 3).
- **Test:** Apple: the pair of small apples. Peach: damaged peach 2.

We obtained an accuracy of 64%. This inconclusive result can be explained by the fact that the difference in permittivity between fruits (peaches and apples) is greater than the one between a healthy and a damaged fruit, which means that we ought to define four classes: healthy apple, damaged apple, healthy peach, and damaged peach. Hence, we move from biclass to multiclass SVM.

We implement the multiclass SVM "BDT-SVM" described in the section "Multiclass SVM" according to the workflow shown in Figure 4. We split the data for the training and the test as settled above. At the first classification stage, we define two classes: the peach class, labeled with -2, and the apple one, labeled with $+2$. Each class contains two subclasses: the healthy class labeled -1 and the damaged one, labeled $+1$. The performances of the classifiers are evaluated with three classification metrics: sensibility (SENS), specificity (SPEC), and accuracy. The SENS is the probability of classifying a sample as true positive whereas the SPEC indicates the probability of classifying it as true negative. SPEC and SENS range from 0 to 1 [55].

First, we must find the optimal values of (C^*, γ^*) for the first classification stage. To this end, we randomly select 80% from the images of the training data set. We proceed to a principal component analysis (PCA) to see how both classes are linked with each other. Figure 5 shows the projection of these data on the three first components. As it has been observed in [39], the data are deeply intricated, which justifies a posteriori the choice of a nonlinear kernel for the SVM. The remaining 20% of the training data set are used for the validation that consists in a pretest to show the efficiency of the hyperplane defined by (C^*, γ^*). This prevents overfitting. Overfitting means that the classifier learns very well on the 80% used for computing (C^*, γ^*) but fails when new data have to be classified. In other words, we have not reached the optimal values of C and γ. In this case, we must carry on with the training and search for new C and γ. The optimal parameters for the first classification stage are found at $(C^*, \gamma^*) = (100, 0.01)$. The results of the classification are shown in Figures 6 and 7 for the training and test data sets, respectively. The confusion matrix,

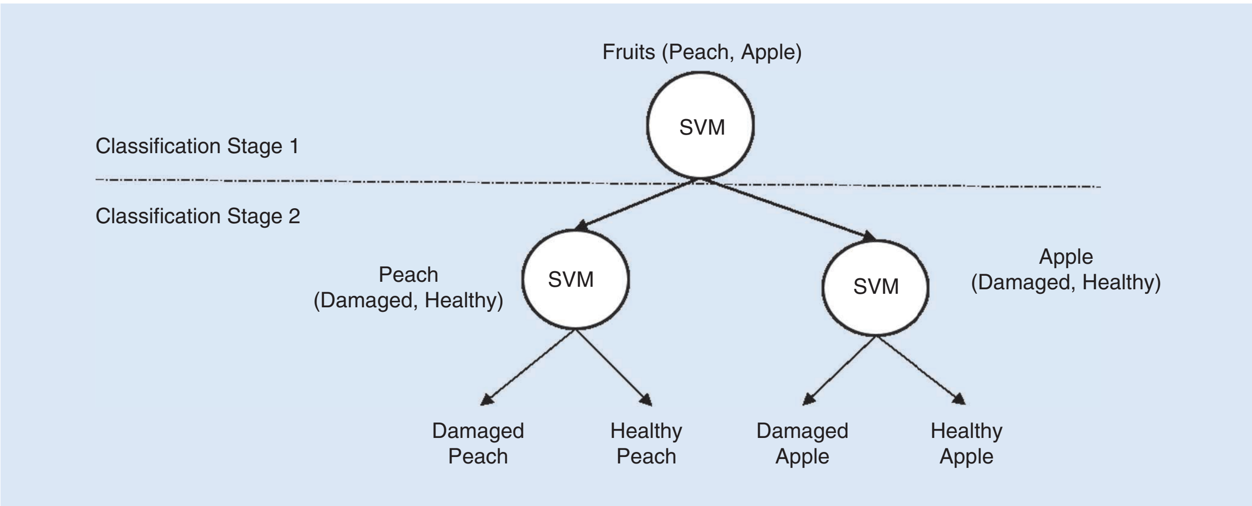

FIGURE 4. The BDT-SVM architecture for sorting healthy from damaged fruits.

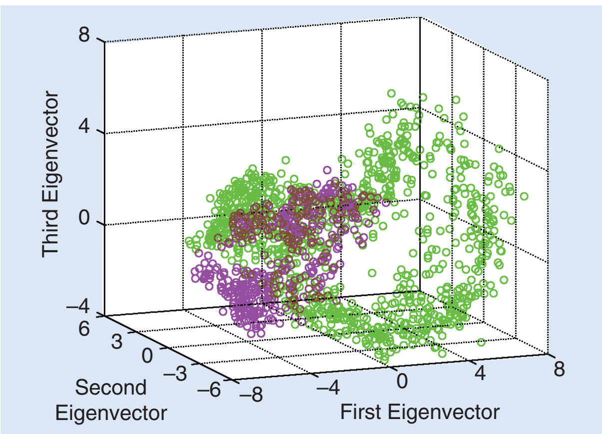

FIGURE 5. A projection of 80% of the training data set on the three most significant eigenvectors. Purple: Peach class; Green: Apple class.

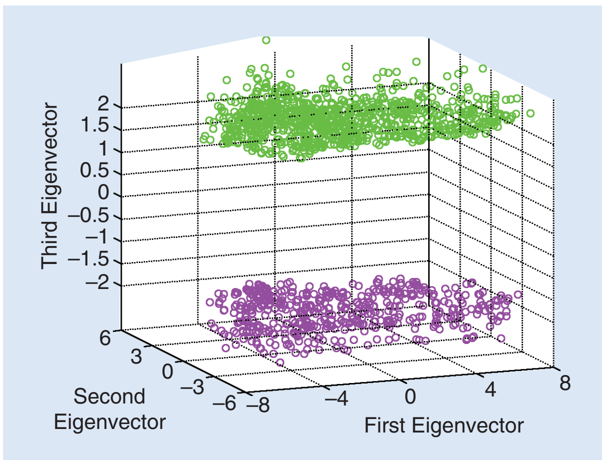

FIGURE 6. The classification results on the training data set. $(C^*, \gamma^*) = (100, 0.01)$. Purple: Peach class; Green: Apple class.

as shown in Table 2, exhibits an accuracy of 100% and values of SPEC and SENS of 1.

The second classification stage is more straightforward since we perform two SVM classifications for which we can reuse the optimal parameters found on peaches and apples separately, as mentioned in the section "Results of First and Second Measurement Campaigns." The confusion matrices for the healthy apple, the damaged one, and the damaged peach of the test data set are shown in Tables 3, 4, and 5, respectively. We perform all our classification experiments on a PC with a core (TM) i5-pro CPU at 1.8 GHz 2.3 GHz. The training phase lasts 6–8 h. This long time is due to the fact that we use what is called the *grid-search* method to find the optimal hyperparameters C^* and γ^* within a grid of 1,200 pairs of (C, γ). The test of one sample is much faster, with 12 ms for a peach sample and 30 ms for an apple's sample.

The total accuracy of the multiclass SVM is the product of the performance of each classification stage and therefore strongly depends on the accuracy of the first stage. In this example, we obtain 100% of total accuracy and values of 1 for SPEC and SENS. However, if we repeat the classification process 30 times while keeping the optimal values of C^* and γ^* and changing the training and test data set, the mean accuracy is 98%. As explained previously, the order of the classification stage is important. To show the impact of the order, we permute and start off by classifying healthy and damaged fruits and then sorting apples and peaches. The total accuracy drops to 71%.

We have proved that the BDT-SVM is efficient for sorting out different features of the fruits, such as their type and state, provided that we know how to arrange the classification stages. There are modeling schemes to overcome this difficulty, but they are time-consuming. Since we deal with problems that are sensitive to the scattered field, i.e., the dielectric constant, we might use the physical variation of ϵ_r as a hint to choose adequately the order of the classification stages, starting with the highest contrast and ending up with the lowest one.

Our next objective is to investigate if a biclass SVM can sort damaged from healthy apples even if they are of different varieties. In addition, we decide to move to low-THz frequencies in view of the realization of the real system. Indeed, our measurement system actually makes use of a vector network analyzer (VNA) that will be replaced with a frequency-modulated continuous wave (FM-CW) radar for practical implementation. One of the main challenges is the tradeoff between the size of the antenna and the image resolution. The increase of frequency naturally reduces the size of the antenna, which improves the system's compactness. Additionally, the lateral resolution is improved due to wavelength reduction.

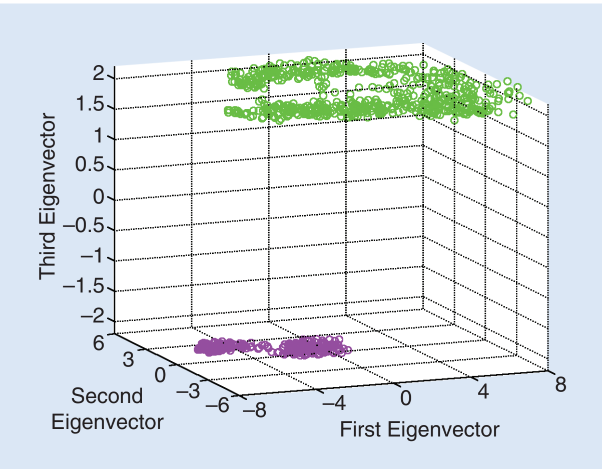

FIGURE 7. The classification results on the test data set. $(C^*, \gamma^*) = (100, 0.01)$. Purple: Peach class; Green: Apple class.

APPLES SORTING IN D-BAND

Here, we work with the apples of the third campaign. Although the increase of frequency has some significant advantages, the penetration depth decreases and there is a risk that the impinging wave remains at the surface of the fruit. This would reduce the advantage of low-THz-waves toward optics since only surface damages would be detectable. To check that the apple is responding, we have conducted range measurements as described in the next section.

TABLE 2. THE CONFUSION MATRIX OF SORTING FRUITS.

Fruit	Peaches	Apples
Peaches	303	0
Apples	0	606

TABLE 3. THE CONFUSION MATRIX FOR THE SMALL HEALTHY APPLE.

Apples	Healthy	Damaged
Healthy	303	0
Damaged	0	0

TABLE 4. THE CONFUSION MATRIX FOR THE SMALL DAMAGED APPLE.

Apples	Healthy	Damaged
Healthy	0	0
Damaged	0	303

TABLE 5. THE CONFUSION MATRIX FOR PEACH 2 (DAMAGED).

Peaches	Healthy	Damaged
Healthy	0	0
Damaged	0	303

RANGE MEASUREMENTS

If the impinging wave passes through the apple, we can estimate some fundamental parameters, such as its size or its average permittivity with a radar measurement. We set the scanner at a fixed position above the apple and apply the following equation:

$$\epsilon_r = (d_{\text{meas}}/d_{\text{apple}})^2, \tag{4}$$

where

- ϵ_r is the *average* permittivity (real part only). We call it the *average* permittivity because the apple is not homogeneous.
- d_{apple} is the real diameter of the apple, which is known in our case.
- d_{meas} is the measured diameter of the apple, which will be estimated with a threshold on the measurements as explained below.

The principle of the measurement and the spatial-frequency processing [56] are quite simple and mimic the FM-CW radar. In FM-CW radar measurements, the range resolution depends on the frequency band whereas the maximum unambiguous distance depends on the parameters of the FM-CW frequency ramp. To get the best range resolution, we set the VNA to the maximum bandwidth of the D-band, which spreads from 110 to 170 GHz, leading to a range resolution of 2.5 mm. We take 401 frequency samples for one measurement. In our measurements, the apple stands on a Rohacell tower located 491 mm from a probe antenna. It has a strong reflection at mm- and sub-mm-waves and contributes to the response of the apple. This is close to the final application where the fruits will be on a conveyor belt that will also contribute to the measured scattered field. Due to the very high space attenuation, the measurement level is close to the noise floor of our system. Moreover, the response of the apple is not strong enough to be directly extracted from the postprocessed radar measurement.

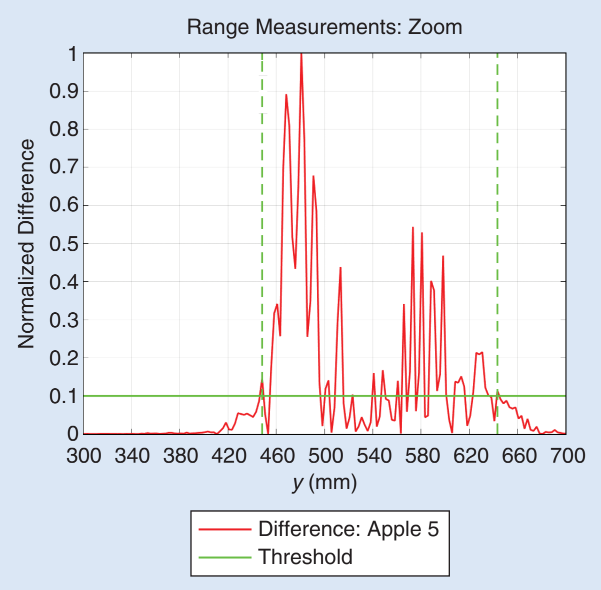

FIGURE 8. The radar signature of a healthy apple of the Golden variety.

To overcome this difficulty, we measure each apple twice. First, we place the apple in a vertical position at the center of the Rohacell tower. Second, we place it horizontally. The only change between both measurements is the apple's positioning. Performing a complex subtraction of the postprocessed radar measurements and applying a threshold that keeps 90% of the signal, we extract the apple's signature over distance, as shown in Figure 8. To ease the readability, we have normalized the curve to the maximum value of the difference. Using the limits given by the threshold (dotted green lines) and accounting for the accuracy of the measurement (range resolution), we find $d_{\text{meas}} = 195 \pm 2.5$ mm, corresponding to $\epsilon_r = 7.3 \pm 0.2$.

This value is much lower than the value of the pulp measured at low frequency [39], but this order of magnitude is consistent with the measured values of the dielectric constant of the water [57] at 100 GHz, which is one of the main components of the apple. We have repeated this experiment with damaged apples and the values range from 7 to 6.5 with no significant difference (which means that the difference is within the 0.2 error margin) between damaged Golden and Gala apples. The difference between healthy and damaged apples is tenuous but still there, and we can carry on our investigation on how the SVM classifier performs with D-band measurements.

D-BAND IMAGES

Figure 9(a)–(e) shows the five apples used for the third measurement campaign. We have selected the damaged apples according to their rotting stage.

- Apple 1 is almost completely rotten.
- Apple 2 has a damaged area 3 cm in diameter. It is not easy to see in the photo. Because it does not affect the appearance of the skin. Indeed, the pulp underneath the skin is smashed. We have chosen this damage to see if low-THz images can outperform optical images.
- Apple 3 has a large damaged area, easy to recognize in the photo.
- Apple 4 has a small damaged area, which is approximately 3 mm and was chosen because its size is of the same order of magnitude as the D-band wavelengths.
- Apple 5 is healthy.

Figure 10 shows the images obtained at 126 GHz for the five apples. It is impossible to work directly with these images for a human eye, contrary to optical ones. However, the images are different for each apple, which is a good starting point for an automatic classification scheme. One of the key points of automatic classification is the ability to produce enough data. That is why we have decided to add the frequency diversity to the spatial one. First, we have to check the images to confirm that there is a difference over frequency. Once again the comparison by human eye is not straightforward as shown in Figure 11(a) and (b). However, if we compute the image difference without changing the dynamic range of the color scale (30 dB in all cases) the response of the apple is clear, and the noise is almost entirely suppressed. Moreover, we will not work with the image differences since the noise is important and useful to the SVM classifier. Indeed, the noise has different realizations depending on the frequency that help the classifier to distinguish the useful signal.

SVM RESULTS IN D-BAND

Once we obtain the D-band images, we follow the workflow as described in Figure 1. We start with placing all Golden apples in the training data set (apples 4 and 5) and the Gala (apples 1–3) in the test. We reoptimize the hyperparameters (C, γ) during the training. Their values are $C^* = 100$ and $\gamma^* = 0.1$, where γ^*

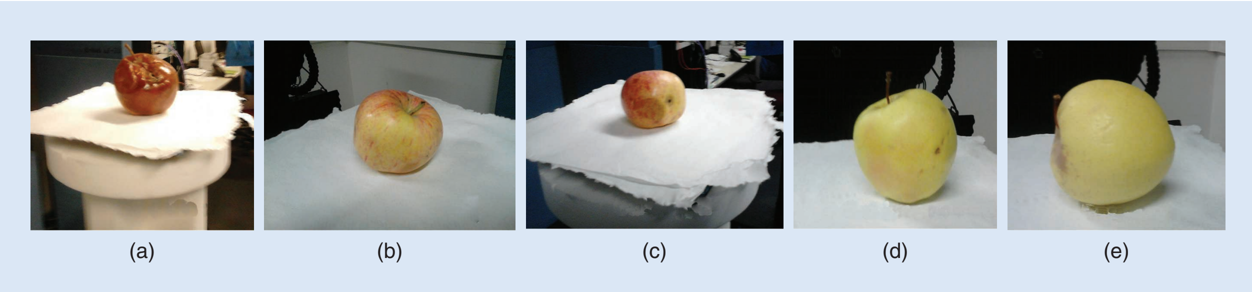

FIGURE 9. The five apples used for the third measurement campaign: (a) Gala apple 1, (b) Gala apple 2, (c), Gala apple 3, (d) Golden apple 4, and (e) Golden apple 5.

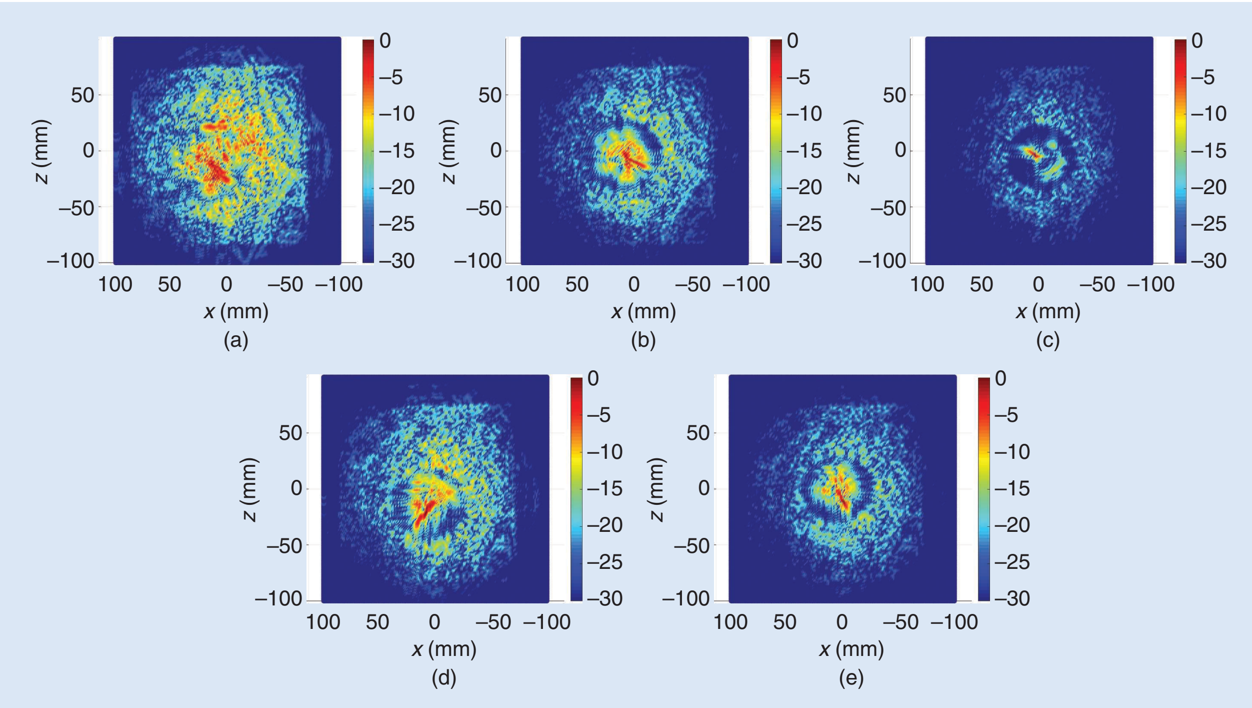

FIGURE 10. The mm-wave images of the five apples over the total scan area: (a) Gala apple 1, 126 GHz; (b) Gala apple 2, 126 GHz; (c) Gala apple 3, 126 GHz; (d) Golden apple 4, 126 GHz; and (e) Golden apple 5, 126 GHz.

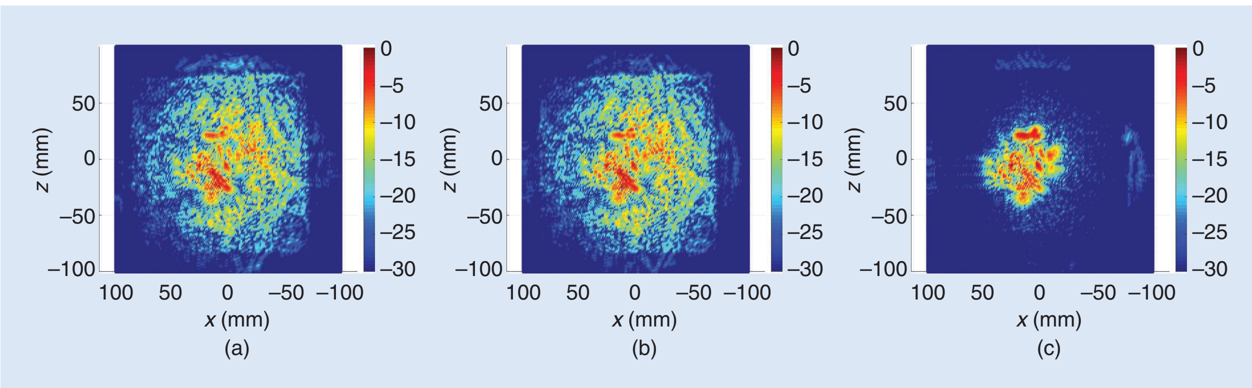

FIGURE 11. The mm-wave images of apple 1 over the total scan area: (a) 124 GHz, (b) 126 GHz, and (c) frequency difference image.

TABLE 6. THE CONFUSION MATRIX FOR SORTING DAMAGED APPLES 1–3.

Apples	Healthy	Damaged
Healthy	0	0
Damaged	0	909

TABLE 7. THE CONFUSION MATRIX FOR SORTING DAMAGED APPLES 2–4.

Apples	Healthy	Damaged
Healthy	0	0
Damaged	0	909

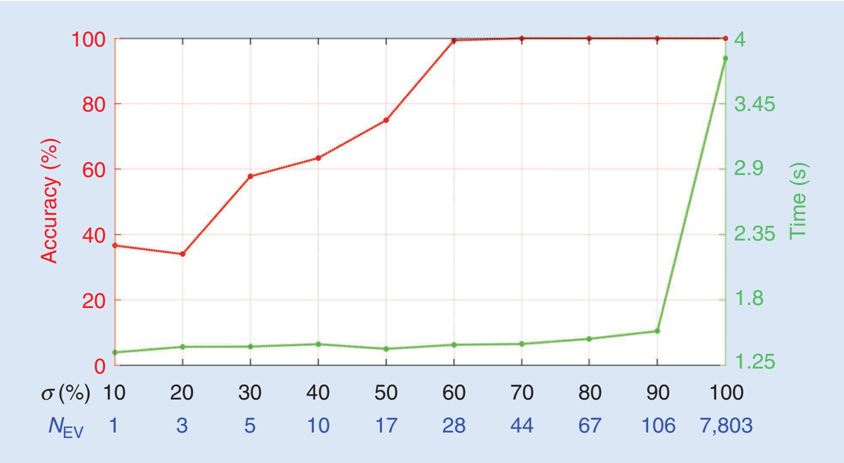

FIGURE 12. The evolution of the accuracy and the computation time for the training data set with the number of eigenvectors N_{EV}.

is significantly higher than in the W-band, which means that the dispersion of the data is larger. It is not surprising since the noise floor is about 10 dB higher than on the W-band, hence the noise level is greater. This also proves the importance of the frequency as an input parameter. In other words, it is crucial to search for new hyperparameters every time the frequency range changes. The confusion matrix in Table 6 shows that all three apples were successfully placed in the damaged category. To get a complete validation of the performance, we had to conduct a second test while mixing Golden and Gala apples in the training and the test data sets.

In the second test, we train with the damaged Gala apple 1 and with the healthy Golden apple 5, and we test the other apples. Without changing the hyperparameters, the confusion matrix in Table 7 shows an accuracy of 100%. From both tests we conclude that we can successfully sort out apples from different varieties without having to perform a multiclass SVM. This is consistent with the results of the average permittivity found in the section "Range Measurements," where the difference between healthy and damaged apples was greater than the one between the apple's varieties. The test of one sample lasts 5.9 s using the same PC described previously. It is about five times faster than for the multiclass SVM because of the more complex architecture of the SVM-BDT.

As a prospective study, we have also investigated the possibility of reducing the degrees of freedom of our training data set for decreasing the computation time. For this, we perform a PCA and keep only the N_{EV} first eigenvectors. Reducing the number of eigenvectors means that we keep only a certain percentage of the information enclosed in the original signal. This percentage is described in the variance (σ), which is the horizontal scale of curve (in black) but we also display the number of corresponding eigenvectors (in blue). Results are shown in Figure 12. As expected, both the accuracy and the computation time increase with the number of eigenvectors. However, we obtain the best accuracy as soon as the variance reaches 70%, which means that we can keep only 44 eigenvectors instead of the 7,083 contained in the original signal. The computation time drops from 3.8 s to 1.35 s (a factor of 3).

The computational load could be further reduced while selecting less than 44 eigenvectors. But this would be done at the expense of a breakdown in terms of accuracy, because it would be less than 100%.

CONCLUSIONS

In this article, we have carried on our investigations on a nondestructive method for sorting healthy from damaged fruits at mm- and low-THz waves. We start off with the scattering measurements. Then we process them and apply a nonlinear SVM with RBF kernel. The performance of the SVM classifier relies on the choice of the hyperparameters C and γ. Our investigations lead us to the following conclusions.

- We have been able to retrieve a difference between the average permittivity of healthy and damaged apples in the low-THz region thanks to range measurements. This justifies the use of images built from the measured scattered field.
- Compared to our previous work, we have successfully increased the frequency band up to the D-band (instead of the W-band). This allows us to get closer to an industrial application that requires a high resolution and a compact system.
- We were able to show that the change of frequency band has an important impact on the value of γ with an order of magnitude of 10 between the W- and D-bands.
- We were able to show that the variety of the apples does not influence the performance of the classifier in recognizing a damaged apple and that the system can be used regardless of the apple's variety.
- We have successfully overcome the limitation encountered in our previous work when we wanted to sort out

the damaged peaches/apples from the healthy ones. For this, we moved from biclass to multiclass SVM and use a BDT Scheme. The mean accuracy is 98%.

However, we hope to improve our method in the following directions:

- use data fusion to tackle the problem of several samples moving on the conveyor belt. In this prospect, a CDD camera would help to count the number of samples on the belt, which would serve as additional input of the SVM.
- produce more complex scenarios such as the detection of damages when the fruit is hidden under the foliage.
- develop optimization techniques to find the optimal values of SVM hyperparameters for improving the computation time of the training.
- take advantage of alternative MLA used in image processing, such as DNNs [58].

We have carried on our investigations on a nondestructive method for sorting healthy from damaged fruits at mm- and low-THz waves.

AUTHOR INFORMATION

Flora Zidane (flora.zidane@univ-cotedazur.fr) is currently pursuing a Ph.D. degree in electronics at the Université Côte d'Azur, Nice, France. Her research interests include millimeter-wave imaging and machine learning algorithms.

Jérôme Lanteri (jerome.lanteri@univ-cotedazur.fr) is currently an associate professor at the Université Côte d'Azur, CNRS, Electronics, Antennas, and Telecommunications Laboratory, Nice, France. His research interests include integrated antennas, reflectarrays and transmitarrays at millimeter-wave frequencies for nondestructive testing, medical or security applications. He is a Member of IEEE.

Julien Marot (julien.marot@univ-amu.fr) is currently an associate professor at Aix Marseille Université, CNRS, Institut Fresnel, Marseille, France. His research interests include applied image processing, multidimensional signal processing, and array processing.

Laurent Brochier (laurent.brochier@univ-cotedazur.fr) is currently an engineer at the Université Côte d'Azur, CNRS, Electronics, Antennas and Telecommunications Laboratory, Nice, France. His research focuses on microwave and millimeter-wave measurements of radiating structures.

Nadine Joachimowicz (nadine.joachimowicz@paris7.jussieu.fr) is currently an associate professor with the Sorbonne University Electrical Engineering and Electronics Laboratory, Paris. Her research interests concern direct and inverse scattering problems and she is currently involved in the development of phantoms for microwave imaging systems. She is a Member of IEEE.

Hélène Roussel (helene.roussel@sorbonne-universite.fr) is a professor at the Sorbonne University, Electrical Engineering and Electronics Laboratory, Paris. Her research interests include numerical methods and fast algorithms for electromagnetic scattering and radiation in complex media. She is a Member of IEEE.

Claire Migliaccio (claire.migliaccio@univ-cotedazur.fr) is currently a professor at the Université Côte d'Azur, CNRS, Electronics, Antennas and Telecommunications Laboratory, Nice, France. Her research interests include wideband antennas, millimeter-wave systems from antenna design to radar and imaging techniques. She is a Member of IEEE.

REFERENCES

[1] J. A. Abbott, "Quality measurement of fruits and vegetables," *Postharvest Biol. Technol.*, vol. 15, no. 3, pp. 207–225, 1999. doi: 10.1016/S0925-5214(98)00086-6.

[2] P. B. Elorza and M. R. Altisent, "Instrumentación de la calidad en frutas y hortalizas frescas," *Horticultura Internacional*, vol. 8, no. 29, pp. 14–20, 2000.

[3] C. Javanaud, "Applications of ultrasound to food systems," *Ultrasonics*, vol. 26, no. 3, pp. 117–123, 1988. doi: 10.1016/0041-624X(88)90001-7.

[4] A. E. Solovchenko, O. V. Avertcheva, and M. N. Merzlyak, "Elevated sunlight promotes ripening-associated pigment changes in apple fruit," *Postharvest Biol. Technol.*, vol. 40, no. 2, pp. 183–189, 2006. doi: 10.1016/j.postharvbio.2006.01.013.

[5] S. Bureau et al., "Rapid and non-destructive analysis of apricot fruit quality using FT-near-infrared spectroscopy," *Food Chem.*, vol. 113, no. 4, pp. 1323–1328, 2009. doi: 10.1016/j.foodchem.2008.08.066.

[6] M. Kopf, R. Gruna, T. Längle, and J. Beyerer, "Evaluation and comparison of different approaches to multi-product brix calibration in near-infrared spectroscopy," in *Proc. Int. Conf. Optical Characterization Materials (OCM)*, 2017, pp. 129–136.

[7] L. S. Magwaza and U. L. Opara, "Investigating non-destructive quantification and characterization of pomegranate fruit internal structure using x-ray computed tomography," *Postharvest Biol. Technol.*, vol. 95, pp. 1–6, Sept. 2014. doi: 10.1016/j.postharvbio.2014.03.014.

[8] N. Kotwaliwale, K. Singh, A. Kalne, S. N. Jha, N. Seth, and A. Kar, "X-ray imaging methods for internal quality evaluation of agricultural produce," *J. Food Sci. Technol.*, vol. 51, no. 1, pp. 1–15, 2014. doi: 10.1007/s13197-011-0485-y.

[9] E. G. Barcelon, S. Tojo, and K. Watanabe, "X-ray computed tomography for internal quality evaluation of peaches," *J. Agric. Eng. Res.*, vol. 73, no. 4, pp. 323–330, Aug. 1999. doi: 10.1006/jaer.1999.0409.

[10] B. Diezma-Iglesias, M. Ruiz-Altisent, and P. Barreiro, "Detection of internal quality in seedless watermelon by acoustic impulse response," *Biosyst. Eng.*, vol. 88, no. 2, pp. 221–230, 2004. doi: 10.1016/j.biosystemseng.2004.03.007.

[11] D. Knorr, M. Zenker, V. Heinz, and D.-U. Lee, "Applications and potential of ultrasonics in food processing," *Trends Food Sci. Technol.*, vol. 15, no. 5, pp. 261–266, 2004. doi: 10.1016/j.tifs.2003.12.001.

[12] P. Chen, M. McCarthy, and R. Kauten, "NMR for internal quality evaluation of fruits and vegetables," *Trans. ASAE*, vol. 32, no. 5, pp. 1747–1753, 1989. doi: 10.13031/2013.31217.

[13] S. R. Bandi, A. Varadharajan, and A. Chinnasamy, "Performance evaluation of various statistical classifiers in detecting the diseased citrus leaves," *Int. J. Eng. Sci. Technol.*, vol. 5, no. 2, pp. 298–307, 2013.

[14] C. Zhang, C. Guo, F. Liu, W. Kong, Y. He, and B. Lou, "Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine," *J. Food Eng.*, vol. 179, pp. 11–18, June 2016. doi: 10.1016/j.jfoodeng.2016.01.002.

[15] W. Guo, F. Zhao, and J. Dong, "Nondestructive measurement of soluble solids content of kiwifruits using near-infrared hyperspectral imaging," *Food Anal. Method*, vol. 9, no. 1, pp. 38–47, 2016. doi: 10.1007/s12161-015-0165-z.

[16] S. R. Dubey and A. S. Jalal, "Detection and classification of apple fruit diseases using complete local binary patterns," in *Proc. 2012 3rd Int. Conf. Computer and Communication Technology*, pp. 346–351. doi: 10.1109/ICCCT.2012.76.

[17] G. ElMasry, N. Wang, and C. Vigneault, "Detecting chilling injury in red delicious apple using hyperspectral imaging and neural networks," *Postharvest Biol. Technol.*, vol. 52, no. 1, pp. 1–8, 2009. doi: 10.1016/j.postharvbio.2008.11.008.

[18] D. Guyer and X. Yang, "Use of genetic artificial neural networks and spectral imaging for defect detection on cherries," *Comput. Electron. Agric.*, vol. 29, no. 3, pp. 179–194, 2000. doi: 10.1016/S0168-1699(00)00146-0.

[19] S. O. Nelson, “Electrical properties of agricultural products-a critical review,” *Trans. ASAE*, vol. 16, no. 2, pp. 0384–0400, 1973. doi: 10.13031/2013.37527.
[20] S. Trabelsi and S. Nelson, “Microwave sensing technique for nondestructive determination of bulk density and moisture content in unshelled and shelled peanuts,” *Trans. ASABE*, vol. 49, no. 5, pp. 1563–1568, 2006. doi: 10.13031/2013.22030.
[21] C. Bernou, D. Rebière, and J. Pistré, “Microwave sensors: A new sensing principle. Application to humidity detection,” *Sens. Actuator B, Chem.*, vol. 68, nos. 1–3, pp. 88–93, 2000. doi: 10.1016/S0925-4005(00)00466-4.
[22] B. Jackson and T. Jayanthy, “A novel method for water impurity concentration using microstrip resonator sensor,” in *Proc. Recent Advances Space Technology Services and Climate Change 2010 (RSTS & CC-2010)*, pp. 376–379. doi: 10.1109/RSTSCC.2010.5712872.
[23] M. A. M. Yunus and S. C. Mukhopadhyay, “Novel planar electromagnetic sensors for detection of nitrates and contamination in natural water sources,” *IEEE Sensors J.*, vol. 11, no. 6, pp. 1440–1447, 2010. doi: 10.1109/JSEN.2010.2091953.
[24] T. Nacke, A. Barthel, C. Pflieger, U. Pliquett, D. Beckmann, and A. Göller, “Continuous process monitoring for biogas plants using microwave sensors,” in *Proc. 2010 12th Biennial Baltic Electronics Conf.*, pp. 239–242. doi: 10.1109/BEC.2010.5630738.
[25] R. Blakey, O. Korostynska, A. Mason, and A. Al-Shamma’a, “Real-time microwave based sensing method for vegetable oil type verification,” *Proc. Eng.*, vol. 47, pp. 623–626, Sept. 9–12, 2012. doi: 10.1016/j.proeng.2012.09.224.
[26] L. Farina et al., “Microwave imaging technology for in-line food contamination monitoring,” in *Proc. 2019 IEEE Int. Symp. Antennas and Propagation and USNC-URSI Radio Science Meeting*, July 2019, pp. 817–818. doi: 10.1109/APUSNCURSINRSM.2019.8889369.
[27] G. Angiulli, M. Cacciola, and M. Versaci, “Microwave devices and antennas modelling by support vector regression machines,” *IEEE Trans. Magn.*, vol. 43, no. 4, pp. 1589–1592, 2007. doi: 10.1109/TMAG.2007.892480.
[28] M. Cacciola, G. Megali, S. Calcagno, M. Versaci, and F. C. Morabito, “Support vector machine for modelling design parameters in circular and triangular microstrip patch antennas,” in *Computational Intelligence Business and Economics Proc. MS’10 Int. Conf.* London: World Scientific, 2010, pp. 657–664. doi: 10.1142/9789814324441_0077.
[29] Z. Zheng, X. Chen, and K. Huang, “Application of support vector machines to the antenna design,” *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 21, no. 1, pp. 85–90, 2011. doi: 10.1002/mmce.20491.
[30] Q.-H. Zhang, B.-X. Xiao, and G.-Q. Zhu, “Inverse scattering by dielectric circular cylinder using support vector machine approach,” *Microw. Opt. Technol. Lett.*, vol. 49, no. 2, pp. 372–375, 2007. doi: 10.1002/mop.22131.
[31] A. Massa, G. Oliveri, M. Salucci, N. Anselmi, and P. Rocca, “Learning-by-examples techniques as applied to electromagnetics,” *J. Electromagn. Waves Appl.*, vol. 32, no. 4, pp. 516–541, 2018. doi: 10.1080/09205071.2017.1402713.
[32] M. Salucci, L. Tenuti, G. Oliveri, and A. Massa, “Efficient prediction of the EM response of reflectarray antenna elements by an advanced statistical learning method,” *IEEE Trans. Antennas Propag.*, vol. 66, no. 8, pp. 3995–4007, 2018. doi: 10.1109/TAP.2018.2835566.
[33] A. Kerhet, M. Raffetto, A. Boni, and A. Massa, “A SVM-based approach to microwave breast cancer detection,” *Eng. Appl. Artif. Intell.*, vol. 19, no. 7, pp. 807–818, 2006. doi: 10.1016/j.engappai.2006.05.010.
[34] M. Salucci, G. Oliveri, and A. Massa, “Real-time electrical impedance tomography of the human chest by means of a learning-by-examples method,” *IEEE J. Electromagn., RF Microw, Med. Biol.*, vol. 3, no. 2, pp. 88–96, 2019. doi: 10.1109/JERM.2019.2893217.
[35] S. Aminikhanghahi, S. Shin, W. Wang, S. H. Son, and S. I. Jeon, “An optimized support vector machine classifier to extract abnormal features from breast microwave tomography data,” in *Proc. 2014 Conf. Research Adaptive and Convergent Systems*, pp. 111–115. doi: 10.1145/2663761.2664230.
[36] T. Yilmaz et al., “Machine learning aided diagnosis of hepatic malignancies through in vivo dielectric measurements with microwaves,” *Phys. Med. Biol.*, vol. 61, no. 13, pp. 5089–5102, 2016. doi: 10.1088/0031-9155/61/13/5089.
[37] M. Salucci et al., “Real-time NDT-NDE through an innovative adaptive partial least squares SVR inversion approach,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6818–6832, 2016. doi: 10.1109/TGRS.2016.2591439.
[38] M. Salucci et al., “A nonlinear Kernel-based adaptive learning-by-examples method for robust NDT/NDE of conductive tubes,” *J. Electromagn. Waves Appl.*, vol. 33, no. 6, pp. 669–696, 2019. doi: 10.1080/09205071.2019.1572546.
[39] F. Zidane, L. Brochier, J. Lanteri, N. Joachimowicz, H. Roussel, and C. Migliaccio, “Damaged apple sorting with mm-w imaging and non-linear support vector machine,” *IEEE Trans. Antennas Propag.*, to be published.
[40] V. Torres et al., “Monitoring water status of grapevine by means of THz waves,” *J. Infrared, Millim., Terahertz Waves*, vol. 37, no. 5, pp. 507–513, 2016. doi: 10.1007/s10762-016-0269-6.
[41] H. B. Zhang, K. Mitobe, and N. Yoshimura, “Application of terahertz imaging to water content measurement,” *Jpn. J. Appl. Phys.*, vol. 47, no. 10, pp. 8065–8070, 2008. doi: 10.1143/JJAP.47.8065.
[42] Y. Ogawa, S. Hayashi, N. Kondo, K. Ninomiya, C. Otani, and K. Kawase, “Feasibility on the quality evaluation of agricultural products with terahertz electromagnetic wave,” in *Proc. 2006 ASAE Annu. Meeting*, p. 1. doi: 10.13031/2013.20868.
[43] A. Zahid et al., “Machine learning driven non-invasive approach of water content estimation in living plant leaves using terahertz waves,” *Plant Methods*, vol. 15, p. 138, Nov. 2019. doi: 10.1186/s13007-019-0522-9.
[44] A. Ren, A. Zahid, M. A. Imran, A. Alomainy, D. Fan, and Q. H. Abbasi, “Terahertz sensing for fruit spoilage monitoring,” in *Proc. 2019 2nd Int. Workshop Mobile Terahertz Systems (IWMTS)*, pp. 1–4. doi: 10.1109/IWMTS.2019.8823735.
[45] A. Ren et al., “Machine learning driven approach towards the quality assessment of fresh fruits using non-invasive sensing,” *IEEE Sensors J.*, vol. 20, no. 4, pp. 2075–2083, 2020. doi: 10.1109/JSEN.2019.2949528.
[46] F. Nsengiyumva, C. Migliaccio, L. Brochier, J. Lanteri, J.-Y. Dauvignac, and C. Pichot, “90 GHz, 3-D scattered field measurements for investigation of foreign object debris,” *IEEE Trans. Antennas Propag.*, vol. 67, no. 9, pp. 6217–6222, 2019. doi: 10.1109/TAP.2019.2922746.
[47] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, 1979. doi: 10.1109/TSMC.1979.4310076.
[48] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 2013.
[49] T. Fletcher, “Support vector machines explained,” Tutorial Paper, Mar. 2009. [Online]. Available: http://sutikno.blog.undip.ac.id/files/2011/11/SVM-Explained.pdf
[50] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2001.
[51] S. Amari and S. Wu, “Improving support vector machine classifiers by modifying Kernel functions,” *Neural Netw.*, vol. 12, no. 6, pp. 783–789, 1999. doi: 10.1016/S0893-6080(99)00032-5.
[52] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, 2002. doi: 10.1109/72.991427.
[53] G. Madzarov, D. Gjorgjevikj, and I. Chorbev, “A multi-class SVM classifier utilizing binary decision tree,” *Informatica*, vol. 33, no. 2, pp. 225–233, 2009.
[54] H. Jawad et al., “Microwave modeling and experiments for non destructive control improved quality of fruit,” in *Proc. 2017 IEEE Conf. Antenna Measurements Applications (CAMA)*, Dec. 2017, pp. 124–127. doi: 10.1109/CAMA.2017.8273375.
[55] E. Pérez-Castaño et al., “Comparison of different analytical classification scenarios: Application for the geographical origin of edible palm oil by sterolic (NP) HPLC fingerprinting,” *Anal. Methods*, vol. 7, no. 10, pp. 4192–4201, 2015. doi: 10.1039/C5AY00168D.
[56] B. Michael, W. Menzel, and A. Gronau, “A real-time close-range imaging system with fixed antennas,” *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 12, pp. 2736–2741, 2000. doi: 10.1109/22.899038.
[57] A. A. Barba and M. D’Amore, “Relevance of dielectric properties in microwave assisted processes,” in *Microwave Materials Characterization*. London: IntechOpen, 2012, ch. 6. [Online]. Available: http://www.intechopen.com/books/microwave-materials-characterization/relevance-of-dielectric-properties-in-microwave-assisted-processes
[58] A. Massa, D. Marcantonio, X. Chen, M. Li, and M. Salucci, “DNNs as applied to electromagnetics, antennas, and propagation: A review,” *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2225–2229, 2019. doi: 10.1109/LAWP.2019.2916369.