



HAL
open science

A polarized hyperspectral imaging system for in vivo detection: multiple applications in sunflower leaf analysis

J.L. Xu, Alexia Gobrecht, Daphné Heran, Nathalie Gorretta, M. Coque, A.A. Gowen, R. Bendoula, D. Sun

► To cite this version:

J.L. Xu, Alexia Gobrecht, Daphné Heran, Nathalie Gorretta, M. Coque, et al.. A polarized hyperspectral imaging system for in vivo detection: multiple applications in sunflower leaf analysis. *Computers and Electronics in Agriculture*, 2019, 158, pp.258-270. 10.1016/j.compag.2019.02.008 . hal-02609247

HAL Id: hal-02609247

<https://hal.inrae.fr/hal-02609247>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **A polarized hyperspectral imaging system for *in vivo* detection:**
2 **multiple applications in sunflower leaf analysis**

3 Jun-Li Xu¹, Alexia Gobrecht², Daphné Héran², Nathalie Gorretta², Marie Coque³, Aoife A.
4 Gowen¹, Ryad Bendoula^{*2} and Da-Wen Sun^{*1}

5 ¹School of Biosystems and Food Engineering, University College Dublin, Belfield, Dublin 4,
6 Ireland.

7 ²ITAP, IRSTEA, Montpellier SupAgro, Université Montpellier, France

8 ³Biogemma SAS, Domaine de Sandreau, Mondoville Blagnac, France

9
10 **Abstract**

11 This study aims to investigate the potential of an original polarized hyperspectral imaging
12 (HSI) setup in the spectral domain of 400-1000 nm for sunflower leaves in real-world.
13 Dataset 1 includes hypercubes of sunflower leaves in two varieties with different life growth
14 stages, while Dataset 2 is comprised of healthy and contaminated sunflower leaves suffering
15 from powdery mildew (PM) and/or septoria leaf spot (SLS). Cross polarised (R_{\perp}), parallel
16 polarised (R_{\parallel}) reflectance signals, R_{BS} ($R_{\parallel}+R_{\perp}$) and R_{SS} ($R_{\parallel}-R_{\perp}$) spectra were obtained and
17 used to develop partial least squares-discriminant analysis (PLS-DA) models. Surface
18 information played an important role in separating two varieties of leaves due to the fact that
19 the best model performance was achieved by using R_{SS} mean spectra, while both surface and
20 subsurface were equally important in classifying leaves between two major growth stages
21 because model of R_{BS} mean spectra outperformed other models. The best classification

*Corresponding author. E-mail: ryad.bendoula@irstea.fr

*Corresponding author dawen.sun@ucd.ie; Websites: <http://www.ucd.ie/refrig>; <http://www.ucd.ie/sun>.

22 model for disease detection was achieved by using pixel R_{\perp} spectra with the correct
23 classification rate (CCR) of 0.963 for both cross validation and prediction, meaning that
24 subsurface spectral features were the most important to detect infected leaves. The resulting
25 classification maps were also displayed to visualize the distribution of the infected regions on
26 the leaf samples. The overall results obtained in this research showed that the developed
27 polarized-HSI system coupled with multivariate analysis has considerable promise in
28 agricultural real-world applications.

29 **Key words:** Hyperspectral imaging; Polarized light spectroscopy; Visible and near infrared;
30 Sunflower leaf; PLSR.

31

32 **Introduction**

33 Hyperspectral imaging (HSI) refers to a technique that can provide both spatial and
34 spectral information by integrating two classical optical sensing technologies of imaging and
35 spectroscopy into one system (Ferrari et al. 2013). It was mainly applied on remote sensing in
36 the early 70's. In recent years, it started blooming in many different disciplines, e.g.,
37 pharmaceutical research (Gendrin et al. 2008), food science (Xu et al. 2015, Gowen et al.
38 2007) and agriculture (Ravikanth et al. 2015). HSI technique acquires spectral data at each
39 pixel of an image forming a three-dimensional array of data with two spatial dimensions and
40 one spectral dimension. As a result, it allows not only to identify and/or quantify the chemical
41 components of the analyzed sample, but also to map their spatial distribution. According to
42 Beer-Lambert law, the concentration of an absorber is directly proportional to the sample
43 absorbance, which is the basis of using spectroscopic signal. Nevertheless, the computed
44 absorbance could be a bad approximation of the Beer-Lambert law absorbance when dealing
45 with the turbid samples containing scattering centers. Highly scattering materials account for
46 additive and multiplicative effects, producing non-linearity in the absorbance-concentration

47 relationship. When this phenomenon dominates the spectra formation, the chemically related
48 absorbance can be severely overlapped by the physically related information (Gobrecht et al.
49 2015). In order to reduce scattering effects, Bendoula et al. (2015) proposed to combine
50 polarized light with near infrared (NIR) spectroscopy using the principle of polarization
51 subtraction. The polarized light spectroscopy method was introduced to reduce the effects of
52 multi-scattering on the measured signal based on the wave theory of light (Backman et al.
53 1999). Polarized hyperspectral imaging system was proposed by Vasefi et al. (2014) as a new
54 multimode dermoscope to accurately determine the spatial distribution of melanin and
55 hemoglobin oxygenation in a skin lesion. As reported, this system separated the contribution
56 of superficial melanin in order to quantify the deep melanin relative concentration so that
57 oxy- and deoxy-hemoglobin distribution can be accurately assessed. Cross-polarization HSI
58 in line scanning configuration was developed to remove glare from the acquired
59 hyperspectral reflectance images of various agricultural products: aubergine, Granny Smith
60 apple, and Royal Gala apple (Nguyen-Do-Trong et al. 2016). Their results showed that cross-
61 polarization enabled to improve the quality of the acquired hyperspectral reflectance images
62 through reducing the unwanted variability due to specular reflection and improving the signal
63 to noise ratios in the data to a factor from 1.1 to 3 times depending on the wavelength regions.
64 More recently, Nkengne et al. (2018) tested a polarized hyperspectral imaging system for
65 repeatable and reproducible in vivo skin quantification of melanin, total hemoglobin and
66 oxygen saturation. Except from the aforementioned researches, the potential of polarized-HSI
67 system has not been fully explored in agriculture.

68 HSI has gained tremendous importance in agriculture with different applications, such as
69 plant disease (Bock et al. 2010, Mahlein et al. 2012), stress detection and yield estimation
70 (Zarco-Tejada et al. 2012, Serrano et al. 2000) and leaf nitrogen content (Vigneau et al. 2011).
71 When applied on the fresh plant leaves, HSI was mainly used to estimate nitrogen and

72 chlorophyll content, and detect different diseases (Jay et al. 2017, Yu et al. 2014). For
73 chlorophyll content prediction, reflectance variables in the red edge region were much better
74 indicators than most other indices (Xiaobo et al. 2011). The position of red edge defined as
75 maximum slope in the vegetation reflectance spectra between the red and near infrared
76 regions is considered as one of the best estimators for chlorophyll content because it is less
77 sensitive to disturbing factors. Some published spectral indices provided relatively poor
78 performance in predicting leaf chlorophyll content when applied across a wide range of
79 species and plant functional types (Xue and Yang 2009). With respect to plant disease
80 detection, Di Gennaro et al. (2016) attempted to discriminate healthy leaves from grapevine
81 leaf stripe disease (*Phaeomoniella chlamydospora*) contaminated leaves with an unmanned
82 aerial vehicles (UAVs) used to acquire very high spatial resolution data of vineyards fields.
83 Even though their resulted showed that normalized difference vegetation index (NDVI)
84 values were useful in discrimination, they claimed that such a method was only reliable if no
85 other factors affected leaf chlorophyll content. Indeed, as different biotic and abiotic stresses
86 may affect leaf chlorophyll content (e.g., nitrogen stress, pests, etc.), the NDVI is not able to
87 discriminate a specific disease from other stresses. For spectral analysis, fresh leaf tissue
88 creates more problems than dried, ground tissue. The cell structure of fresh plant material
89 scatters light as it passes through multiple air and water interfaces with different refractive
90 indices. These phenomena are likely to obscure the subtle absorption features caused by
91 bending and stretching of chemical bonds.

92 In this context, we proposed a new modality that combines light polarization with
93 hyperspectral imaging to illuminate sunflower leaves with both parallel and perpendicular
94 polarization images of the remitted light recorded. The objective of this study was to assess
95 the potential of different combination of images to characterize different leaf properties, i.e.

96 variety, growth stage and presence/absence of diseases. This polarized-HSI system was tested
97 on two datasets of hyperspectral images of sunflower leaves collected on *in vivo* plants.

98

99 **Materials and methods**

100 **2.1 Dataset 1: variety and growth stage discrimination**

101 Leaves of sunflower (France) were collected for experiment. All investigated plants were
102 grown in greenhouses of Irstea, France. The grow chamber is composed of multispectral
103 lighting controlled by the Herbro automaton (GreenHouseKeeper entreprise). Herbro PLC
104 measures the temperature and humidity every 30 seconds, and takes care of watering the
105 sunflower plants every 48 hours. In terms of lighting, the Herbro PLC manages the light
106 spectrum emitted by the four lamps in a time interval. The lamps are equipped with 5
107 channels corresponding to 5 wavelengths (450 nm, 560 nm, 660 nm, 730 nm and 6000 ° k).

108 In this work, two different genotypes labelled as B and R were used. For B variety, plants
109 of 8, 7 and 6 weeks old were recorded as B1, B2, B3, while R variety plants of 8, 7, 6, 5 and
110 4 weeks old were labelled as R1, R2, R3, R4 and R5. Three leaves were collected at the upper,
111 middle and lower parts of each plant. On each leaf, four regions of interest (ROIs) were
112 selected and cut: one close to petiole, one close to tip, two in the blade region located at each
113 side of the midrib. As a result, $8 \times 3 \times 4 = 96$ images were acquired in the first
114 measurement and parts of these were used as the calibration set for variety and growth stage
115 discrimination. To form a test set, one more leaf were collected at the middle of each plant
116 and four ROIs were also selected according to the aforementioned criterion, which makes
117 $8 \times 1 \times 4 = 32$ hypercubes in the second measurement. For the purpose of classifying plant
118 variety, calibration set was obtained by B1-3 and R1-3 images in the first measurement
119 excluding R4 and R5 images in order to avoid imbalanced class distribution. Test set 1

120 includes only B1-3 and R1-3 images in the second measurement and Test set 2 includes all
121 the R4 and R5 images from the first and second measurements. As a result, there are $6 \times 3 \times$
122 $4 = 72$ observations for the calibration set, $6 \times 1 \times 4 = 24$ for the Test set 1 and $2 \times 4 \times$
123 $4 = 32$ for the Test set 2. With respect to identify growing stage, it was noticed that B1-3 and
124 R1-3 have similar height, around 100 cm, while R4-5 were much shorter, around 80 cm.
125 Therefore, B1-3 and R1-3 were recognized as the same class: Old, while R4 and R5 as the
126 other class: Young. To make a balanced dataset, only R1-5 from the first measurement were
127 used as calibration set and R1-5 from the second measurement generated Test set 1, while all
128 the B1-3 formed Test set 2, which makes $5 \times 3 \times 4 = 60$ observations for the calibration set,
129 $5 \times 1 \times 4 = 20$ for the Test set 1 and $3 \times 4 \times 4 = 48$ for the Test set 2. The data partition
130 details are summarized in Table 1.

131 **2.2 Dataset 2: disease detection**

132 Powdery mildew (PM) is a fungal disease that affects a wide range of plants but is an
133 emerging disease in sunflowers. Contaminated sunflower leaves exhibited symptoms of white
134 to grey powdery patches. Severely infected areas lose luster, curls, turn yellow and may dry
135 up. Septoria leaf spot (SLS) is another disease affecting sunflower leaves with black necrosis
136 and yellow spots, that can be observed at all stages of plant growth. Black or brown streaks
137 are usually noticed in a typical SLS affected plant. In this work, 20 contaminated regions
138 were found and cut from leaves in the B and R varieties during different growth stage from 4-
139 8 weeks. Yellow or brown spots were found both on the healthy and diseased leaves. In order
140 to discriminate yellow area from the infected and healthy areas, pixels from yellow area were
141 also extracted and form a Yellow category. To test the developed polarized-HSI system, 1875
142 pixel spectra of healthy leaf, 625 of yellow spot, 1188 of PM and 695 SLS were extracted,
143 among which 67% were randomly selected as calibration set and the remaining as the test set.

144 **2.3 Polarized-HSI setup**

145 Spectral images of the prepared leaf samples were acquired in the reflectance mode by
146 using a laboratory-based line scanning hyperspectral imaging system coupled with light
147 polarization. The developed polarized-HSI system was composed of a light source, an
148 illumination optical setup, a translation rail, and a detection system. A Xenon arc lamp
149 (300W, Newport 6508, Newport Corporation, California, USA) was mounted in lamp
150 housing with an F/1 aspheric condenser (Newport 67005, Newport Corporation, California,
151 USA). The optical illumination system was composed of a linear grid polarizer (Thorlabs
152 WP12L-UB, Thorlabs, INC., Newton, USA), a plano-convex cylindrical lens (THORLABS
153 LJ1810L2, Thorlabs, INC., Newton, USA) and a dielectric mirror (Thorlabs BB2-E02,
154 Thorlabs, INC., Newton, USA) to shape the incident beam as a vertical line (typically 10 cm
155 x 4 cm) on the sample. The sample was placed on a translation rail, synchronized with the
156 acquisition software which can record images when sample was scanned under the
157 hyperspectral camera (NEO Hypspec VNIR-1600 with 30 cm-objective, Norsk Elektro Optikk
158 AS, Skedsmokorest, Norway). A part of the light backscattered by the sample reached a
159 second linear grid polarizer (Thorlabs WP25M-UB, Thorlabs, INC., Newton, USA), also
160 known as analyzer, mounted in a rotation mount in front of the hyperspectral camera. The
161 rotation of this polarizer allowed choosing the polarization state of the acquired signal: either
162 parallel to the incident polarization or perpendicular to it. Spectral data were acquired in the
163 400-1000 nm wavelength range at 3.6 nm intervals.

164 **2.4. Reflectance calibration**

165 Due to the imperfections of some components (e.g., light source, lens, spectrograph, and
166 camera) and different measurement environments, the acquired raw hyperspectral images
167 generally contain noises and artifacts. Many factors, such as non-uniform illumination, pixel-

168 to-pixel sensitivity variations of the detector, and dust on the lens, will contribute to different
 169 image artifacts, leading to the raw images not being suitable for quantitative analysis.
 170 Therefore, proper calibration procedure is necessary. For each sample, two hypercubes were
 171 acquired with the parallel and cross polarization modes where analyzer was set parallel ($I_{||raw}$)
 172 and perpendicular ($I_{\perp raw}$) with respect to the polarization of the illumination light,
 173 respectively. A diffuse reflectance white standard (Spectralon®, [SRS-50-010](#), Labsphere,
 174 about reflectance 50%) was used to standardize spectra from non-uniformities with two
 175 images, $I_{||white}$ and $I_{\perp white}$. Two images of dark current were also recorded and indicated as
 176 $I_{||dark}$ and $I_{\perp dark}$ for parallel and cross polarization modes, respectively. Reflectance
 177 calibration was performed by comparing the acquired images according to the following
 178 formulas adapted from Bendoula *et al.* (2015):

$$R_{||corrected} = \frac{I_{||raw} - I_{||dark}}{(I_{||white} + I_{\perp white}) - (I_{||dark} + I_{\perp dark})} \quad (1)$$

$$R_{\perp corrected} = \frac{I_{\perp raw} - I_{\perp dark}}{(I_{||white} + I_{\perp white}) - (I_{||dark} + I_{\perp dark})} \quad (2)$$

179 **2.5 Spectral computations**

180 According to Gobrecht *et al.* (2015), the backscattered reflectance (R_{BS}) is comprised of
 181 the calibrated parallel ($R_{||}$) and cross images (R_{\perp}), and therefore can be obtained by:

$$R_{BS} = R_{||} + R_{\perp} \quad (3)$$

182 At the same time, it should be noticed that the backscattered reflectance (R_{BS}) is the sum of
 183 the multi-scattered (R_{MS}) and the weakly scattered (R_{SS}) parts of light:

$$R_{BS} = R_{MS} + R_{SS} \quad (4)$$

184 Due to multiple scattering events, the linearly polarized incident light loses its initial
185 polarization state and oscillates randomly in all the planes. However, photons that have
186 undergone weakly scattering event retain their initial polarization status. Therefore, multi-
187 scattered light is isotropically depolarized and half of its intensity passes through the analyzer
188 when oriented parallel to the polarizer and the other half when oriented perpendicular, while
189 weakly scattered light retains initial polarization state and only passes through the analyzer
190 when oriented parallel to the polarizer. As a result,

$$R_{\perp} = \frac{1}{2} R_{MS} \quad (5)$$

$$R_{\parallel} = \frac{1}{2} R_{MS} + R_{SS} \quad (6)$$

191 Based on the aforementioned relations, the weakly scattered (R_{SS}) reflectance can be obtained
192 from light polarization subtraction:

$$R_{SS} = R_{\parallel} - R_{\perp} \quad (7)$$

193 The detailed theoretical background and inference process can be found in Gobrecht *et al.*
194 (2015). Technically, the backscattered reflectance (R_{BS}) which has been included for
195 comparison in this work is similar as reflectance from the regular (or traditional)
196 hyperspectral imaging. The rationale behind this can be found in Appendix A. Therefore, R_{BS}
197 spectra will be used to represent the performance of traditional HSI. It should be noted that
198 no spectral pre-processing methods were used in this work.

199 **2.6 Digital microscopic images**

200 Microscopic images can present visually differences among different leaves on a
201 microscale. The acquired information from microscopic images will help interpret spectral
202 profiles and can be linked to model performances. Therefore, a digital microscope (Olympus

203 BX43) was used to acquire microscopic images of B1, B3, R1 and R5. The images were
204 acquired with Transmitted Koehler Illuminator and x10 microscope objective. Leaf samples
205 were first cut from the central of each leaf, scanned by polarized-HSI, and placed on the
206 regular glass slide on which a drop of water was placed. It was ensured that leaf specimen
207 was flat and contact well with the glass slide. Excess water was removed with the paper towel.
208 Texture analysis of the microscopic images was conducted by applying entropy filter in this
209 work.

210 **2.7 Unsupervised multivariate analysis**

211 Principal components analysis (PCA) is one of the most widely used unsupervised
212 multivariate analysis techniques for hyperspectral image analysis. Normally, it is used as an
213 exploratory technique, because it enables the reduction of the many spectral dimensions to a
214 smaller number of principal component (PC) scores which capture the maximum variation in
215 the data. In this work, PCA was performed on Dataset 2 to investigate the image of the
216 infected sunflower leaf. The upper and lower surfaces of the leaf were both used, each
217 surface was scanned in two sequence to obtain parallel and cross images. Two images in the
218 same mode (either parallel or cross) from upper and lower surfaces were first unfolded and
219 then concatenated to form a single augmented matrix. PCA was applied on this augmented
220 matrix.

221 **2.8 Discriminant analysis and model performance evaluation**

222 Discriminant models were constructed by using partial least squares-discriminant
223 analysis (PLS-DA) between the X matrix containing the spectra for calibration, and the
224 corresponding Y matrix containing the belonging identity for each class. It should be noted
225 that Dataset 1 uses mean spectrum of each ROI while Dataset 2 uses pixel spectrum to form

226 X matrix. The performance of each developed model was evaluated by the correct
227 classification rate (CCR), which is expressed according to the following equation (Teye et al.
228 2013):

$$\text{CCR} = \frac{N_1}{N_2} \times 100\% \quad (8)$$

229 where N_1 is the number of correctly classified samples and N_2 is the total number of samples.
230 The optimal number of latent variables (LVs) was determined by classification error of cross
231 validation with 10-fold venetian blinds. In addition, classification map was also generated to
232 evaluate the performance of developed classification models. The tested hypercube was
233 unfolded such that the three-dimensional information was rearranged in two-dimensional
234 matrix where each row represents the spectrum from each pixel and each column refers to a
235 certain wavelength variable. The obtained model was applied on this two-dimensional matrix
236 to classify each pixel into one category. The obtained prediction matrix was transformed to
237 produce a classification map where each pixel was assigned to a predicted category.

238 **Results and discussion**

239 **3.1 Dataset 1: variety and growth stage discrimination**

240 **3.1.1 Microscopic image**

241 Fig. 1a shows the microscopic images of B1, B3, R1 and R5. Visually, B1 has a darker
242 shade of green than R1, indicating the color difference between B and R varieties. While the
243 young leaf (R5) has a lighter shade of green than the old leaf (R1), the difference between B3
244 and B1 is less obvious. The entropy filter can detect subtle variations in the local gray level
245 distribution and therefore it was used to extract texture information at green channel as shown
246 in Fig. 1b. As shown, the intricate leaf veins (vascular bundles) running across the surface of
247 the leaf are highlighted by the entropy filter. Vein distribution patterns differ markedly

248 between R5 and the rest of images (B1, B3 and R1). Old leaves (B1, B3 and R1) have well-
249 structured vein distribution, while young leaf (R5) has not developed a fully functioning vein
250 system. Fig. 1c displays the histogram distribution of intensity values in the green plane. All
251 histograms present a bimodal property with one peak for the cells with chloroplasts and the
252 other for vein. It is observed that the B variety has lower intensity for the peak of chloroplasts
253 than R variety, and young leaf has higher intensity than the older leaves. This is partly
254 because young leaves have lower photosynthetic capacity compared to the matured leaves
255 (Ölçer et al. 2001).

256 3.1.2 Spectral feature analysis

257 The mean reflectance spectra of cross, parallel, R_{BS} and R_{SS} were computed from each
258 calibration and test set of Dataset 1 and are exhibited in Fig. 2. The mean spectra of B and R
259 variety are shown in Fig. 2a and Fig. 2b for calibration set and Test set 1, respectively. It is
260 first noticed that all the cross, parallel and R_{BS} have similar spectral characteristics of the
261 “green plants” (Wu et al. 2008, Zhao et al. 2016). The reflectance is minimum in the blue
262 (around 450 nm) and red (around 670 nm) spectral domains and shows a peak in the green
263 (around 550 nm) wavelength region, indicating that leaves absorb red and blue light and
264 reflect green light. It then demonstrates a sharp edge from 670 nm to 740 nm (the so-called
265 “red edge”) before reaching a plateau in the NIR region (Filella and Penuelas 1994). When
266 focusing on the spectral difference between the two varieties, it is observed that B variety has
267 slightly higher reflectance in the green region, higher absorption in the red and blue regions,
268 lower reflectance after 740 nm in the calibration set (Fig. 2a). However, consistent
269 differences were not found in the Test set 1 (Fig. 2b). R_{SS} spectra also show a major
270 absorption peak around 670 nm (red), a minor absorption peak at 530 nm (blue) and
271 reflectance peak at 550 nm (green). However, the global reflectance values of R_{SS} have been
272 largely reduced compared to R_{\perp} , R_{\parallel} and R_{BS} . This is expected because with R_{SS} only a small

273 portion of the signal is selected: the weakly scattered one. Multi-scattered light vibrates in all
274 planes and half of its intensity passes through the analyzer when oriented parallel to polarizer
275 and the other half when oriented perpendicular, yet weakly scattered light retains its initial
276 polarization state and therefore only passes through the analyzer when positioned parallel to
277 polarizer, as described in Section 2.5. In this sense, R_{\perp} images obtained when the analyzer
278 was oriented perpendicular to polarizer shows the enhanced information from deeper layers
279 due to the rejection of superficial reflectance and preferentially selecting deeper penetrating
280 light, while the R_{SS} images preserve only the surface information after polarization
281 subtraction. When compared to R_{\perp} , R_{\parallel} or R_{BS} spectra, the R_{SS} spectra exhibit greater
282 differences between the two varieties and more consistency between the calibration and Test
283 set 1, with the B variety showing slightly higher reflectance values both in the green and red
284 regions.

285 With respect to spectral difference between young and old plant leaves (Fig. 2 (c) and (d)),
286 it can be noticed that cross reflectance of young leaves is lower than that of old ones in the
287 spectral domain of 400 to 700 nm both for calibration and Test set 1. This reflectance
288 difference is reduced for parallel and R_{BS} spectra. On the other hand, R_{SS} spectra demonstrate
289 an opposite trend with young leaves having higher reflectance in this spectral range. The
290 difference in the R_{SS} spectra is more pronounced compared to the other spectra types,
291 indicating that a marked difference between surface features of young and old leaves due to
292 the combined effect of color and physical structure as observed in the microscopic images in
293 Fig.1.

294 **3.1.3 Classification modelling of variety discrimination**

295 PLS-DA was applied to build classification models to classify B and R varieties and the
296 results are shown in Table 2. As shown, there is a higher prediction accuracy in Test 2 than

297 Test 1. Comparing model performances, it can be seen that the model developed with R_{SS}
298 spectra outperformed the others, while cross spectra resulted in the worst model performance
299 based on the two test datasets. In detail, the R_{SS} PLS-DA model yielded CCR of 0.750 for
300 Test set 1 and 0.906 for Test set 2, while the CCR of the R_{\perp} PLS-DA model were 0.667 and
301 0.844 for Test set 1 and 2, respectively. This result suggests that the major spectral difference
302 between B and R variety is mostly due to the superficial reflectance of leaves rather than
303 subsurface information.

304 R_{BS} and R_{SS} PLS-DA models were selected due to their better performances and then
305 applied on each pixel of the images to produce prediction maps. To approximate how the
306 selected leaves would appear to humans, their RGB pseudo-color images were produced by
307 the combined R_{BS} images at three wavelengths located at red (670 nm), green (550 nm), and
308 blue (460 nm) wavelengths. The pseudo-color images are displayed in the Fig. 3 and the
309 corresponding prediction maps built from R_{BS} and R_{SS} spectra are shown in Fig. 4a and b,
310 respectively. As shown, no obvious difference between R and B varieties can be observed by
311 visual inspection of the pseudo-color images. R5 seems to present a different vein structure
312 compared to the rest of leaves, which agrees with the microscopic images in Fig. 1. On top of
313 each prediction map, the predicted class for this ROI was achieved by applying the model on
314 its mean spectrum. As seen, among these 8 observations, 7 are correctly classified by using
315 both R_{BS} and R_{SS} PLS-DA models. It should be noted that all PLS-DA models were built
316 with mean spectra and, to create the prediction maps, were applied to the pixel spectra.
317 Therefore, it is reasonable that there are many incorrect predictions for pixels within one
318 sample. **The accuracy of pixel-wise prediction might be improved by spectral pre-processing**
319 **methods, such as smoothing.** For the B variety, a larger portion of pixels were predicted as B
320 instead of R variety. In addition, R4 and R5 prediction maps present mostly yellow pixels

321 (predicted as R), which supports the result from the Table 1 that prediction performance for
322 Test set 2 is better.

323 **3.1.4 Classification modelling of age discrimination**

324 PLS-DA modelling was also used to classify leaves into two growth stages (Old and
325 Young) and the results are exhibited in Table 3. Compared to variety discrimination, it is
326 more accurate to distinguish between these two leaf age stages. Interestingly, the worst
327 performance was observed on the model of R_{SS} with CCR of 0.750 for Test 1 and 0.854 for
328 Test 2, while the best performance was achieved by using R_{BS} spectra with high CCR of
329 1.000 for Test 1 and 0.970 for Test 2, implying that only surface information is not sufficient
330 to accomplish this task. At the same time, it can be noted that the cross model overall
331 performs better than parallel, especially for the Test 1 where cross achieves CCR of 1.000
332 while parallel has CCR of 0.800. Contrarily, higher CCR (equivalent to 1.000) is found in the
333 Test 2 by using parallel spectra, while it is 0.917 for cross. These results indicate that deep
334 information captured by cross spectra are crucial in discriminating between leaves of two
335 growth stages. However, superficial information captured by parallel spectra also play a part
336 in explaining the difference between two growth stages, especially when B variety leaves are
337 included (Test 2). As a consequence, when using the combined information from subsurface
338 and surface of leaf, R_{BS} spectra (the sum of multi-scatter and weakly scattered light) produces
339 the best performance in classification.

340 Likewise, the best two models (cross and R_{BS}) were employed to generate classification
341 maps of the 8 aforementioned observations and respectively displayed in Fig. 4c and d. As
342 observed, all the 8 observations are correctly classified by applying both PLS-DA models on
343 the mean spectra. Better performance of R_{BS} is confirmed by prediction maps: R1-3 images

344 have large areas wrongly predicted as Young class using the cross model. It is also noted that
345 vein structure is highlighted in the prediction maps via predicting vein pixels as the Old class.

346 **3.2 Dataset 2: disease detection**

347 **3.2.1 Spectral feature analysis**

348 Mean spectra of each class from healthy and contaminated leaves are shown in Fig. 5a and
349 b for calibration and test sets, respectively. There are obvious reflectance value differences
350 between different types. Compared to healthy leaves, cross, parallel and R_{BS} spectra of
351 yellow spot demonstrate higher reflectance in the visible spectral range with a broad
352 reflectance peak observed around 550 to 600 nm and narrow absorption peak at 670 nm. The
353 PM infected leaves share similar spectral shape as healthy leaves, but with higher reflectance
354 which is probably due to the white color of pustules. The SLS infected leaves have distinctive
355 spectral difference with the rest of spectra: the green peak around 550 nm is barely observed
356 and the sharp shoulder of red edge is lost. With respect to R_{SS} spectra, yellow and SLS
357 infected leaves show no resolved absorption peaks, while healthy and PM infected leaves
358 demonstrate absorption peak at 670 nm.

359 **3.2.2 Unsupervised PCA exploration**

360 The upper and lower surfaces of a contaminated leaf were scanned and PCA was
361 performed on the cross and R_{SS} hypercubes as shown in Fig. 6. Likewise, the RGB pseudo-
362 color images were created by the combined cross and R_{SS} images at three wavelengths of red
363 (670 nm), green (550 nm), and blue (460 nm) and exhibited in Fig. 6a. It should be noted that
364 only the pseudo-color image of the backscattered reflectance (R_{BS}) hypercube will appear
365 similar to their natural color in the real world, not cross, parallel or R_{SS} hypercubes. As can
366 be seen, there's some SLS infected dark spots spread out on the upper surface of cross
367 hypercube, while these spots are less visually observed at the corresponding locations of the

368 lower surface. In addition to SLS contamination, two major PM infected areas covered with
369 whitish powdery mycelial growth can be well observed on the upper surface. Interestingly,
370 PM and SLS contaminated regions are overlapped with some part of SLS infected area
371 covered by PM (SLS+PM). R_{SS} images display superficial texture information of the selected
372 leaf. PM spots can be also seen on the upper surface of R_{SS} image, while SLS specks could
373 not be found. It can be seen from Fig. 6a that the upper and lower surfaces of a leaf differ in
374 color and texture as confirmed by cross and R_{SS} images.

375 Fig. 6b shows the first three score images and loadings of cross images. **The first three PCs**
376 **have explained more than 99% of variance.** PC 1 loading implies this PC describes the major
377 spectral difference caused by red-edge with most negative loading at 680 nm and most
378 positive one at 750 nm. Loading values below 720 nm are negative and above positive. Based
379 on Fig. 5, it can be seen that healthy and PM infected leaves exhibit a sharp increase after 680
380 nm, while SLS infected spectrum increases slowly. In addition, SLS cross spectrum have
381 higher reflectance values than healthy leaf spectrum below 720 nm, and become lower when
382 above 720 nm due to a much smaller rate of increase. Combined with loading values and the
383 mean cross spectra from Fig. 5, it is not difficult to speculate that SLS spots are supposed to
384 present very negative values on the PC1 score images. As expected, SLS infected regions are
385 shown on the upper surface of leaves and their locations correspond well with the RGB
386 pseudo-color images. In addition, it is noticed that SLS infected regions overlapped with PM
387 regions are also revealed. More importantly, SLS spots are also displayed on the lower
388 surface though many of these spots cannot be visually detected in the pseudo-color image,
389 indicating that cross spectra reveal subsurface information from deep layers. It is also
390 observed the shape of the major SLS area on the left of midrib on the upper surface is the
391 same as that on the lower surface. It can also be concluded that SLS first occurs on the upper
392 surface due to bigger infected areas observed on the PC1 score image of upper surface. PC 2

393 mainly presents the spectral difference between SLS uncovered and covered by PM regions.
394 As shown in Fig. 5, PM cross spectrum is overall higher than SLS spectrum, the biggest two
395 differences are located around 550 and 720 nm. PC2 loading values are all negative with two
396 valleys evidenced at 550 and 720 nm. Therefore, SLS uncovered region should have high
397 positive values, while influenced by the top PM region should be negative. It is noted that
398 PC2 score images show the visible SLS infected regions on both upper and lower surfaces.
399 Nevertheless, PC3 is more complicated due to different characteristic properties shown on
400 score images. PC3 loading plot has the opposite shape as the PC2 loading, while PC3 loading
401 has both positive and negative values. The upper surface of PC3 highlights the whitish
402 powdery mycelial structure in negative values. Yet the features highlighted in very positive
403 value on both upper and lower surfaces are not straightforward to interpret. To further
404 confirm the aforementioned interpretation, threshold value equals to -0.5 was used to segment
405 all the SLS infected area on cross PC1 score image ($Mask_{PC1}$), threshold value of 0.25 to
406 identify the uncovered SLS infected area on PC2 score image ($Mask_{SLS}$), and threshold value
407 of -0.1 to detect all the whitish powdery mycelial structure on PC3 score image ($Mask_{PC3}$).
408 Mask of SLS infected region covered by PM (SLS+PM) should have both '1' in $Mask_{PC1}$
409 and $Mask_{PC3}$, while mask of PM infected region is supposed to have '1' in $Mask_{PC3}$ but '0'
410 in $Mask_{PC1}$. All the created masks can be seen in Fig. 7. Cross and R_{SS} mean spectra are also
411 shown in Fig. 7. The shapes of SLS and PM spectra (R_{\perp} and R_{SS}) are similar to those from
412 the mean of calibration set in Fig. 5. When SLS infected region is covered by PM, the mean
413 spectra (SLS+PM) are in the middle between SLS and PM spectra.

414 When it comes to PCA results of R_{SS} hypercubes, PC1 score images highlight vein
415 structure in both upper and lower surfaces. As seen, the surface texture is different between
416 the upper and lower surfaces: the upper surface seems smoother than the lower surface. PC2
417 and PC3 score images also show little difference between the contaminated and healthy

418 region, yet the overall images are noisy due to pixel subtraction. For a better comparison,
419 PCA scatter plots obtained from R_{\perp} and R_{SS} hypercubes are both shown in Fig. 8. It should
420 be noted that each dot from the plot corresponds to a pixel on the upper surface. Pixels
421 extracted from SLS infected, SLS infected region covered by PM (SLS+PM) and PM
422 infected masks are respectively highlighted in black, magenta and red colors. As indicated,
423 SLS and PM infected pixels form distinct clusters on R_{\perp} spectra while R_{SS} spectra are unable
424 to obtain separable clusters, indicating that depth information is vital for disease detection on
425 sunflower leaf.

426

427 3.2.3 Classification modelling of disease detection

428 Classification models were developed based on the pixel spectra of the leaf upper surface
429 and their results are shown in Table 4. The best model performance is noticed by using cross
430 spectra with CCR of 0.963 both for cross validation and prediction, followed by using R_{BS}
431 spectra with CCR of 0.948 for cross validation and 0.951 for prediction, respectively. R_{SS}
432 spectra contribute to the worst model prediction ability, which probably due to the noisy
433 characteristics after pixel subtraction. It is also noted that model built with parallel spectra
434 performs worse than that of R_{BS} spectra. As described in the Section 2.5, cross spectrum
435 equals to half of the multi-scattered light, and parallel spectrum is the sum of half multi-
436 scattered and weakly-scattered light. As a consequence, R_{BS} spectrum is the sum of multi-
437 scattered and weakly-scattered light, which means the influence of weakly-scattered light is
438 reduced in the R_{BS} spectrum compared to parallel spectrum. Since R_{BS} model outperforms
439 parallel model, it can be indicated that subsurface features are more important than superficial
440 information in detecting disease on sunflower leaves.

441 Classification maps were subsequently produced by using cross model on the leaf upper
442 surface due to the best performance. RGB pseudo-color images of 4 contaminated leaves are
443 generated on the R_{BS} reflectance image and shown in Fig. 9. It should be noted that Leaf 1 is
444 the same leaf that was used in unsupervised PCA exploration as shown in Fig. 6 and 7. PM
445 infected regions can be well observed as white patches on each leaf, while SLS infected
446 region is only visually observed on Leaf 1 based on the pseudo-color images. As can be seen,
447 slightly overestimation problem happens when it comes to predicting SLS infected pixels:
448 SLS+PM pixels are mostly predicted as SLS class and some nearby pixels without obvious
449 SLS symptom are wrongly predicted as SLS class. This classification model succeeds in
450 predicting PM pixels in Leaf 2, 3 and 4 and the whitish powdery mycelial structures are well
451 presented. However, some pixels belong to veins are misclassified as PM class.

452 **Conclusions**

453 This work investigated the feasibility and usefulness of an original polarized light
454 hyperspectral imaging setup as a noninvasive technique for sunflower leaf assessment. The
455 proposed polarized-HSI system has advantages compared with conventional HSI, enabling
456 collection and separation of backscattered reflectance into subsurface and surface spectral
457 features. Specifically, cross polarization image presents deep information due to the rejection
458 of superficial reflectance, while the R_{SS} spectrum, measuring weakly scattered light contains
459 mostly superficial information. Based on Dataset 1, it can be concluded that superficial
460 information contributes more in discriminating between B and R varieties. Both surface and
461 subsurface features are equally important in growth stage classification since the best model
462 was obtained from the R_{BS} spectra. With respect to the disease detection on Dataset 2, the
463 classification model performance indicates that subsurface information captured by cross
464 spectrum is the most desired feature in detecting powdery mildew and septoria leaf spot on
465 sunflower leaves. The combination of light polarization and hyperspectral imaging will

466 facilitate the enhancement of the current applications in proximal remote sensing and many
467 other disciplines. Many real-world applications will be investigated by using this proposed
468 polarized-HSI in the future work.

469 **Acknowledgements**

470 This work was conducted within the OPTIPAG Project supported by the grant ANR-16-
471 CE04-0010 from the French Agence Nationale de la Recherche. The authors would also like
472 to acknowledge China Scholarship Council (CSC) and University College Dublin (UCD) for
473 financial support of this study under CSC-UCD Scheme and the Georges Frêche
474 L'Association, for financial support.

475 **References**

- 476 Backman, V., Gurjar, R., Badizadegan, K., Itzkan, I., Dasari, R.R., Perelman, L.T. & Feld, M.S. (1999)
477 Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular
478 structures in situ. *IEEE Journal of Selected Topics in Quantum Electronics*, **5**, 1019-1026.
- 479 Bendoula, R., Gobrecht, A., Moulin, B., Roger, J.-M. & Bellon-Maurel, V. (2015) Improvement of the
480 chemical content prediction of a model powder system by reducing multiple scattering using
481 polarized light spectroscopy. *Applied spectroscopy*, **69**, 95-102.
- 482 Bock, C., Poole, G., Parker, P. & Gottwald, T. (2010) Plant disease severity estimated visually, by
483 digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in*
484 *Plant Sciences*, **29**, 59-107.
- 485 Di Gennaro, S.F., Battiston, E., Di Marco, S., Facini, O., Matese, A., Nocentini, M., Palliotti, A. &
486 Mugnai, L. (2016) Unmanned Aerial Vehicle (UAV)-based remote sensing to monitor
487 grapevine leaf stripe disease within a vineyard affected by esca complex. *Phytopathologia*
488 *Mediterranea*, **55**, 262.
- 489 Ferrari, C., Foca, G. & Ulrici, A. (2013) Handling large datasets of hyperspectral images: reducing data
490 size without loss of useful information. *Analytica Chimica Acta*, **802**, 29-39.
- 491 Filella, I. & Penuelas, J. (1994) The red edge position and shape as indicators of plant chlorophyll
492 content, biomass and hydric status. *International Journal of Remote Sensing*, **15**, 1459-1470.
- 493 Gendrin, C., Roggo, Y. & Collet, C. (2008) Pharmaceutical applications of vibrational chemical imaging
494 and chemometrics: a review. *Journal of pharmaceutical and biomedical analysis*, **48**, 533-553.
- 495 Gobrecht, A., Bendoula, R., Roger, J.-M. & Bellon-Maurel, V. (2015) Combining linear polarization
496 spectroscopy and the Representative Layer Theory to measure the Beer–Lambert law
497 absorbance of highly scattering materials. *Analytica chimica acta*, **853**, 486-494.
- 498 Gowen, A., O'donnell, C., Cullen, P., Downey, G. & Frias, J. (2007) Hyperspectral imaging—an
499 emerging process analytical tool for food quality and safety control. *Trends in Food Science &*
500 *Technology*, **18**, 590-598.
- 501 Jay, S., Gorretta, N., Morel, J., Maupas, F., Bendoula, R., Rabatel, G., Dutartre, D., Comar, A. & Baret,
502 F. (2017) Estimating leaf chlorophyll content in sugar beet canopies using millimeter-to
503 centimeter-scale reflectance imagery. *Remote Sensing of Environment*, **198**, 173-186.

- 504 Mahlein, A.-K., Steiner, U., Hillnhütter, C., Dehne, H.-W. & Oerke, E.-C. (2012) Hyperspectral imaging
505 for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods*,
506 **8**, 3.
- 507 Nguyen-Do-Trong, N., Keresztes, J.C., De Ketelaere, B. & Saeys, W. (2016) Cross-polarised VNIR
508 hyperspectral reflectance imaging system for agrifood products. *Biosystems Engineering*,
509 **151**, 152-157.
- 510 Nkengne, A., Robic, J., Seroul, P., Gueheunneux, S., Jomier, M. & Vie, K. (2018) SpectraCam®: A new
511 polarized hyperspectral imaging system for repeatable and reproducible in vivo skin
512 quantification of melanin, total hemoglobin, and oxygen saturation. *Skin Research and
513 Technology*, **24**, 99-107.
- 514 Ölçer, H., Lloyd, J.C. & Raines, C.A. (2001) Photosynthetic capacity is differentially affected by
515 reductions in sedoheptulose-1, 7-bisphosphatase activity during leaf development in
516 transgenic tobacco plants. *Plant Physiology*, **125**, 982-989.
- 517 Ravikanth, L., Singh, C.B., Jayas, D.S. & White, N.D. (2015) Classification of contaminants from wheat
518 using near-infrared hyperspectral imaging. *Biosystems Engineering*, **135**, 73-86.
- 519 Segtnan, V.H., Høy, M., Lundby, F., Narum, B. & Wold, J.P. (2009) Fat distribution analysis in salmon
520 fillets using non-contact near infrared interactance imaging: a sampling and calibration
521 strategy. *Journal of Near Infrared Spectroscopy*, **17**, 247-253.
- 522 Serrano, L., Filella, I. & Penuelas, J. (2000) Remote sensing of biomass and yield of winter wheat
523 under different nitrogen supplies. *Crop Science*, **40**, 723-731.
- 524 Teye, E., Huang, X., Dai, H. & Chen, Q. (2013) Rapid differentiation of Ghana cocoa beans by FT-NIR
525 spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular
526 and Biomolecular Spectroscopy*, **114**, 183-189.
- 527 Vasefi, F., Mackinnon, N., Saager, R.B., Durkin, A.J., Chave, R., Lindsley, E.H. & Farkas, D.L. (2014)
528 Polarization-sensitive hyperspectral imaging in vivo: a multimode dermoscope for skin
529 analysis. *Scientific Reports*, **4**, 4924.
- 530 Vigneau, N., Ecartot, M., Rabatel, G. & Roumet, P. (2011) Potential of field hyperspectral imaging as
531 a non destructive method to assess leaf nitrogen content in Wheat. *Field Crops Research*,
532 **122**, 25-31.
- 533 Wu, C., Niu, Z., Tang, Q. & Huang, W. (2008) Estimating chlorophyll content from hyperspectral
534 vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, **148**, 1230-
535 1241.
- 536 Xiaobo, Z., Jiyong, S., Limin, H., Jiewen, Z., Hanpin, M., Zhenwei, C., Yanxiao, L. & Holmes, M. (2011)
537 In vivo noninvasive detection of chlorophyll distribution in cucumber (*Cucumis sativus*)
538 leaves by indices based on hyperspectral imaging. *Analytica chimica acta*, **706**, 105-112.
- 539 Xu, J.L., Riccioli, C. & Sun, D.W. (2015) An Overview on Nondestructive Spectroscopic Techniques for
540 Lipid and Lipid Oxidation Analysis in Fish and Fish Products. *Comprehensive Reviews in Food
541 Science and Food Safety*, **14**, 466-477.
- 542 Xue, L. & Yang, L. (2009) Deriving leaf chlorophyll content of green-leafy vegetables from
543 hyperspectral reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**, 97-106.
- 544 Yu, K.-Q., Zhao, Y.-R., Li, X.-L., Shao, Y.-N., Liu, F. & He, Y. (2014) Hyperspectral imaging for mapping
545 of total nitrogen spatial distribution in pepper plant. *PloS one*, **9**, e116205.
- 546 Zarco-Tejada, P.J., González-Dugo, V. & Berni, J.A. (2012) Fluorescence, temperature and narrow-
547 band indices acquired from a UAV platform for water stress detection using a micro-
548 hyperspectral imager and a thermal camera. *Remote Sensing of Environment*, **117**, 322-337.
- 549 Zhao, Y.-R., Li, X., Yu, K.-Q., Cheng, F. & He, Y. (2016) Hyperspectral imaging for determining pigment
550 contents in cucumber leaves in response to angular leaf spot disease. *Scientific Reports*, **6**,
551 27790.
- 552

553

Table 1. Data partition for classification model development of Dataset 1.

Objective	Calibration set	Test set 1	Test set 2
Plant variety discrimination	B1-3 and R1-3 FM ($6 \times 3 \times 4 = 72$)	B1-3 and R1-3 SM ($6 \times 1 \times 4 = 24$)	R4-5 FM and SM ($2 \times 4 \times 4 = 32$)
Growth age discrimination	R1-5 FM ($5 \times 3 \times 4 = 60$)	R1-5 SM ($5 \times 1 \times 4 = 20$)	B1-3 FM and SM ($3 \times 4 \times 4 = 48$)

554 **Note:** FM: first measurement composed of B1-3, R1-5 (8 plants \times 3 locations \times 4 ROIs=96 images); SM: second measurement composed of B1-3, R1-5 (8
555 plants \times 1 location \times 4 ROIs=32 images).

556

557

558

559

Table 2. Performance of PLS-DA model to classify between two sunflower leaf varieties.

Methods	LV	Calibration	Cross validation	Test 1	Test2
R_{\perp}	10	0.944	0.914	0.667	0.844
R_{\parallel}	10	0.972	0.930	0.708	0.875

560

R_{BS}	11	0.972	0.943	0.708	0.875
R_{SS}	10	0.958	0.914	0.750	0.906

561

562

563

564

Table 3. Performance of PLS-DA model to classify between two sunflower leaf growth stages.

Methods	LV	Calibration	Cross validation	Test 1	Test2
R_{\perp}	7	0.933	0.900	1.000	0.917
R_{\parallel}	7	0.933	0.933	0.800	1.000
R_{BS}	7	0.933	0.933	1.000	0.970
R_{SS}	8	0.933	0.900	0.750	0.854

565

Table 4. Performance of PLS-DA model to classify between healthy leaf, yellow spots, SLS and PM infected pixels.

Methods	LV	Calibration	Cross validation	Test
R_{\perp}	6	0.963	0.963	0.963
R_{\parallel}	6	0.913	0.921	0.930
R _{BS}	6	0.948	0.948	0.951
R _{SS}	7	0.665	0.656	0.656

List of Figures

Fig. 1. Visualization of microscopic images of B1, B3, R1 and R5 (a); their corresponding texture images (b) and histograms (c) at green channel.

Fig. 2. Plot of mean spectra of two varieties for **calibration** set (a) and Test set 1 (b); of two growth stages for **calibration** set (c) and Test set 1 (d).

Fig. 3. RGB pseudo-color images of the ROI from B1-3 and R1-5 categories.

Fig. 4. Prediction maps to classify between two varieties by using R_{BS} (a) and R_{SS} (b) spectra; classify between two growth stages by using cross (c) and R_{BS} (d) spectra. **On top of each prediction map, the label of each sample is placed on the left side of the equal sign, the predicted class on the mean spectrum is sit on the right side of the equal sign. B class includes B1-3, while R class includes R1-5; Old class includes B1-3 and R1-3, while Young class includes R4-5.**

Fig. 5. Plot of mean spectra of heathy leaf, yellow spots, PM and SLS infected pixels for calibration (a) and test set (b).

Fig. 6. Visualization of RGB pseudo-color images **from SLS and PM infected sunflower leaf** (a); the obtained score images and loadings of PCA conducted on the hypercube of R_{\perp} (b) and R_{SS} (c). **The explained variance by each PC is indicated in the bracket on top of loading plot.**

Fig. 7. Obtained masks from thresholding PCA score images and the plot of mean spectra of SLS, SLS covered by PM (SLS+PM) and PM infected pixels.

Fig. 8. PCA scatter plot obtained from R_{\perp} and R_{SS} spectra. Each dot corresponds to a pixel on the upper surface of sunflower leaf. The explained variance by each PC is indicated in the bracket.

Fig. 9. Prediction map to classify SLS, PM, yellow spots and healthy leaf pixels by using cross PLS-DA model.

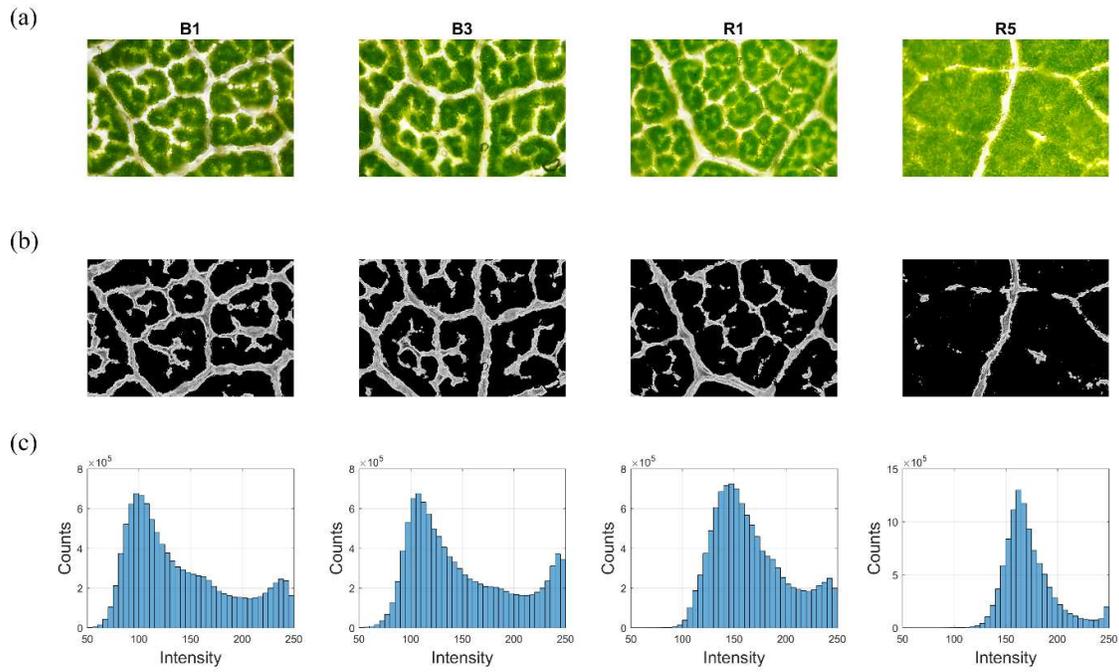


Fig. 1. Visualization of microscopic images of B1, B3, R1 and R5 (a); their corresponding texture images (b) and histograms at green channel (c).

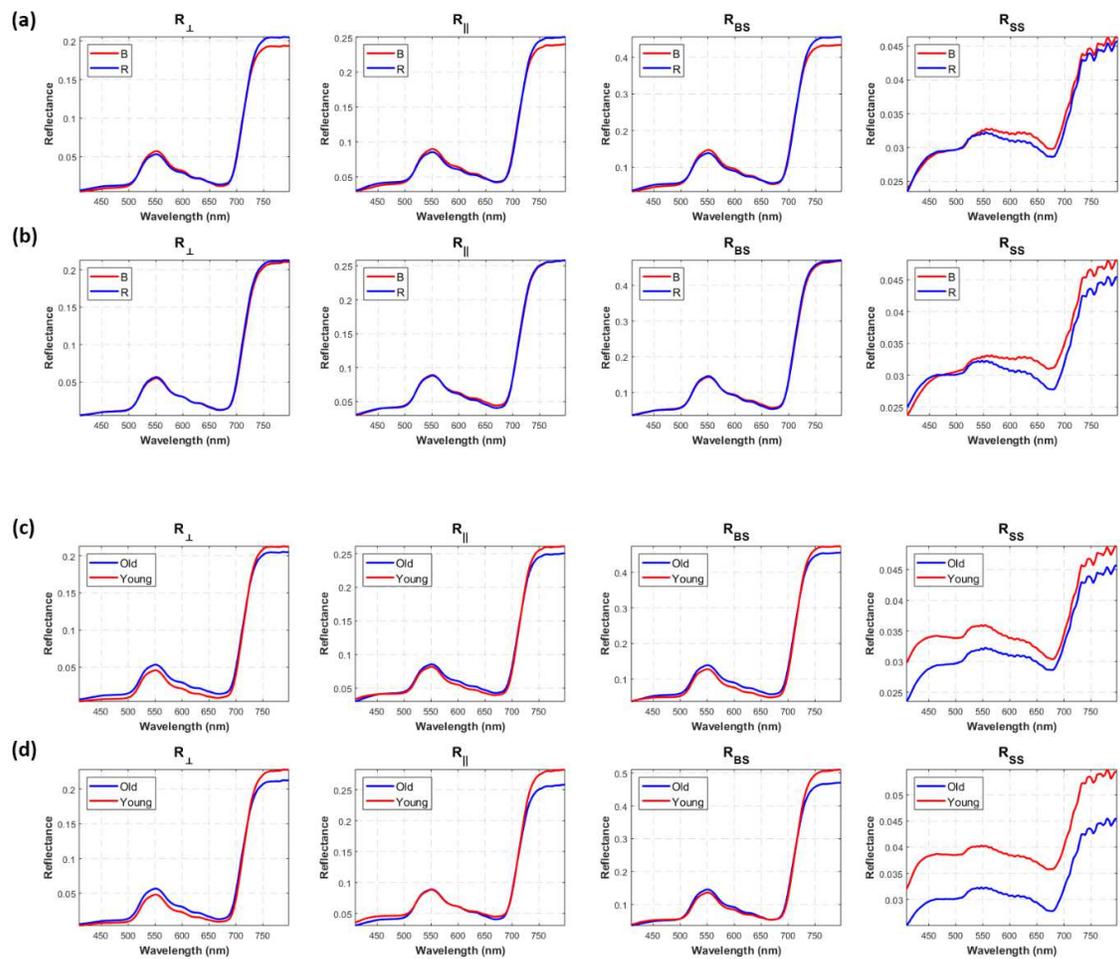
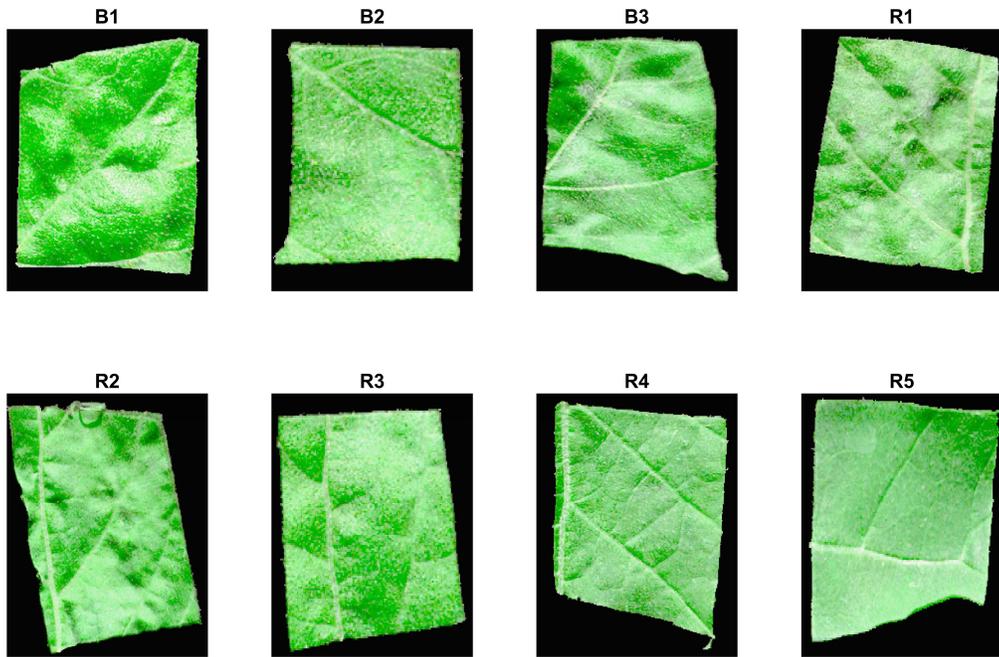


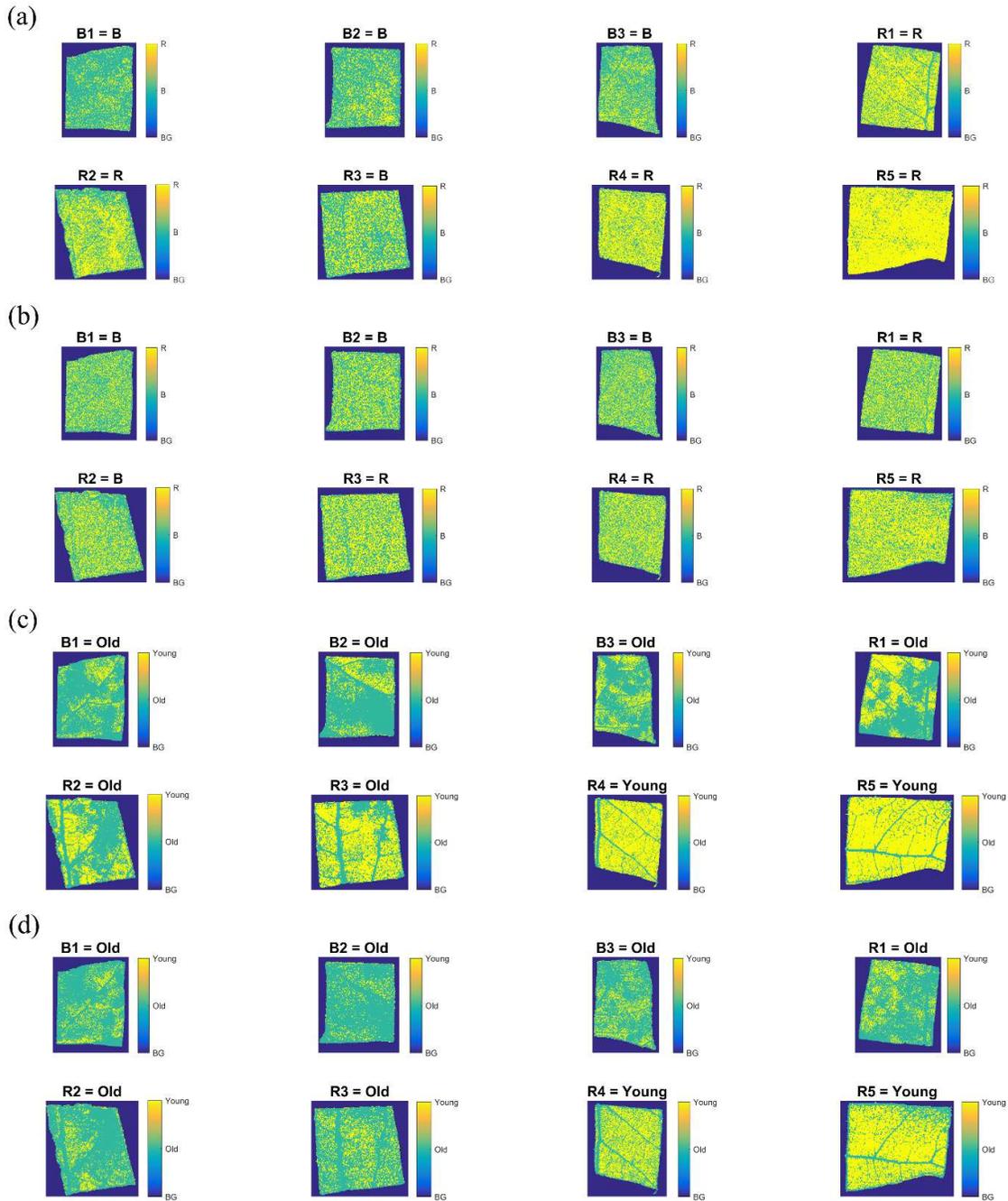
Fig. 2. Plot of mean spectra of two varieties for calibration set (a) and Test set 1 (b); of two growth stages for calibration set (c) and Test set 1 (d).



1

2 Fig. 3. RGB pseudo-color images of the ROI from B1-3 and R1-5 categories.

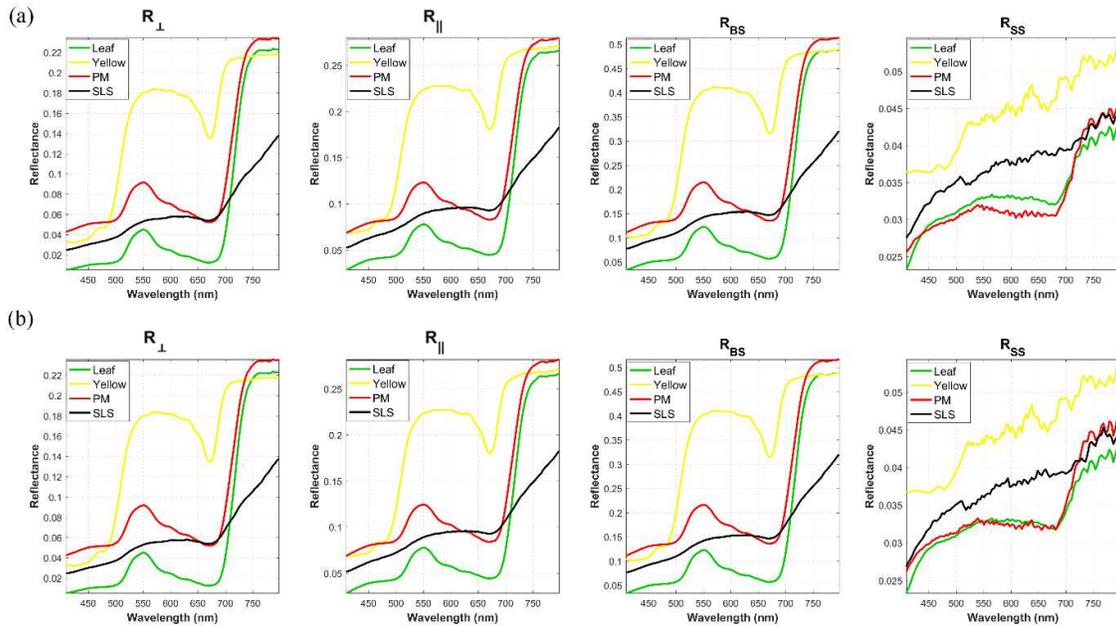
3



4
5
6
7
8
9
10
11
12

Fig. 4. Prediction maps to classify between two varieties by using R_{BS} (a) and R_{SS} (b) spectra; classify between two growth stages by using R_{\perp} (c) and R_{BS} (d) spectra. On top of each prediction map, the label of each sample is placed on the left side of the equal sign, the predicted class on the mean spectrum is sit on the right side of the equal sign. B class includes B1-3, while R class includes R1-5; Old class includes B1-3 and R1-3, while Young class includes R4-5.

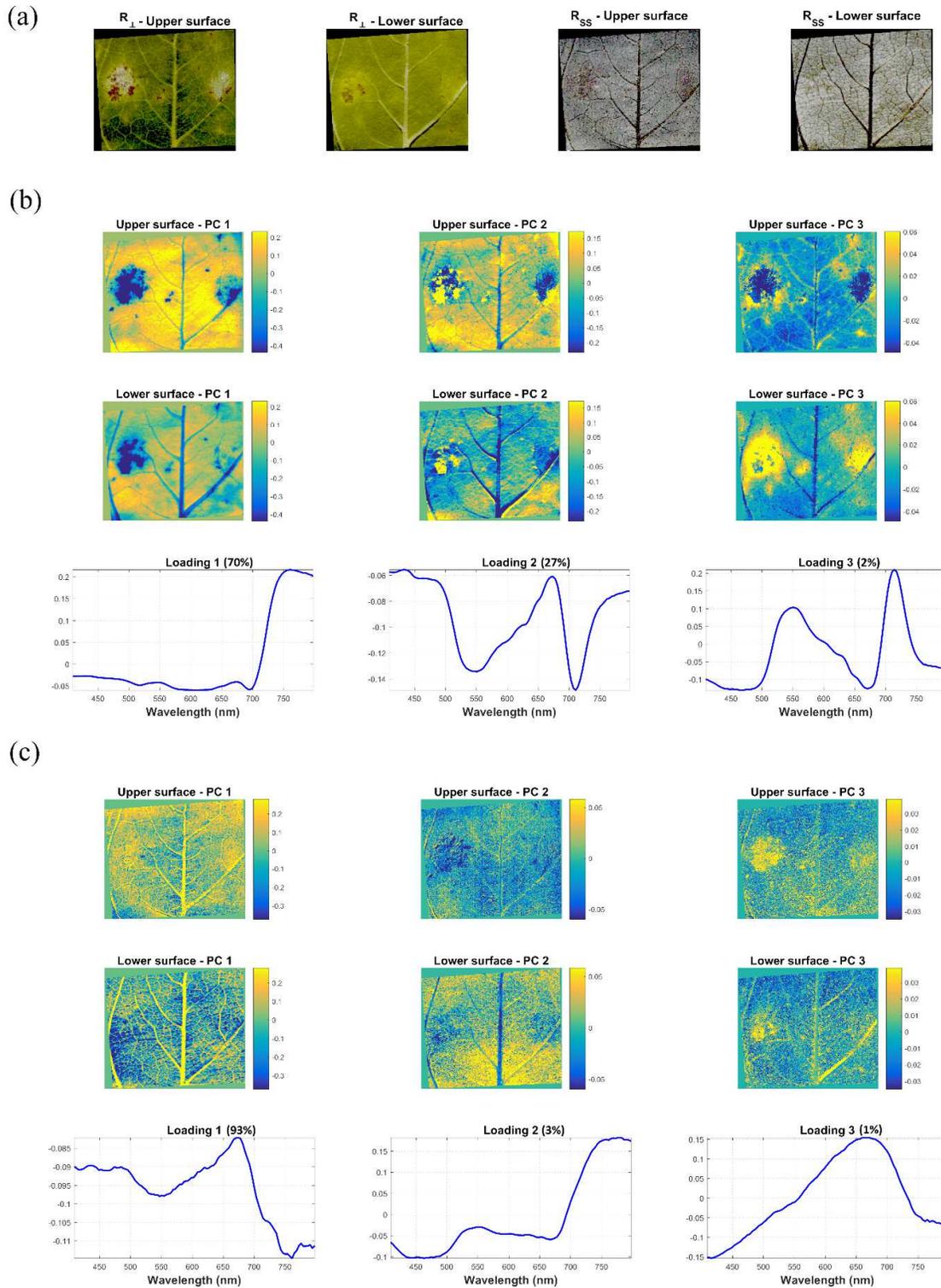
13



14

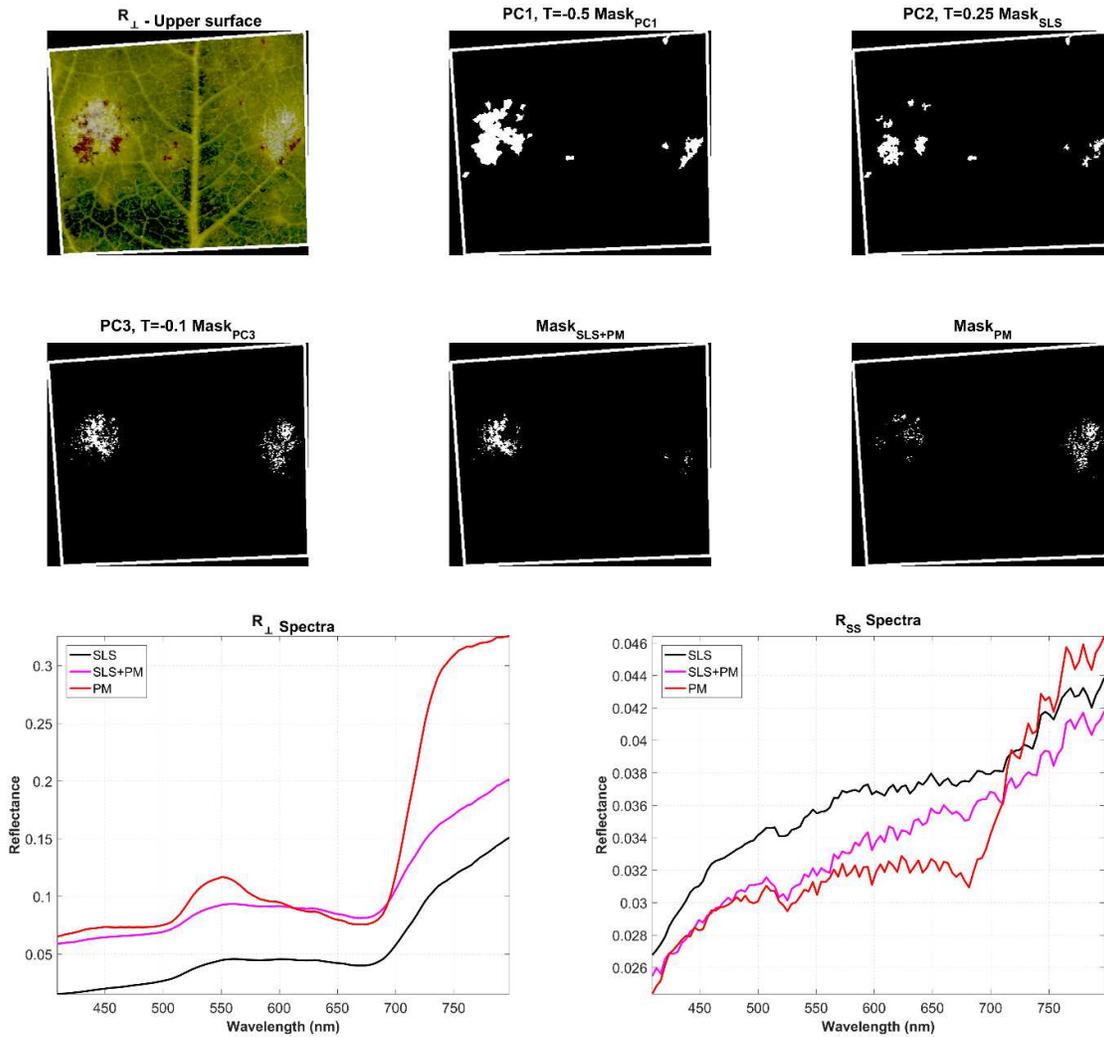
15 Fig. 5. Plot of mean spectra of healthy leaf, yellow spots, PM and SLS infected pixels for
16 calibration (a) and test set (b).

17



18

19 Fig. 6. Visualization of RGB pseudo-color images from SLS and PM infected sunflower leaf
 20 (a); the obtained score images and loadings of PCA conducted on the hypercube of R_{\perp} (b)
 21 and R_{SS} (c). The explained variance by each PC is indicated in the bracket on top of loading
 22 plot.

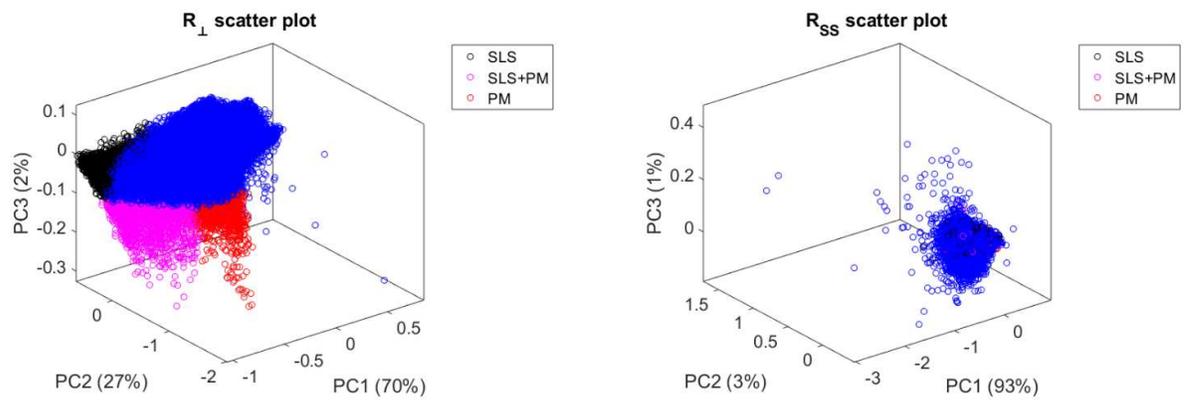


24

25 Fig. 7. Obtained masks from thresholding PCA score images and the plot of mean spectra of
 26 SLS, SLS covered by PM (SLS+PM) and PM infected pixels.

27

28



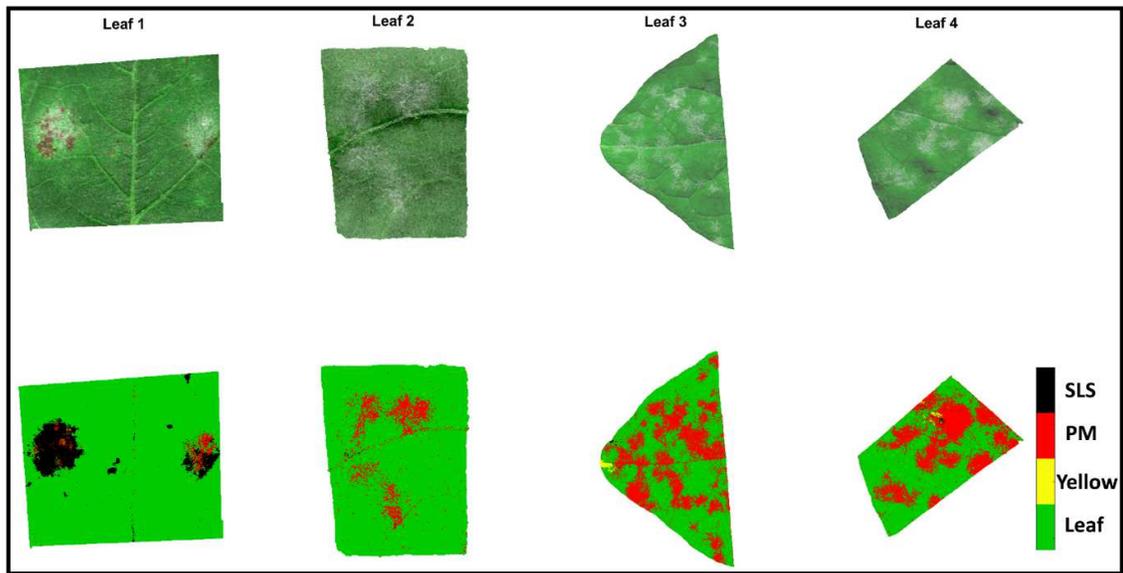
29

30 Fig. 8. PCA scatter plot obtained from R_{\perp} and R_{SS} spectra. Each dot corresponds to a pixel on
31 the upper surface of sunflower leaf. The explained variance by each PC is indicated in the
32 bracket.

33

34

35



36

37 **Fig. 9.** Prediction map to classify SLS, PM, yellow spots and healthy leaf pixels by using
38 cross PLS-DA model.

39

40

41

42 **Appendix A**

43 Light of arbitrary polarization can be represented by four numbers known as the Stokes
44 parameters, I , Q , U and V . I refers to the irradiance or the intensity of the light; the parameters
45 Q , U and V represent the extent of horizontal linear, 45 deg linear and circular polarization,
46 respectively [A1,A2].

47 In polarimetry, the Stokes vector \mathbf{S} of a light beam is constructed based on six flux
48 measurements obtained with different polarization analyzers in front of the detector as
49 follows:

$$50 \quad S = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} I_H + I_V \\ I_H - I_V \\ I_{+45^\circ} - I_{-45^\circ} \\ I_R - I_L \end{pmatrix} \quad (\text{A.1})$$

51 Where $I_H = R_{\parallel}$, $I_V = R_{\perp}$, I_{+45° , I_{-45° , I_R and I_L are the light intensities measured with
52 horizontal linear polarizer, a vertical linear polarizer, a $+45^\circ$ linear polarizer, -45° linear
53 polarizer, a right circular analyzer, and a left circular analyzer in front of detector,
54 respectively.

55

56 In traditional reflectance, we work with an arbitrary polarization and we measured the first
57 element I of the Stokes vector. And in accordance with the theory, this first element is equal
58 to the backscattered light which equal to the sum of horizontal linear intensity and vertical
59 linear intensity.

60

61 **References:**

62 [A1] V.V. Tuchin, L Wang, DA Zimnyakov, “ Optical polarization in biomedical
63 applications”, Springer – Verlag, New York, 2006.

64 [A2] M Born and E Wolf, “Principles of optics: electromagnetic theory of propagation,
65 interference and diffraction of light”, Pergamon Press, Sixth Edition, 1993.

66

67