

1 Appendix A: HSI-PP software tutorial

2 Contents

3	1. Introduction.....	3
4	2. Software main window	4
5	3. Hyperspectral image analysis (HSI analysis)	5
6	3.1 Load image.....	6
7	3.2 Image preprocessing.....	9
8	3.2.1 Radiometric calibration	11
9	3.2.2 Vegetation indices.....	12
10	3.2.3 Image segmentation.....	13
11	3.2.4 Image denoising.....	24
12	3.2.5 Image compression.....	27
13	3.3 Feature extraction.....	34
14	3.3.1 Spectral features	35
15	3.3.2 Image augmentation	35
16	3.3.3 Vegetation indices.....	36
17	3.3.4 Textural features	38
18	3.3.5 Morphological features.....	38
19	3.4 Analysis by model.....	41

20	3.5 Batch processing	42
21	4. Data analysis	47
22	4.1 Data preprocessing	48
23	4.2 Data reduction	50
24	4.2.1 Correlation-based feature selection (CFS)	50
25	4.2.2 Principle component analysis (PCA).....	52
26	4.2.3 Sequential forward selection (SFS)	53
27	4.2.4 Genetic Algorithm (GA).....	53
28	4.2.5 Vegetation index band selection (VI select)	54
29	4.3 Model generation.....	55
30	4.3.1 Random forests.....	56
31	4.3.2 Support vector machines	56
32	4.3.3 Multi-layer perceptron	57
33	4.3.4 k-nearest neighbor	58
34	4.3.5 Linear discriminant analysis.....	59
35	4.3.6 Multiple linear regression.....	60
36	4.3.7 Partial least square regression.....	60
37	4.4 Evaluation of generated models	61
38	5. Import project.....	64
39	6. References.....	65

41 **1. Introduction**

42 The software was designed to provide a user-friendly program for plant researchers and decision-
43 makers to analyze hyperspectral images. It needs to be reliable and flexible for analyzing
44 commonly-used hyperspectral image formats, including band interleaved by line (BIL), band
45 sequential (BSQ), and band interleaved by pixel (BIP), and different extensions. Various features
46 can be extracted from hyperspectral images that can be used in plant phenotyping for research
47 purposes or decision-making. Cost-efficiency, extensibility, usability, and reusability were also
48 considered in the design of this program. HSI-PP software was developed in Python 3.7, which is
49 free and open-source software as a graphical user interface (GUI) using the Tkinter package (Meier,
50 2015), and executed using PyInstaller tools (Cortesi, 2020). The Python script and stand-alone
51 version can be found in (<https://github.com/AI-Elmanawy/Hyperspectral-image-analysis>). Users
52 can install HSI-PP on Windows 10 by running the downloaded file. Once installation is over the
53 software will run automatically and create a shortcut icon on the desktop.

54 The datasets introduced in material and methods (HSI-PP: A flexible open-source software for
55 hyperspectral imaging-based plant phenotyping) will be used during this tutorial to show the main
56 functions of HSI-PP. HSI-PP is composed of 4 working windows: 1) HSI-PP main window, 2)
57 Hyperspectral image analysis, 3) Data analysis, and 4) Import project. There are also some sub-
58 windows that help users to visualize every single step of the analysis and preview the evaluation
59 of different models.

60 **2. Software main window**

61 The main window is shown in (Figure S1). This screen is informative with three different buttons:

62 1) HSI analysis: load hyperspectral images, conduct analysis, extract features and batch

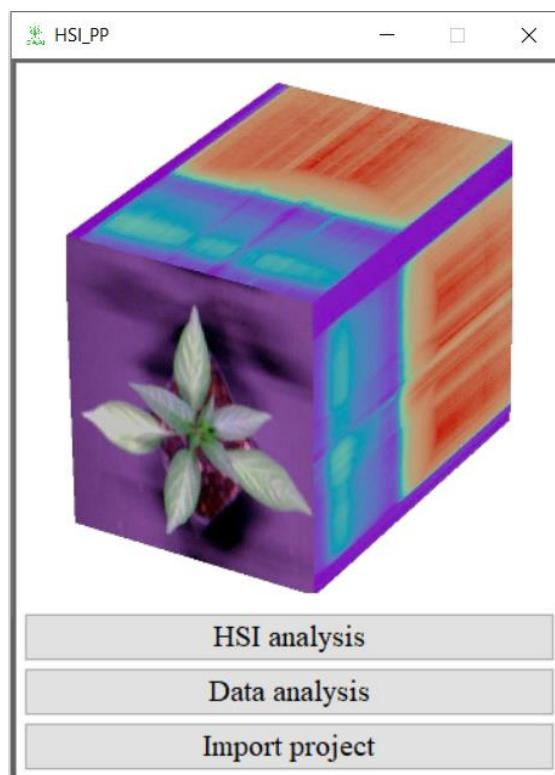
63 processing.

64 2) Data analysis: data preprocessing, data reduction, train different machine learning models

65 (regression and classification), conduct comparison analysis and select the optimal model.

66 3) Import project: run projects previously saved, automatic analysis single or batch of images

67 and analysis image by pretrained model for making decision.



68

69 **Figure S1.** Software's main window.

70 **3. Hyperspectral image analysis (HSI analysis)**

71 In this window, users can import hyperspectral images, analyze these images and extract different
72 features. This window as shown in (Figure S2) contains plot and image display area, list of
73 wavelengths, imported image directory, status bar, and five buttons:

74 1) Load image: specify the directory of captured images and load image. HSI-PP
75 automatically identifies and reads files of the following extensions: .raw, .dat, .img, .tif,

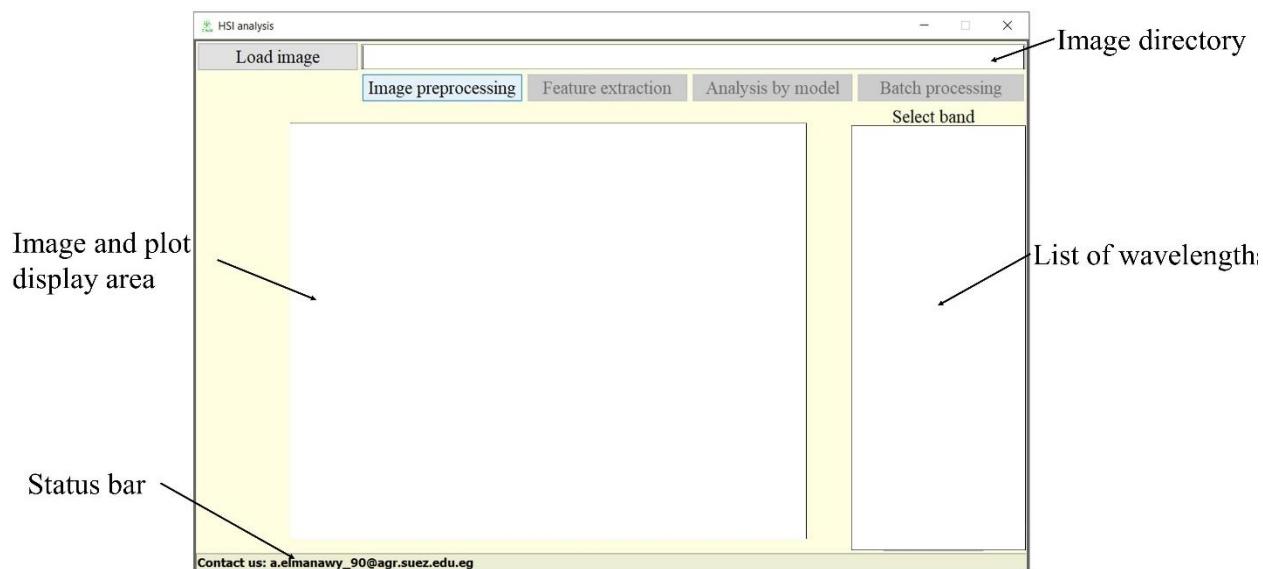
76 and .mat. It is worth mentioning that this software is able to import data with different
77 formats (i.e., BIL, BSQ, BIP),

78 2) Image preprocessing: select different image preprocessing procedures,

79 3) Feature extraction: conduct feature extraction function after image preprocessing,

80 4) Analysis by model: analyzing preprocessed hypercube by pre-trained model,

81 5) Batch processing: preprocessing/analyzing a batch of hyperspectral images in the directory
82 file of the first image. Inactivated button will be activated after image preprocessing.



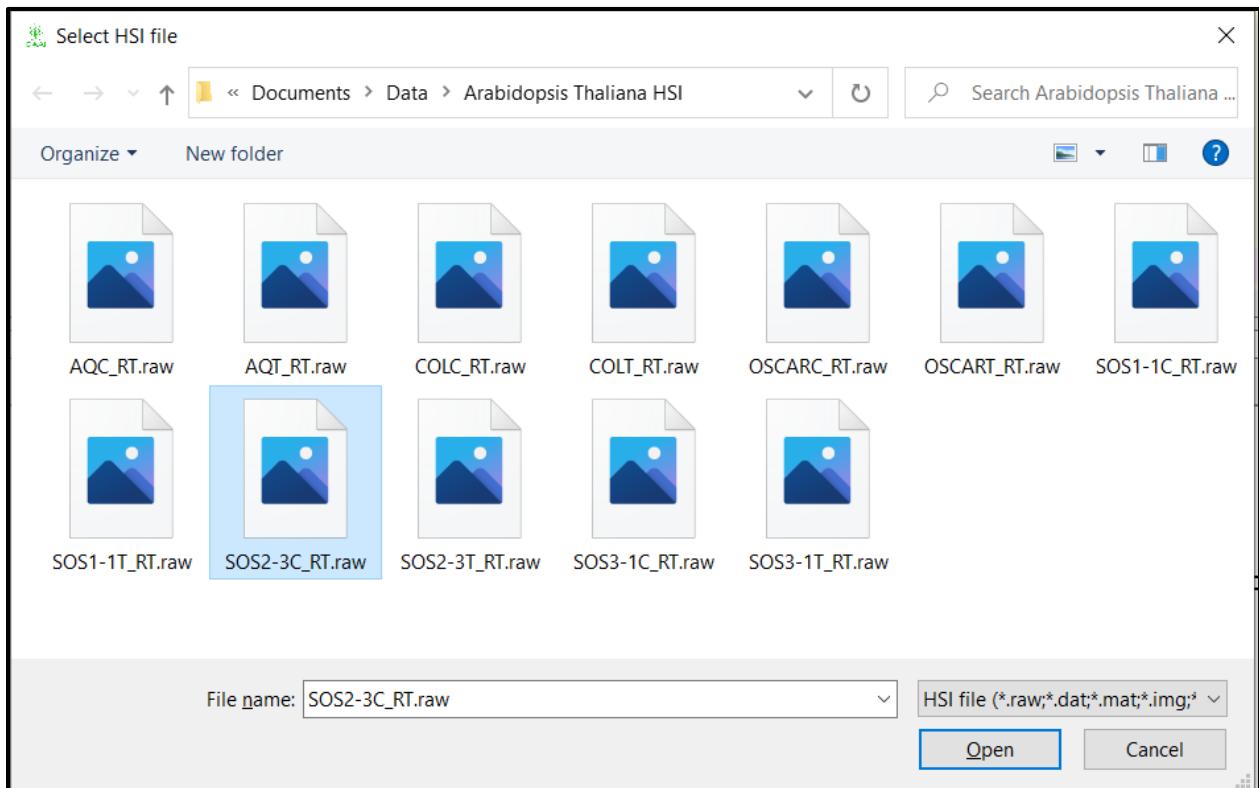
83

84

Figure S2. HSI analysis window.

85 *3.1 Load image*

86 After clicking the “**Load image**” button when the **Select HSI file** dialog appears (Figure S3),
 87 choose the image format of preference. Finish loading the image by double-clicking on it or
 88 clicking on its name and clicking **Open**. When importing images, users need to ensure that every
 89 image has a header file that contains format information of the image. The header file and the
 90 image data file of a hyperspectral image should have the same file name but with a different suffix
 91 (.hdr).

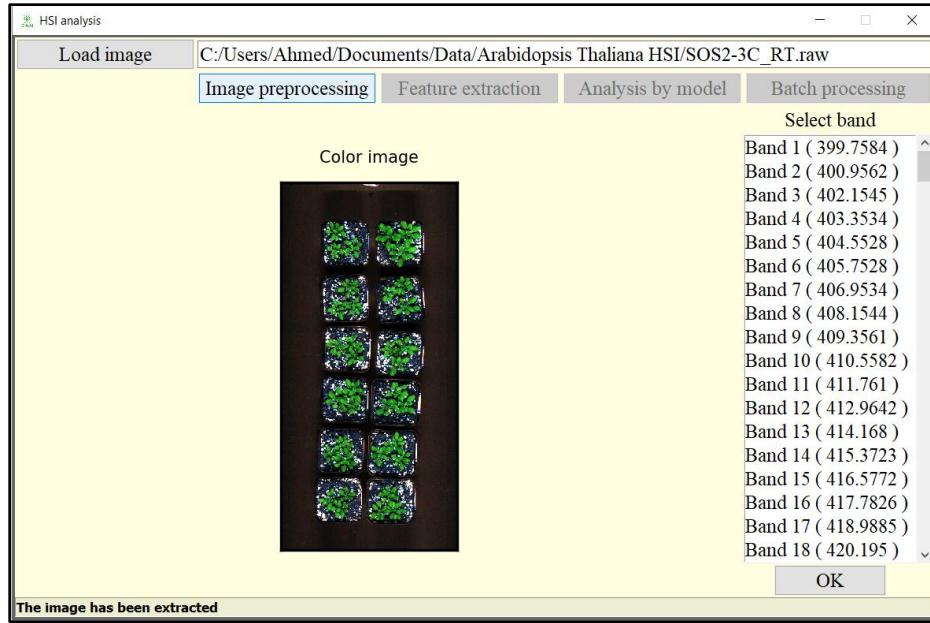


92

93 **Figure S3.** Select HSI file dialog for selecting hyperspectral image file.

94 Once the hyperspectral image is selected, the software will automatically search and read the
95 header file for the format information of the image. Then the software will extract all bands from
96 the image, and display the RGB image (true\false-color), which is automatically saved to the
97 predefined directory. As shown in (Figure S4), the bands' names and values will be listed in the
98 list of wavelengths.

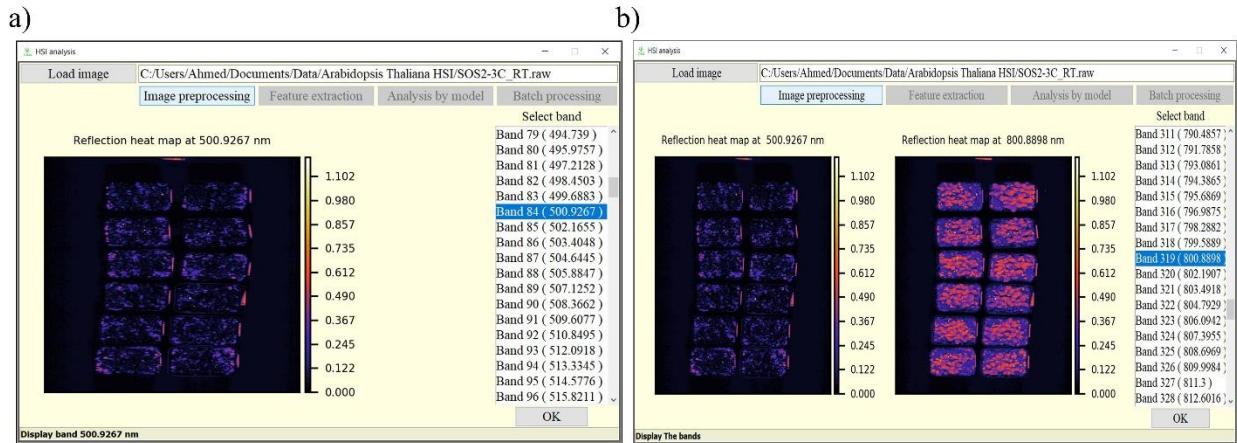
99



100

Figure S4. Open hyperspectral image in HSI-PP.

101 At this point by a double click on any wavelength from the list of wavelengths, the color map
 102 of this band will be displayed in the plot area as shown in (Figure S5a). In addition, it can display
 103 more than one band simultaneously from the list in the file selection dialog by using the **Shift-**
 104 click or **Ctrl**-click before clicking the “**OK**” button as shown in (Figure S5b). This feature will be
 105 helpful for users to visualize different information from different bands at the same time, and useful
 106 in segmentation.



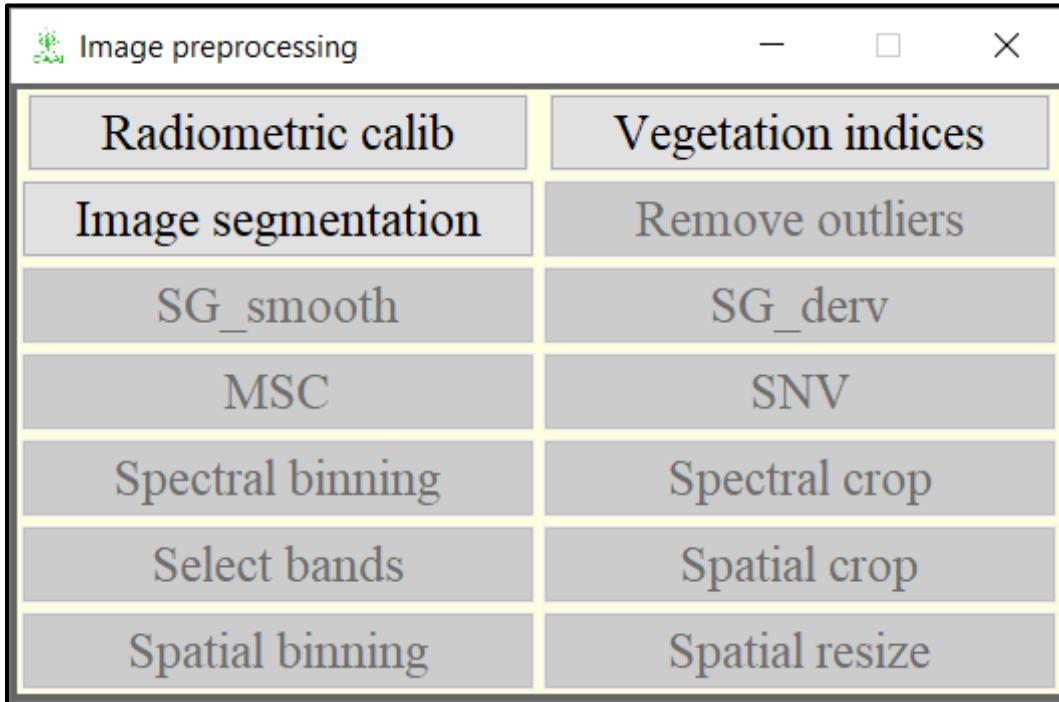
107

108

Figure S5. Display heat map of one band (a) or multiple bands (b).

109 *3.2 Image preprocessing*

110 In HSI-PP, several preprocessing steps were incorporated, including image
 111 denoising, and image resizing and visualization tools to present data before and after preprocessing
 112 to help users in selecting the optimal parameters with the suitable preprocessing procedure. Figure
 113 S6 shows image preprocessing window which contains 14 buttons for image preprocessing
 114 procedures include:



115

116

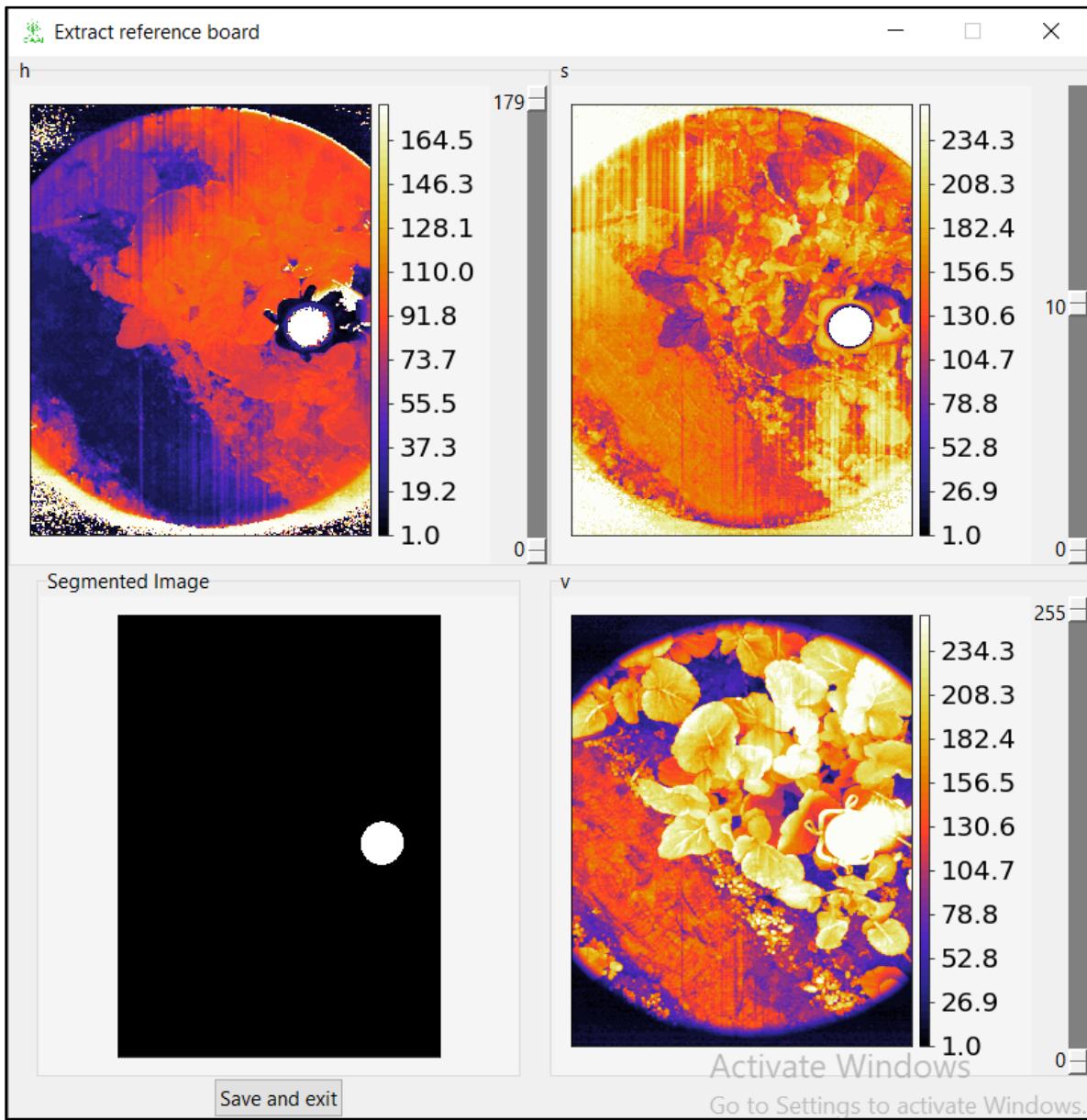
Figure S6. Image preprocessing window.

- 117 1) Radiometric calib: extract reference calibration pan from image and make radiance
118 calibration automatically,
- 119 2) Vegetation indices: calculate different vegetation indices for image segmentation,
- 120 3) Image segmentation: compare between different segmentation methods and select the
121 optimal one,
- 122 4) Remove outliers: remove outlier pixels and increase the signal to noise ratio (SNR),
- 123 5) SG_smooth: smooth image using Savitzky-Golay algorithm,
- 124 6) SG_derv: calculate first and second derivatives using Savitzky-Golay algorithm,
- 125 7) MSC: calculate multiplicative scatter correction (MSC),
- 126 8) SNV: calculate standard normal variate (SNV),

- 127 9) Spectral binning: bin image in spectral dimension,
- 128 10) Spectral crop: remove bad bands,
- 129 11) Select bands: select specific band from an image at a specific wavelength,
- 130 12) Spatial crop: split image into sub-images,
- 131 13) Spatial binning: bin image in spatial dimensions,
- 132 14) Spatial resize: resize an image at specific spatial dimension.

133 3.2.1 *Radiometric calibration*

134 HSI-PP can calibrate the hyperspectral images, but users should ensure that the reference panel is
135 acquired with the target (plant in our case) in the same image. Click on the “**Radiometric calib**”
136 button, select threshold values to segment reference-board from the image using scroll bars as
137 shown in Figure S7, once it is done click the “**Save and exit**” button for image calibration.



138

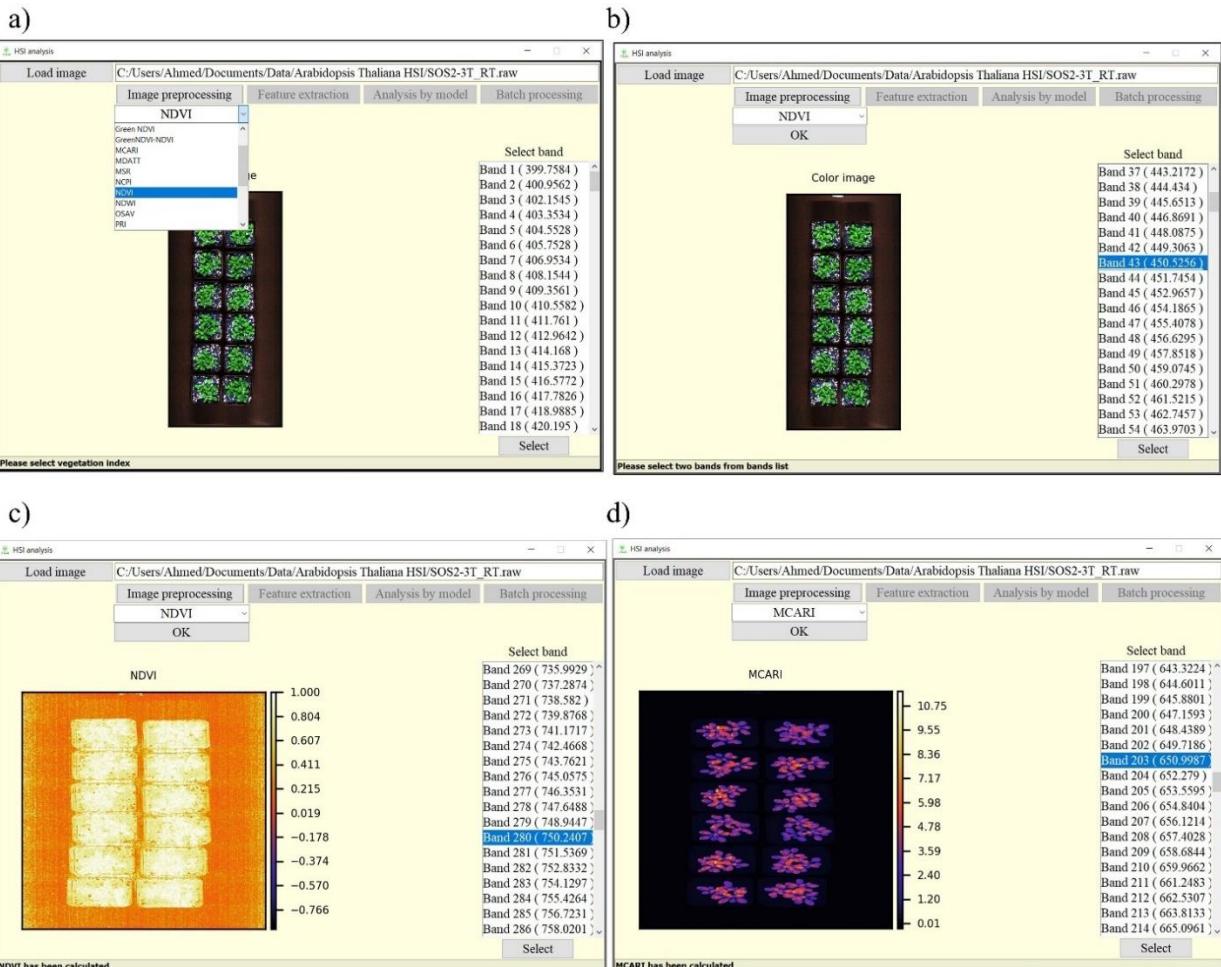
139

Figure S7. Segment reference-board from image.

140 3.2.2 *Vegetation indices*

141 Click on the “**Vegetation indices**” button, users can select index (Table S2 in notes S1) from the
 142 drop-down box (Figure S7a) and by **Ctrl**-click, users can select related bands from the list of
 143 wavelengths (Figure S8b). Then the software will calculate the selected index after clicking the

144 “Select” button. This step is very important in comparing between different indices and select the
 145 optimal one for segmentation Figure (S8 c,d).



146

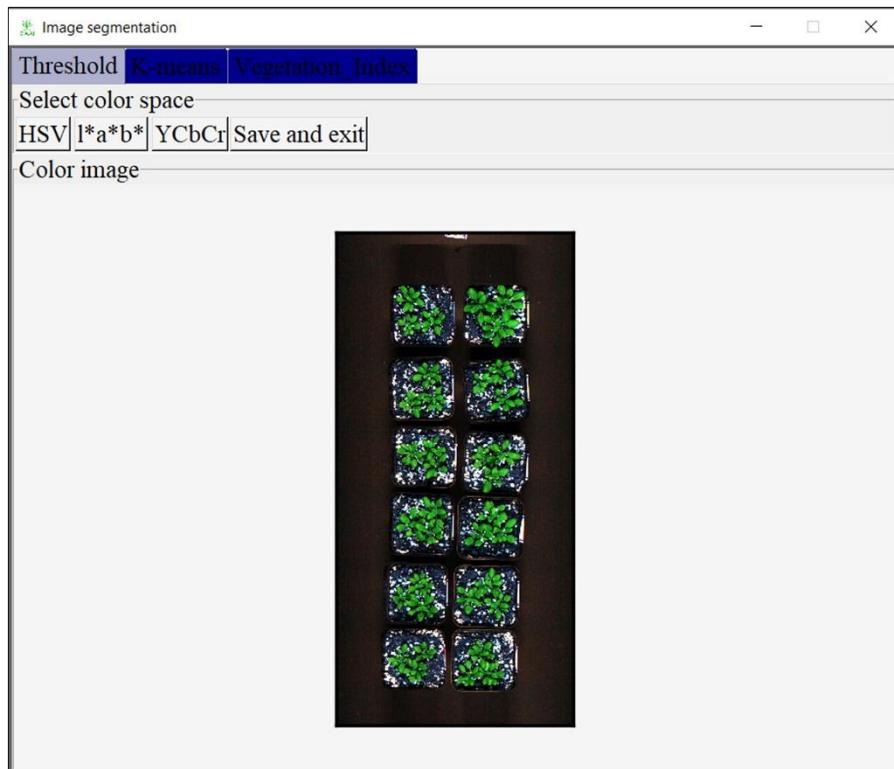
147

148 **Figure S8.** Calculation of vegetation index, select index type (a), select bands from list (b), calculate normal
 149 vegetation index (NDVI) at band 450 and 750nm (c), and calculate Modified Chlorophyll Absorption in Reflectance
 150 Index at bands 450, 650 and 750nm.

151 3.2.3 Image segmentation

152 The software contains three segmentation methods namely threshold-based, learning-based (K-
 153 means) and vegetation index-based segmentation method. Once users click on the “Image

154 **segmentation”** button, the image segmentation window appears which contains three tabs
155 (**Threshold** tab, **K-means** tab, and **vegetation_index** tab) as shown in (Figure S9). In threshold,
156 tab users have to select amongst three-color spaces (**HSV**, **CIELab**, or **YCbCr**), which shows the
157 highest variability between foreground and background (Figure S10 a-c). Then apply various
158 threshold values using the scroll bar (in red box) and the result of segmentation will be shown in
159 the same window (Figure S10d right) if the segmented result accepted click on the “**Done**” button,
160 then click “**Save and exit**” button in segment window (Figure S9).

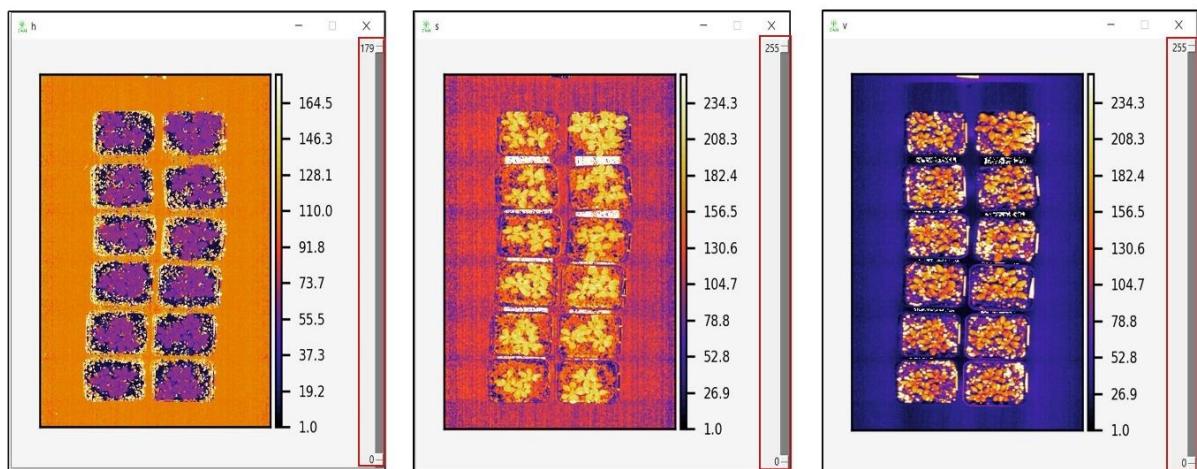


161

162

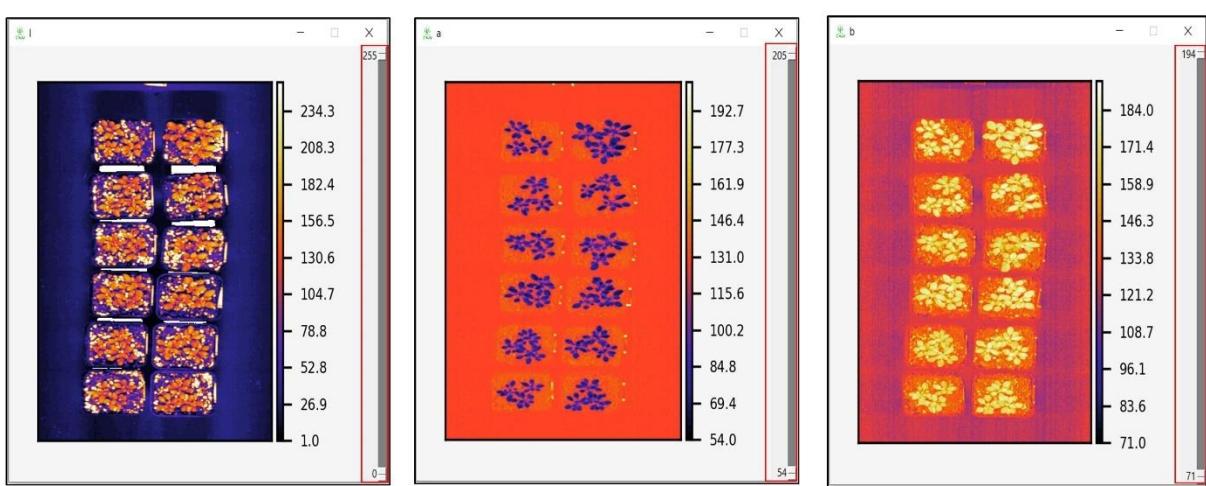
Figure S9. Image segmentation window.

a)



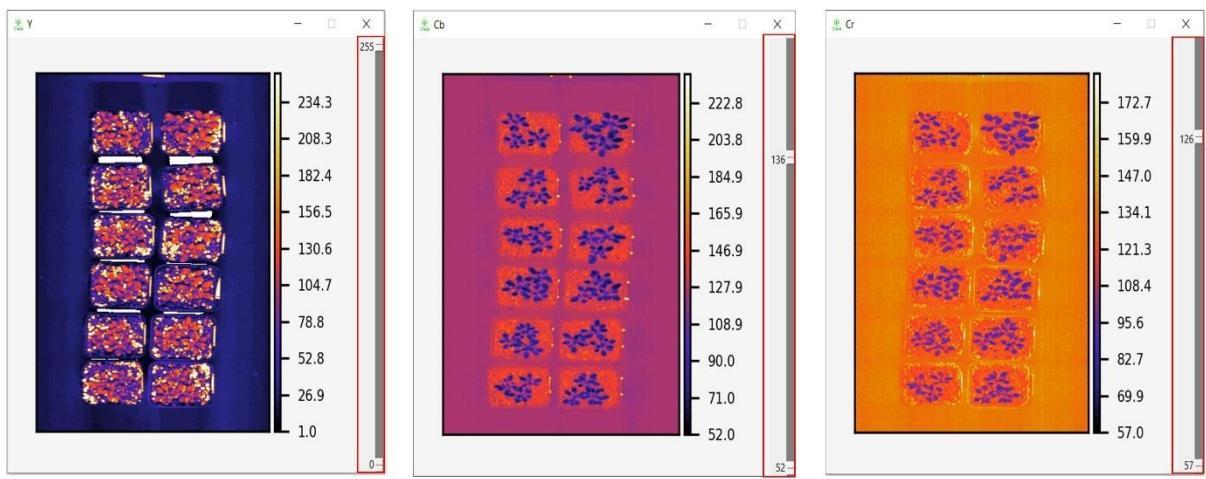
163

b)

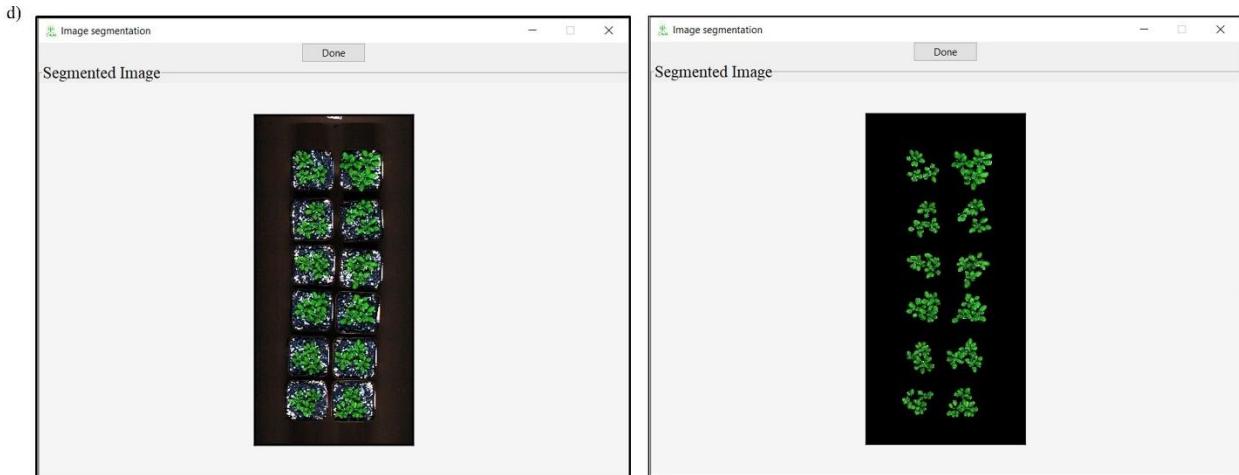


164

c)



165

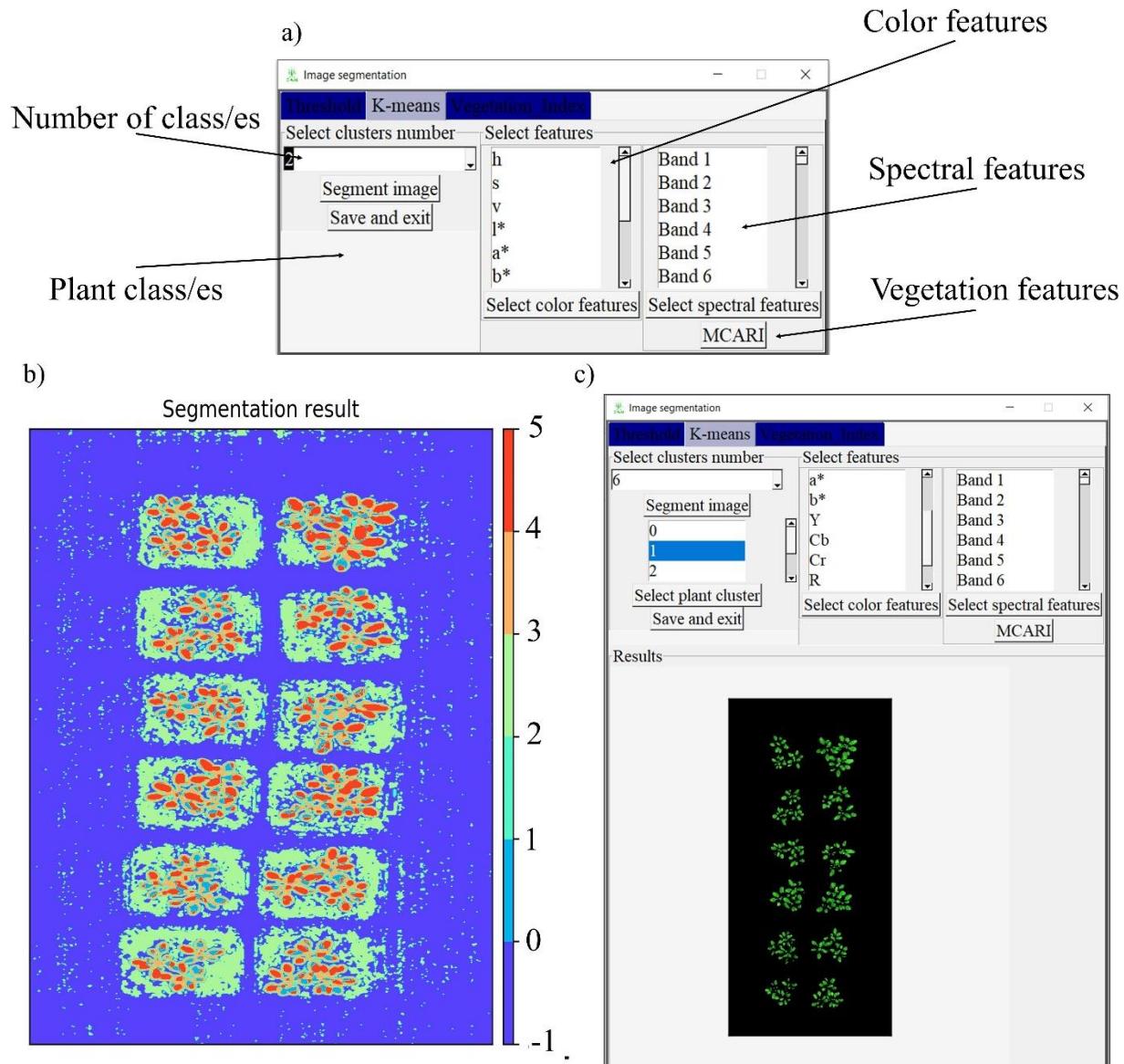


166

167 **Figure S10.** Compare between three color space HSV (a), CIELab (b) and YCbCr (c) and segmentation result (d).

168 In the K-means tab users can select one feature or more from two lists of features (color
 169 features (i.e., H, Y, Cb, Cr, R, G and B), or spectral features), then click the “**Select color features**”
 170 or the “**Select spectral features**” button, respectively. There are also more option for using
 171 vegetation feature by click a button with vegetation index name (Figure S11a). Then select the
 172 number of the clusters (ranged from 2 to 20) default is (2). By clicking the “**Segment image**”
 173 button the K-means algorithm will divide an image into a selected number of classes as shown in
 174 (Figure S11b). As it is shown in (Figure S11b) there is more than one class representing the plant
 175 shoot area so the software combines these classes (class 4 and 5 after selecting it from the list) to
 176 represent foreground (Figure S11c) by clicking the “**Select plant cluster**” button.

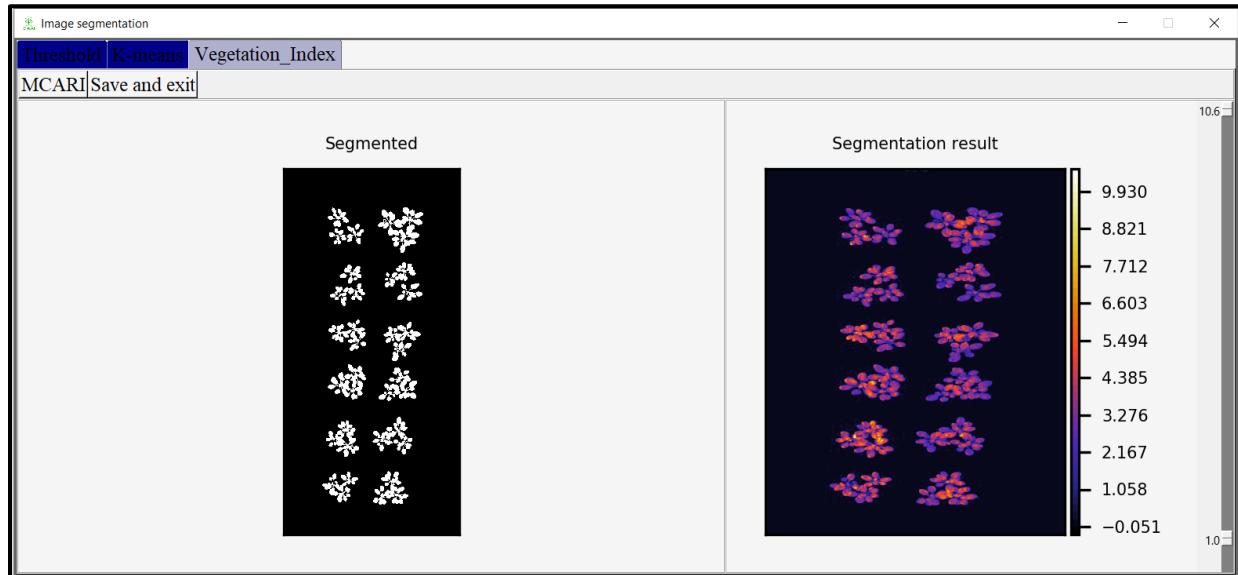
177



179
180 **Figure S11.** K-means for image segmentation. a) K-means tab component, b) K-means cluster result, and c)

181 select classes which represent foreground.

182 As it is shown in Figure S7c, MCARI is better than NDVI for segmentation. For threshold
183 using vegetation index after choosing the optimal VI for segmentation, click the third tab then click
184 “MCARI” button (the button name according the selected VI) select the threshold value using
185 scroll bar as shown in (Figure S12).



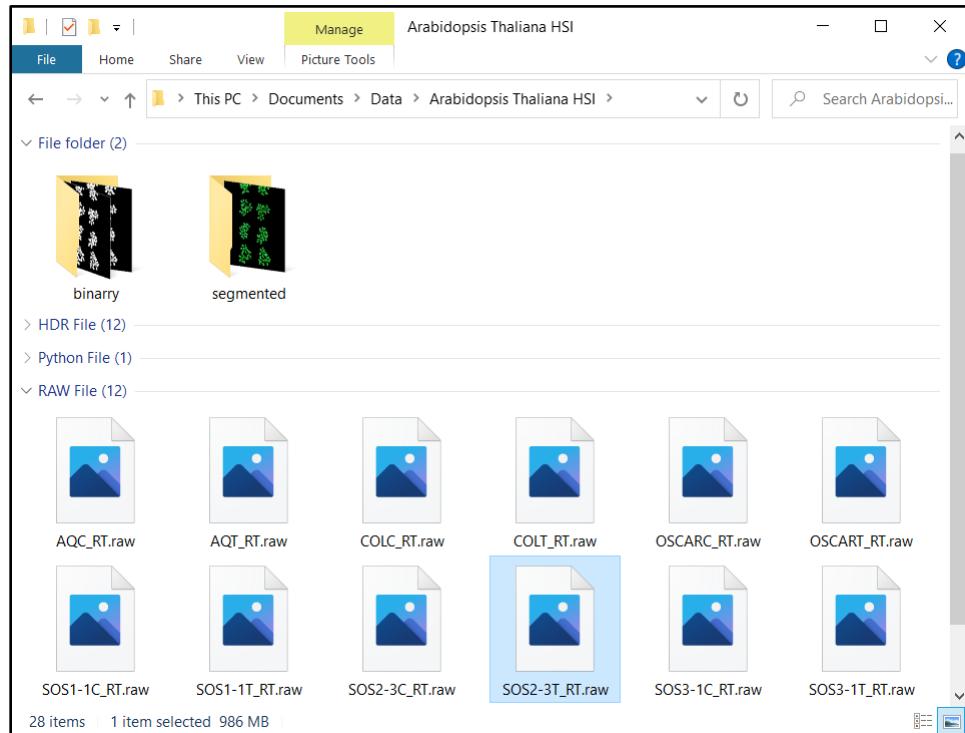
186

187

Figure S12. Vegetation tab in image segmentation window.

188 Once the segmentation is done, the connected components method is used on the processed
 189 image to remove spurious outliers and noise from the image (e.g., plant debris on soil). By clicking
 190 the “**Save and exit**” button the masked image will be automatically saved in the “**binary**” file and
 191 the segmented RGB in the “**segmented**” file in the predefined directory (these files are created
 192 automatically) (Figure S13), then the masked image will be applied to HSI as region of interest
 193 (ROI). Once the hyperspectral image is segmented, HSI-PP automatically detects the smallest
 194 rectangle containing the ROI and eliminates the area outside this rectangle that will reduce the
 195 image’s spatial dimension. After masking, the transition between foreground and background is
 196 very sharp. Pixels at this transition include parts of both classes and are depicted as “mixed pixels”.
 197 In this step, the “**image filter**” sub-window will be open, select a specific band from the list by a
 198 double click on the selected band. Then using the scroll bar to choose a minimum and/or maximum

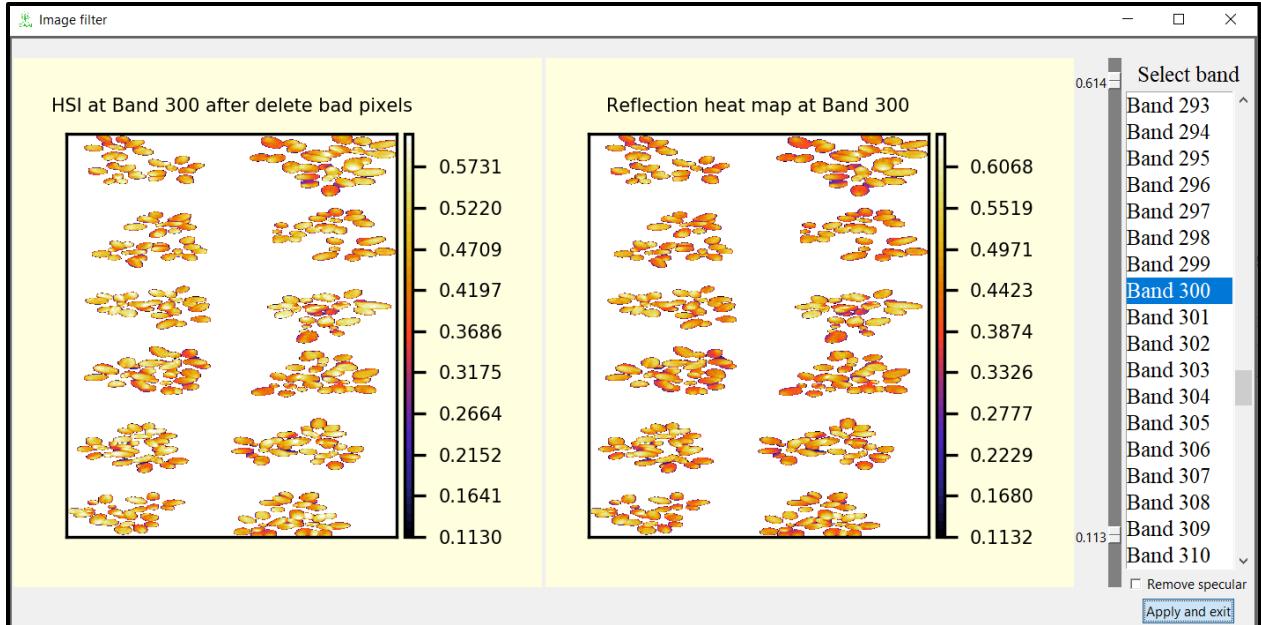
199 value till these mixed pixels in the image vanished. This step could be used to build a mask from
200 any band; by passing the aforementioned steps and go to this step (Figure S14). The specular pixels
201 will be removed automatically if the user check “**Remove specular**” check box. If the threshold
202 value is suitable click “**Apply and exit**” button.



203

204

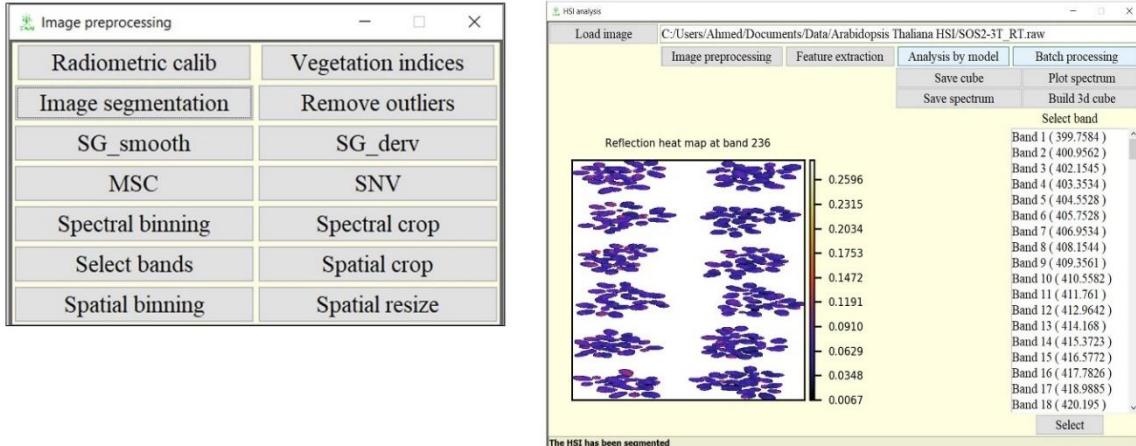
Figure S13. Save binary and segmented RGB in the predefined directory.



205

206 **Figure S14.** Image filter window for band threshold (removing bad pixels) the selected band before removing
207 bad pixels (right) after removing bad pixels (left).

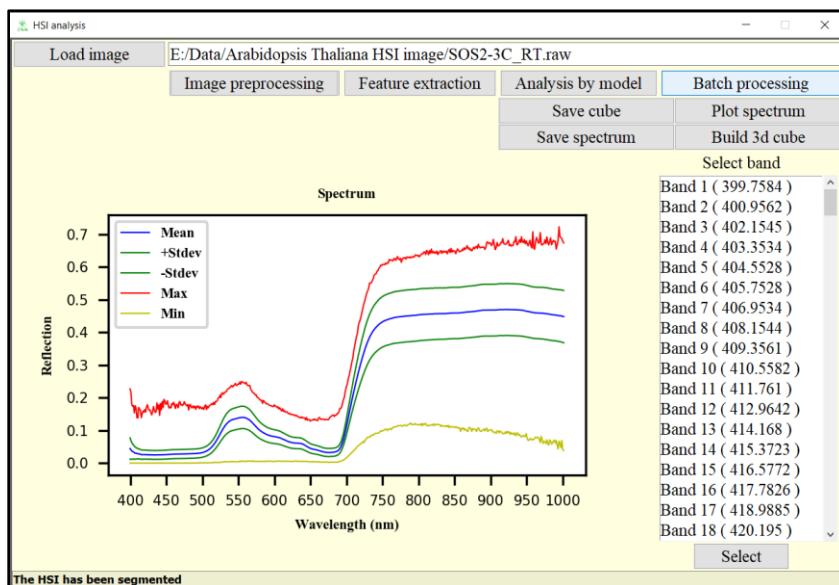
208 After image filtration, all buttons in the image preprocessing window and HSI analysis window
209 will be activated (Figure S15). Also there are four more button appear: “**Save cube**” for saving
210 hypercube as .mat file; “**Plot spectrum**” button for plotting mean spectrum; “**Save spectrum**”
211 button for save mean spectrum as excel file; and “**Build 3d cube**” button for plot hypercube as
212 three dimension plot.



213

214 Figure S15. Image preprocessing window(left) and HSI analysis window(right) after image segmentation.

215 In this step, users have four options. The first one plot spectrum to check if the denoising
216 procedure is necessary or not by click “**Plot spectrum**” button. The mean spectrum, maximum
217 values, minimum values and positive and negative stander will be plotting (Figure S16).



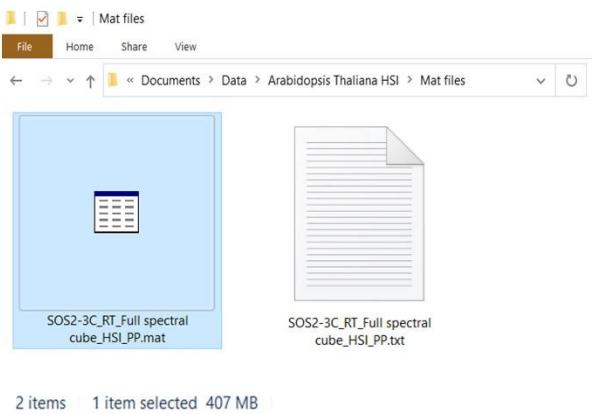
218

219 Figure S16. HSI-PP main window after image segmentation and click plot spectrum button.

220 The first one is saving spectral and spatial features for the processed cube as three dimensions
221 cube by click “**Save cube**” button, the softwrae will save the new hypercube as a mat file with the

222 same image name and predefined directory (Figure S17a). The software will also save the metadata
223 of this hypercube (i.e., spatial and spectral dimension and wavelength values) as txt file with the
224 same image name (Figure S17b). The software reduces hypercube size from 986 MB to 407 MB
225 (about 60% reduction) just after segmentation.

a)



b)

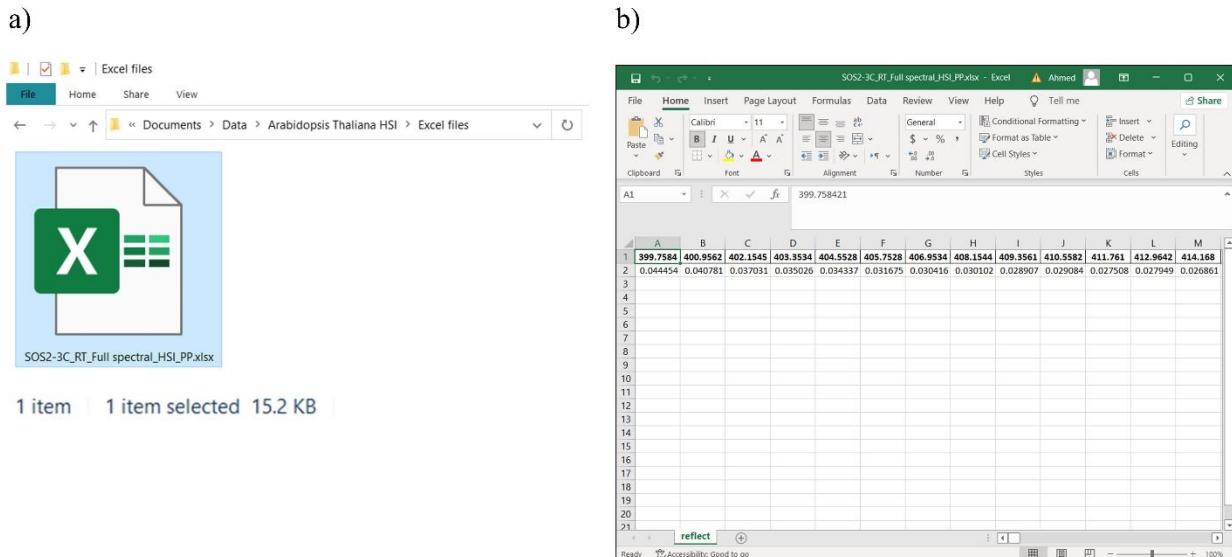
A screenshot of a Notepad window titled 'SOS2-3C_RT_Full spectral cube_HSI_PP.txt - Notepad'. The window contains the following text:

```
File Edit Format View Help
interleave = bil
samples = 292
lines = 864
bands = 473
default bands={193,122,50}
wavelength = {399.7584,
400.9562,
402.1545,
403.3534,
404.5528,
405.7528,
406.9534,
408.1544,
409.3561,
410.5582,
411.761,
412.9642,
414.168,
415.3723,
416.5772,
417.7826,
418.9885,
420.195,
421.402,
422.6095,
423.8176,
425.0262,
426.2252
<
```

The window includes standard Notepad controls at the bottom: 'Ln 7, Col 48', '100%', 'Windows (CRLF)', and 'UTF-8'.

226
227 **Figure S17.** Save hypercube as mat file a) create and save mat file with the same image name, b) metadata for saved
228 hypercube as txt file.

229 The second option calculating the mean spectrum for all pixels in ROI and saves it as an excel
230 file with the same image name and predefined directory by click “**Save spectrum**” button (Figure
231 S18a). The software will save the value of wavelength as column names and the first raw will be
232 the mean spectrum (Figure S18b).

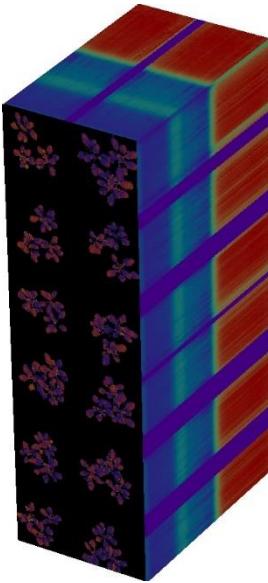


233

234 **Figure S18.** Save mean spectrum as excel file a) create and save excel file with the same image name, b)

235 wavelength values and mean spectrum in excel file.

236 The third option plotting and saving hypercube as 3D cube plot by clicking the “**Build 3d**
 237 **cube**” button (Figure S19). The last option, if the spectrum is not good enough and require
 238 preprocessing the users can go to next step for image denoising.



239

240

Figure S19. plot new hypercube as 3d cube.

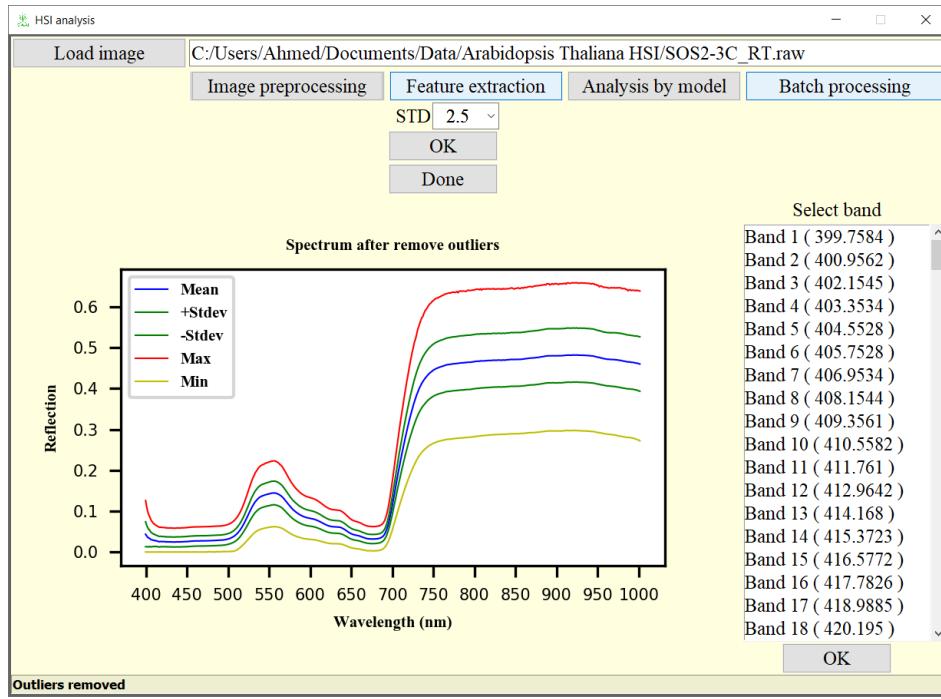
241 *3.2.4 Image denoising*

242 The denoising procedure is essential for reducing the noise in the hyperspectral image before
243 feature extraction. In Figure S16 the pattern of min and max reflection is not right which means
244 that the new hypercube contains contamination and it requires data denoising. There are a number
245 of alternatives for preprocessing routines of spectral data such as smoothing, first and second
246 derivatives, multiplicative scatter correction, and standard normal variate which implemented in
247 this software. The user can compare between these methods to select the optimal one according to
248 spectrum pattern and the preprocessing method should produce a robust model with the best
249 predictive ability.

250 **3.2.4.1 Remove outlier pixels**

251 Click the “**Remove outliers**” button, select the standard-deviation threshold’s value, then the

252 result will be shown as (Figure S20) after clicking the “**ok**” button. If the plot pattern is right, click
253 the “**Done**” button. If it is not, reselect the standard-deviation threshold’s value. The user has
254 option to save hypercube as mat file and save mean spectrum as excel file (Figure S17, S18).

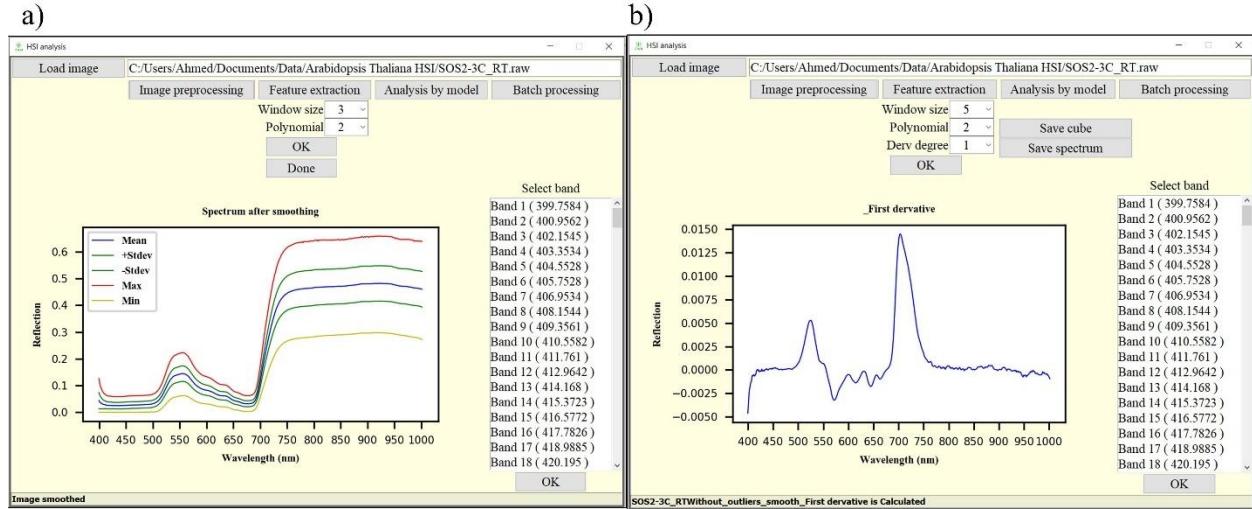


255
256 **Figure S20.** Spectrum pattern after remove outliers’ pixels by select stander-deviation threshold’s value (2.5).

257 3.2.4.2 Image smoothing and differentiation

258 In this software, Savitzky-Golay (SG) was used for smoothing images and calculating the first
259 and second derivative (Savitzky and Golay, 1964, Rinnan *et al.*, 2009). Click the “**SG_smooth**”
260 button, then select the window size and polynomial degree to smooth (Figure S21a). The SG will
261 be applied to the spectrum at each pixel position independently to reduce the random noise. If the
262 plot pattern is right, click the “**Done**” button, if it is not, changes window size or polynomial
263 degree. The user has option to save hypercube as mat file and save mean spectrum as excel file

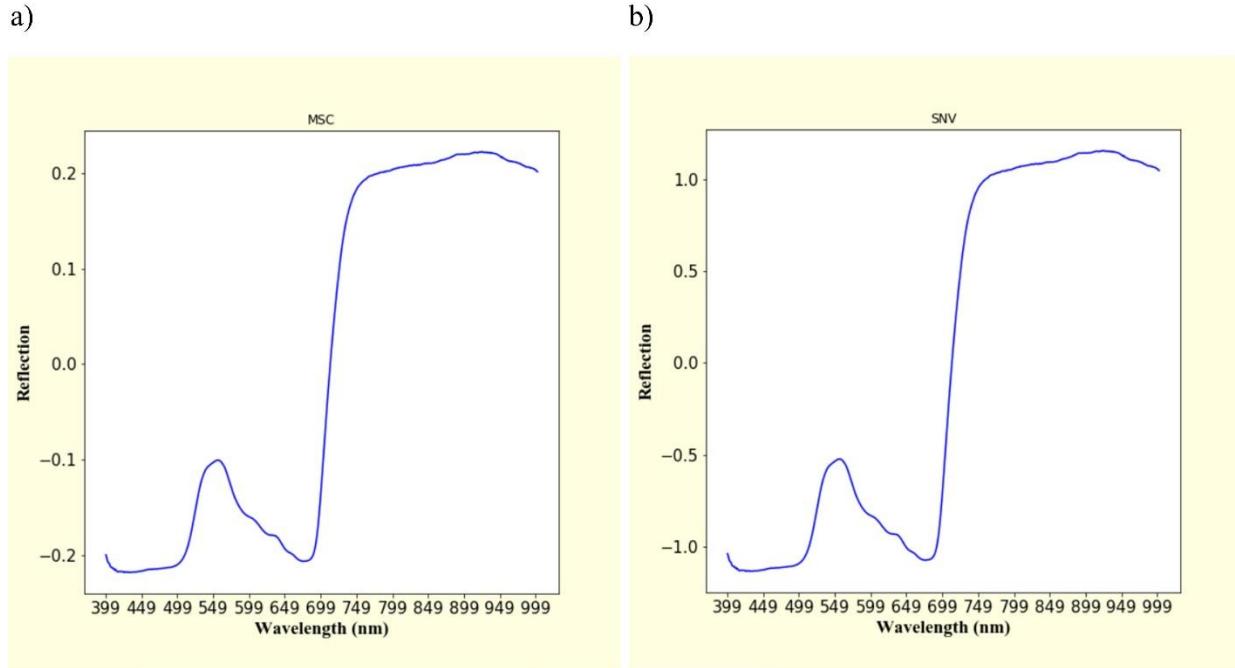
264 (Figure S17, S18). Click the “SG_derv” button and select window size, polynomial degree and
 265 the derivative degree to calculate derivatives (Figure S21b). The user has option to save derivative
 266 hypercube as mat file and save mean derivative spectrum as excel file (Figure S17, S18).



267
 268 **Figure S21.** Image denoising a) image smoothing, and b) image derivative (first derivative).

269 3.2.4.3 Scatter correction

270 Click the “MSC” button or the “SNV” button for multiplicative scale correction (Figure S22). All
 271 of these preprocessing procedures are available in HSI-PP.py in functions (11-13). The user has
 272 option to save MSC/SNV hypercube as mat file and save mean MSC/SNV spectrum as excel file
 273 (Figure S17, S18). It is important to mention that, the spectrum pattern before and after processing
 274 procedures aforementioned will be visualize to check whether the preprocessing procedure is
 275 correctly selected or not, since a wrong selection of the preprocessing method may distort the
 276 useful information in the spectra.



277

278 **Figure S22.** Image multiplicative scale correction by calculating a) multiplicative scatter correction (MSC) and

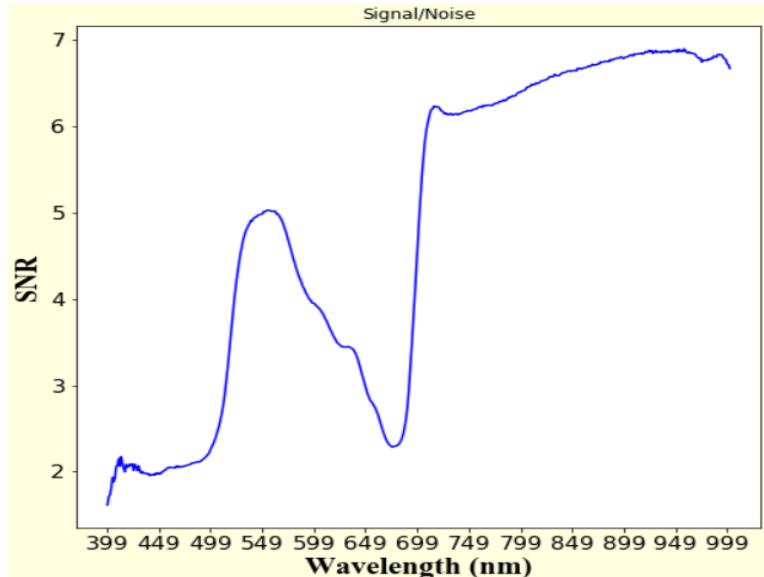
279 b) standard normal variate (SNV).

280 *3.2.5 Image compression*

281 An image compression feature related to the spatial and spectral crops, spatial and spectral binning,
 282 spatial resize, and band selection was also added in HSI-PP, which makes sharing and analyzing
 283 hyperspectral images easier.

284 **3.2.5.1 Image crop**

285 For spectral cropping, HSI-PP detects signal-to-noise ratio (SNR) for every band, then visualizing
 286 it (Figure S23).

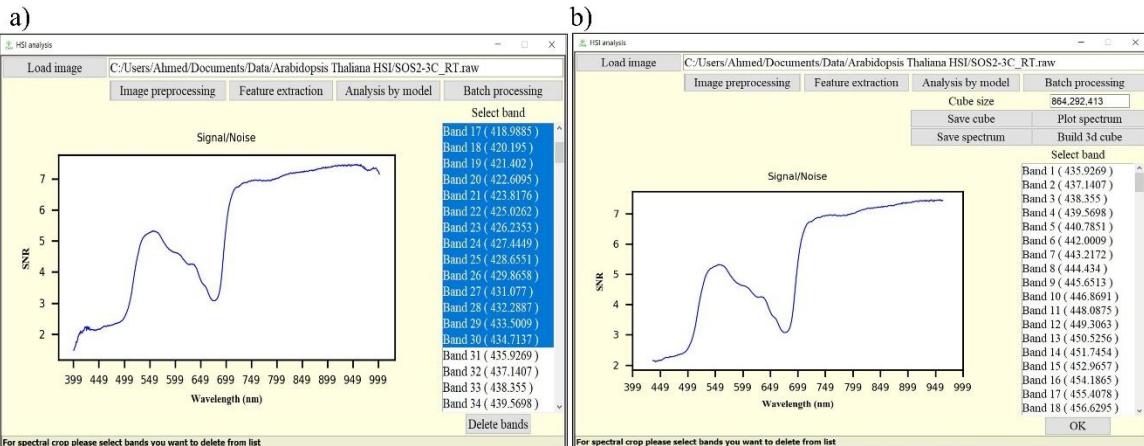


287

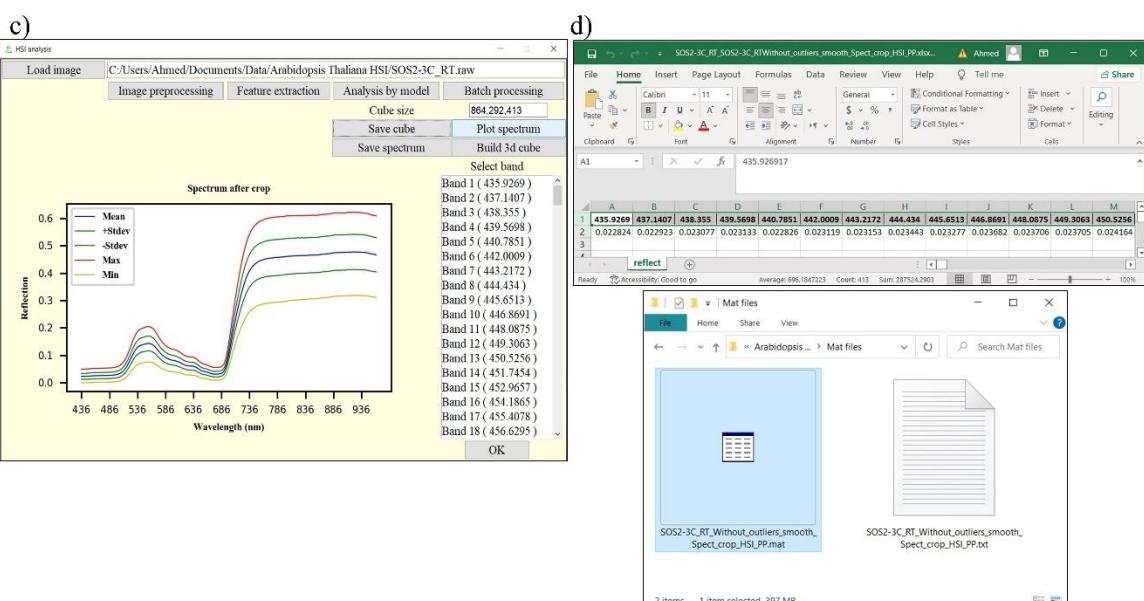
288

Figure S23. Signal noise ratio for all wavelengths.

289 In this step, users can compare and eliminating bands with low SNR by selecting bands (to be
 290 deleted) from the list by **Ctrl-click**, then clicking the “**Delete bands**” button (in this hypercube
 291 the first and last 30 bands have to be deleted) new hypercube size will be displayed (Figure S24a).
 292 If it is ok click “**Done**” button the hypercube will be ready for next step or saved as mat file (Figure
 293 S24b), if it not users can reselect the undesirable bands by re-clicking the “**Spectral crop**” button.
 294 By click “**Plot spectrum**” button the mean spectrum for new hypercube will be displayed (Figure
 295 S24c). The main spectrum for new hypercube can be saved as an excel file, also hypercube can be
 296 saved a mat file by click “**Save spectrum**” and “**Save cube**” buttons, respectively (Figure S24d).
 297 The size of hypercube reduced to 397 MB (about 60% reduction) after this process. This process
 298 has to be done before spatial crop not after.



299



300

301 **Figure S24.** Spectral crop by HSI-PP a) select the bands with lower SNR, then b) image size and wavelength list

302 after spectral crop, c) plot spectrum after spatial crop and d) save mean spectrum as excel file and hypercube as mat

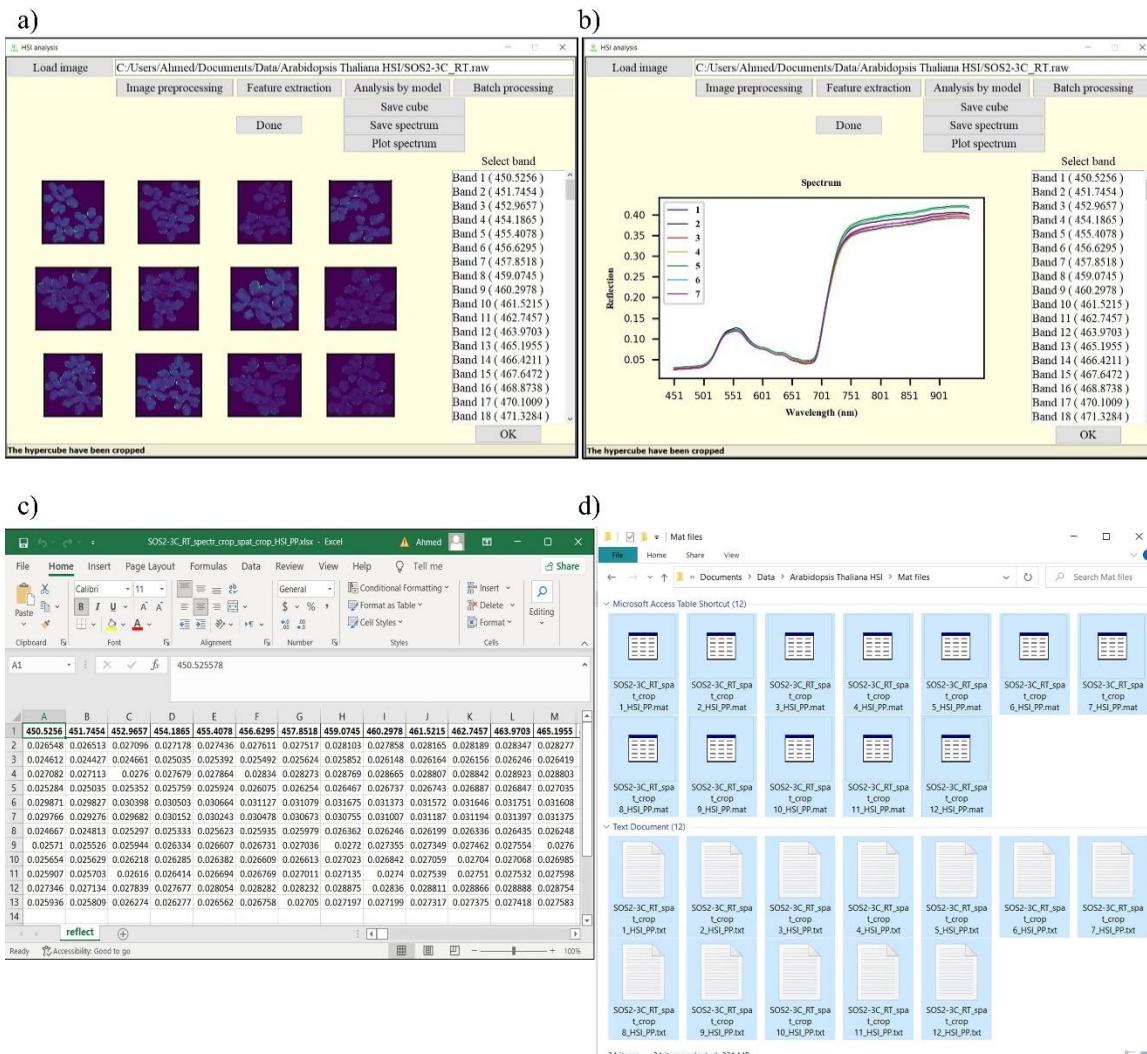
303 file.

304 For spatial crop, click the “**Spatial crop**” button in the image preprocessing window, the image

305 will automatically split into sub-images sorted from top-right to bottom-left (Figure S25a). Click

306 “Done” button then click “**Plot spectrum**” button to display the mean spectrum for every sub-

307 image (first seven sub-image) will be displayed (Figure S25b). The main spectrum for sub-images
 308 can be saved as an excel file and every sub-image can be saved as a separated mat file by click
 309 “Save spectrum” and “Save cube” buttons, respectively (Figure S25c, d). As it is shown in (Figure
 310 S25d) every sub-image has its txt file and the size of hypercube reduced to 221 MB (about 75%
 311 reduction) after this process.

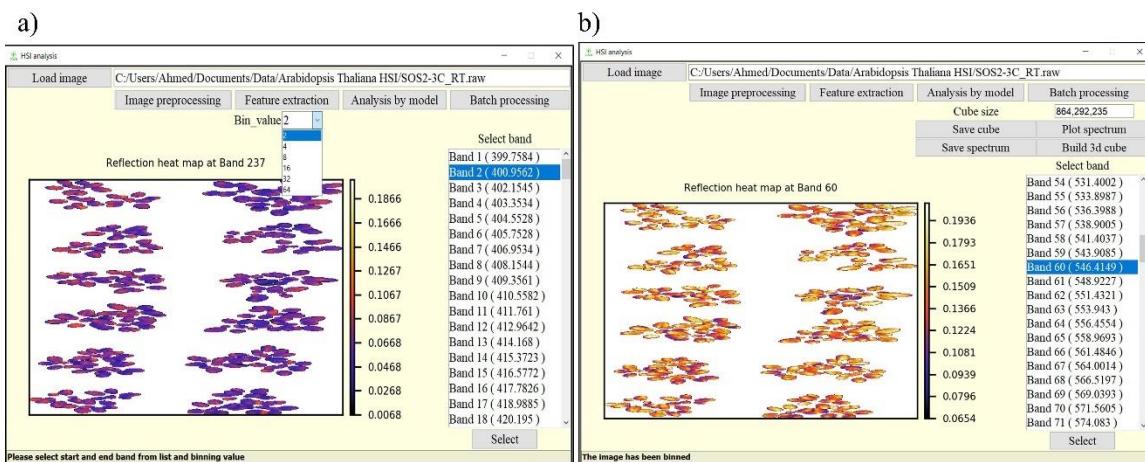


313
 314 **Figure S25.** Split image into sub-images a) display sub-image cube, b) plot spectrum for sub-
 315 image, c) save spectrum in excel file, d) save every sub-image as separated mat file (12 file).

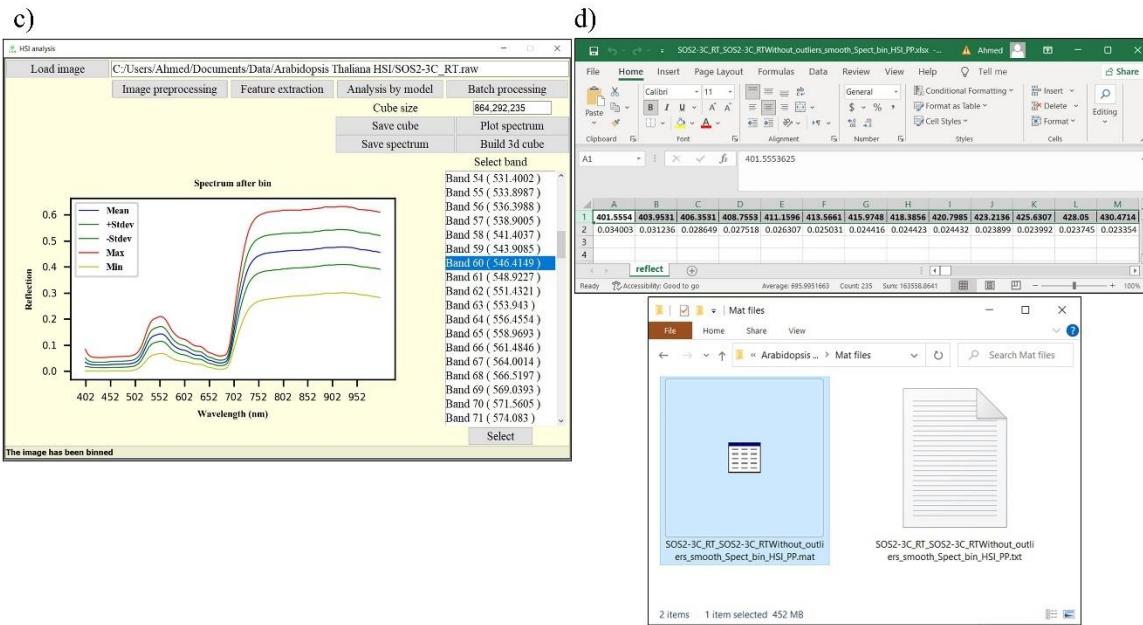
316

3.2.5.2 Image binning

317 For spectral binning, click the “**Spectral binning**” button then from wavelength list choice start
 318 and end band using **Ctrl-click**. After that, select a bin value from the drop-down combo (Figure
 319 S26a) (which is the power of 2) then clicks the “**Select**” button. The new hypercube size after
 320 binning will be display and also the value of wavelength after binning will be listed (Figure S26b).
 321 If it is ok click “**Done**” button the hypercube will be ready for next step or saved as mat file (Figure
 322 S26b), if it not users can repeat these steps by re-clicking the “**Spectral binning**” button. By click
 323 “**Plot spectrum**” button the mean spectrum for new hypercube will be displayed (Figure S26c).
 324 The main spectrum for new hypercube can be saved as an excel file, also hypercube can be saved
 325 a mat file by click “**Save spectrum**” and “**Save cube**” buttons, respectively (Figure S26d). The
 326 size of hypercube reduced to 452 MB (about 50% reduction) after this process. This process has
 327 to be done before spatial crop not after.



328



329

330 **Figure S26.** Image spectral binning a) select binning value and start and end band, b) image size and wavelength

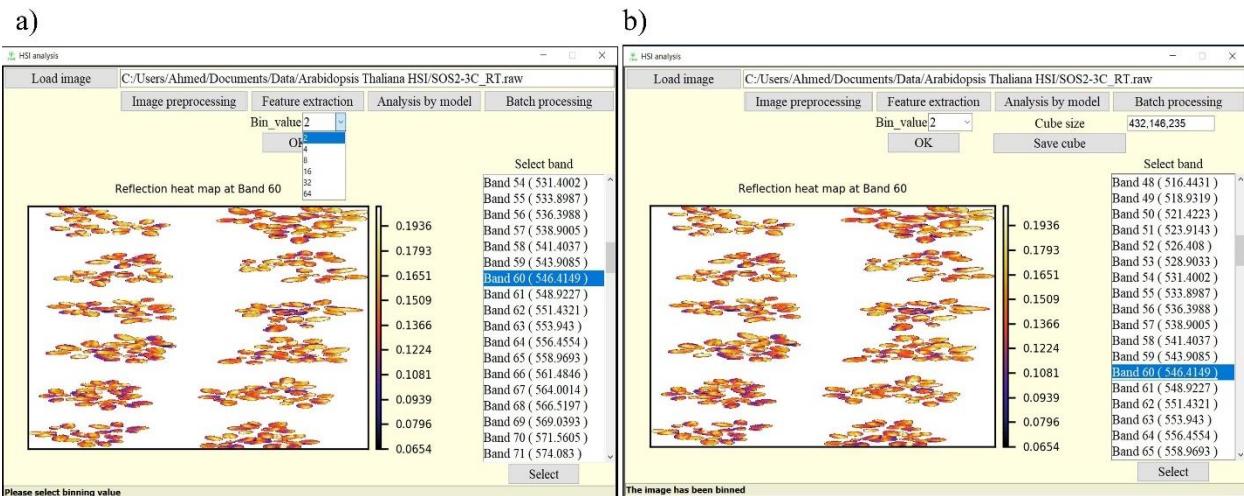
331

list after spectral binning.

332 For spatial binning, click the “**Spatial binning**” button, then only select bin value from the

333 drop-down combo then clicks the “**OK**” button to bin contiguous pixels (Figure S27). The size of

334 hypercube reduced to 150 MB (about 85% reduction) after this process.



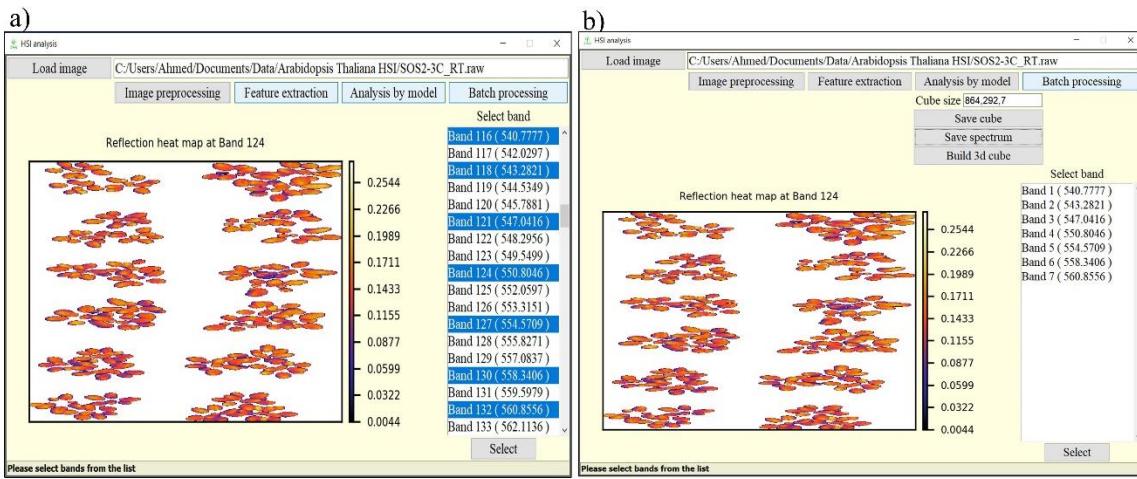
335

336 **Figure S27.** Image spatial binning a) select binning, b) image size after spectral binning.

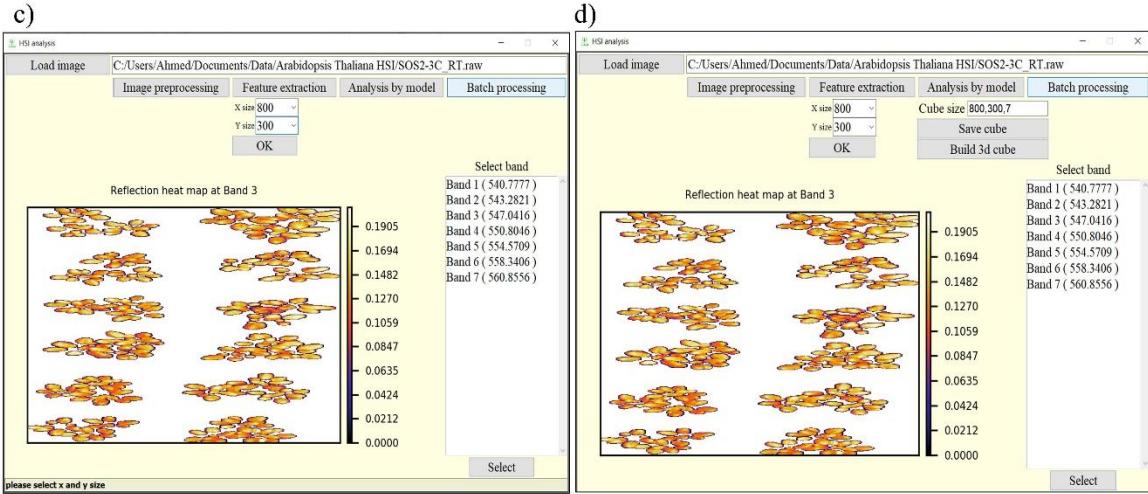
337

3.2.5.3 Image resizes

338 For resizing the HSI cube in spectral dimension, click the “**Select bands**” button, then select
 339 required bands from the list by **Ctrl-click** as shown in (Figure S28a). If it is ok click “**Done**” button
 340 the hypercube will be ready for next step or saved as mat file (Figure S28b), if it not, users can
 341 repeat these steps by re-clicking the “**Select bands**” button and reselect bands again. For spatial
 342 resize, click the “**Spatial resize**” button then select/insert the value of the new spatial dimension
 343 (Figure S28c), then click “OK” button the new hypercube dimensions will be display and ready
 344 for save as mat file (Figure S28d). This procedure (spectral/spatial resize) is necessary for
 345 preparing the hypercube to analysis by deep-learning.



346



347

348

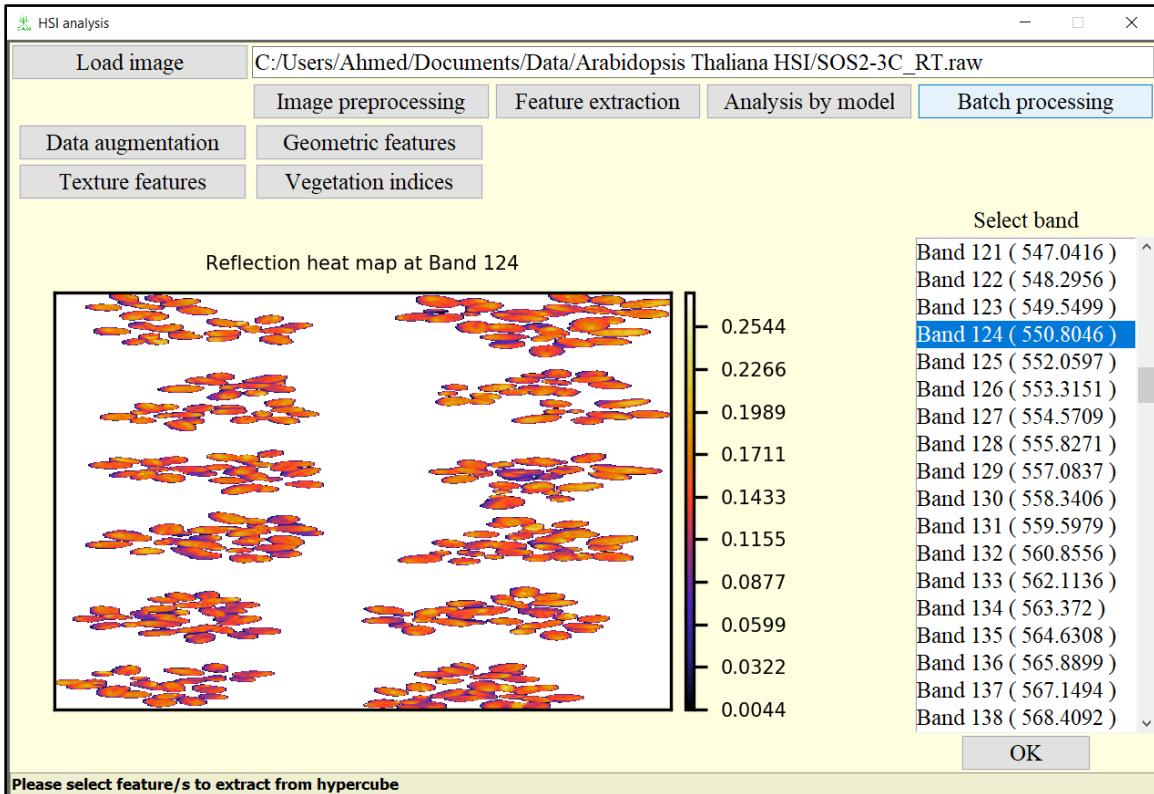
Figure S28. Image resize spectral domain a) select bands from the list b) after click “Done” button.

349

c) insert spatial size, and b) hypercube size after spectral and spatial resize.

350 3.3 Feature extraction

351 After image preprocessing, a new hypercube together with a wavelength vector, a segmented RGB
 352 image and an object mask is obtained, from which spectral, textural and morphological features
 353 can then be extracted. Click the “**Feature extraction**” button then select which features to be
 354 extracted from the new hypercube. (Figure S29).



355

356

Figure S29. Extract features from new hypercube.

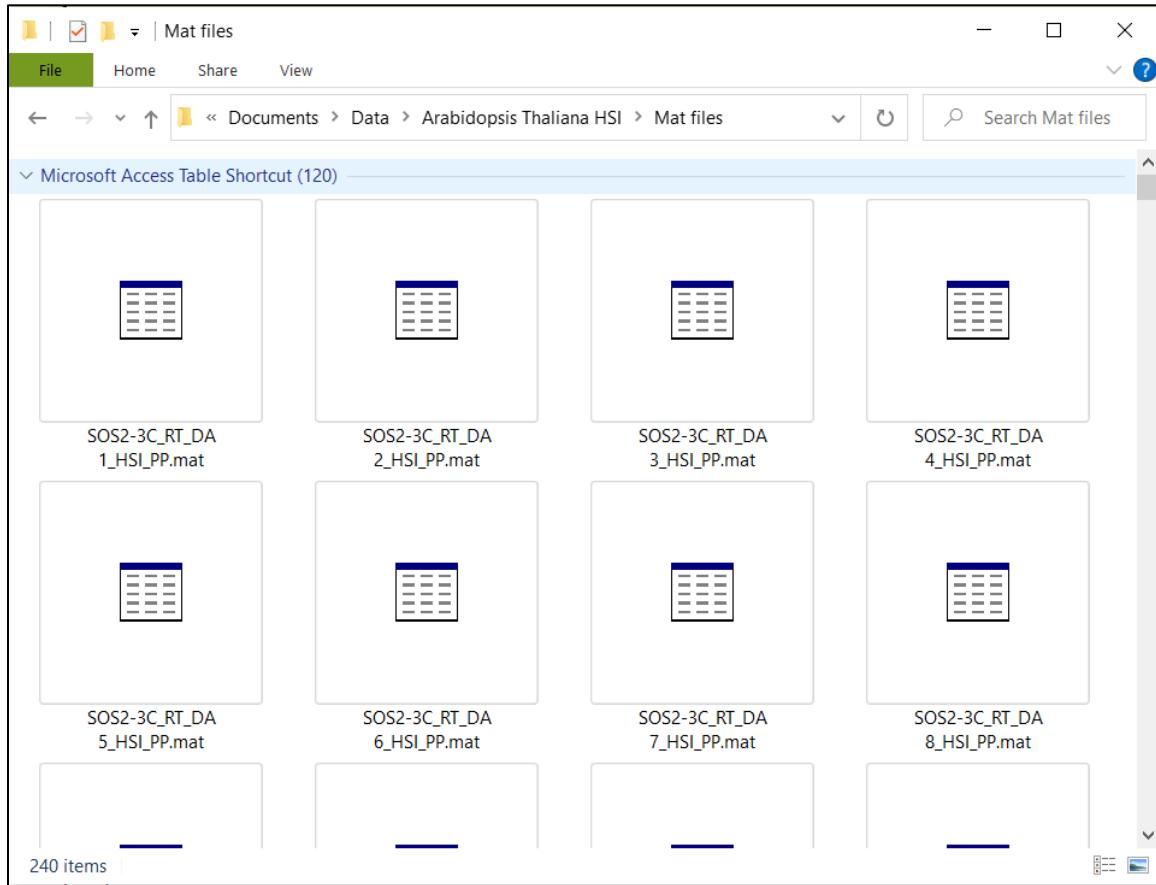
357 *3.3.1 Spectral features*

358 At every aforementioned image preprocessing step, users have options to save the spectral feature
 359 as mean spectrum of the image as an excel file by clicking the “**Save spectrum**” or save new
 360 hypercube as a mat file by clicking the “**Save mat**” button Figures (S17, S18, S24d, S25(c, d) and
 361 S26d).

362 *3.3.2 Image augmentation*

363 For increasing the size of the dataset by flipping, shifting, shearing and zooming each image in the
 364 dataset click the “**Data augmentation**” button the dataset will be increased ten times and saved as

365 mat file (Figure S30).



366

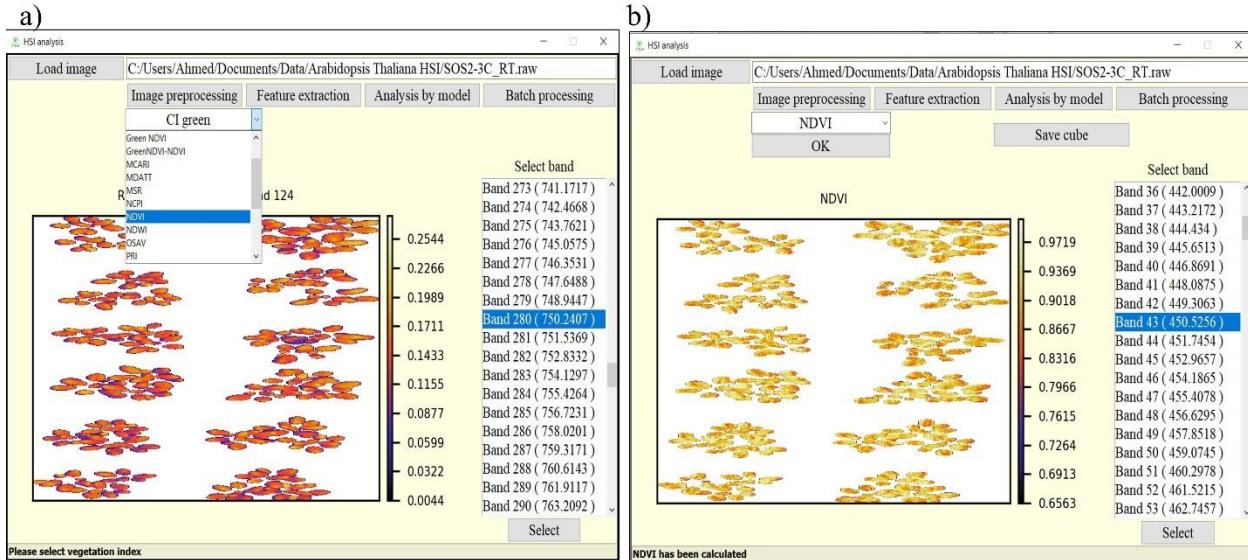
367 **Figure S30.** Save dataset as mat file after augmentation and spatial crop.

368 3.3.3 Vegetation indices

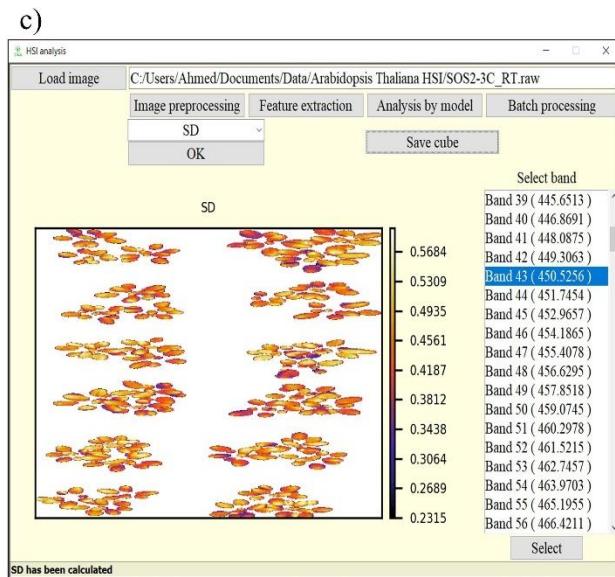
369 To calculate hyperspectral Vegetation Indices, click on the “**Vegetation indices**” button, users can
370 select index (Table S2 in notes S1) from the drop-down box; which are arranged alphabetically
371 then click the “**OK**” button and by **Ctrl-click**, users can select related bands from the list of
372 wavelengths (Figure S31a). Then the software will calculate the selected index after clicking the
373 “**Select**” button (Figure S31 b,c). The calculated vegetation index can be saved as a mat file by
374 clicking the “**Save cube**” button and the metadata shows spatial dimension and the selected bands

375 wavelength values (Figure S31d).

376



377



378

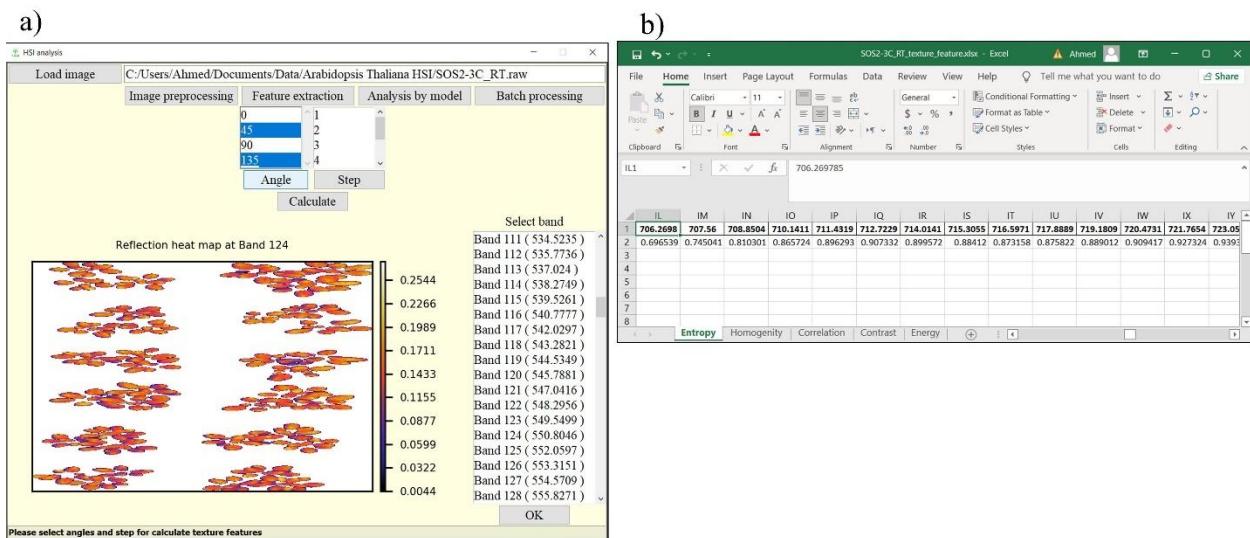
379 **Figure S31.** Calculate vegetation index a) select vegetation and bands, and b) display NDVI, c) display SD, and

380

d) save vegetation index as mat file with metadata as txt file.

381 3.3.4 *Textural features*

382 Various texture features (i.e., contrast, correlation, energy, entropy, and homogeneity) can be
 383 extracted from hypercube at different distances and different directions by click “**Texture feature**”
 384 button. Users can select the distance between 1 to 10 pixel/s and select between four directions (0°,
 385 45°, 90°, and 135°) as shown in (Figure S32a), then click “Calculate” button the texture features
 386 will be calculated and averaged for all steps and angles to give only one value for each band and
 387 be saved as excel file (Figure S32b). These texture features are derived from gray-level co-
 388 occurrence matrix (GLCM).



389

390 **Figure S32.** Select pixels steps and direction angles for calculate texture features.

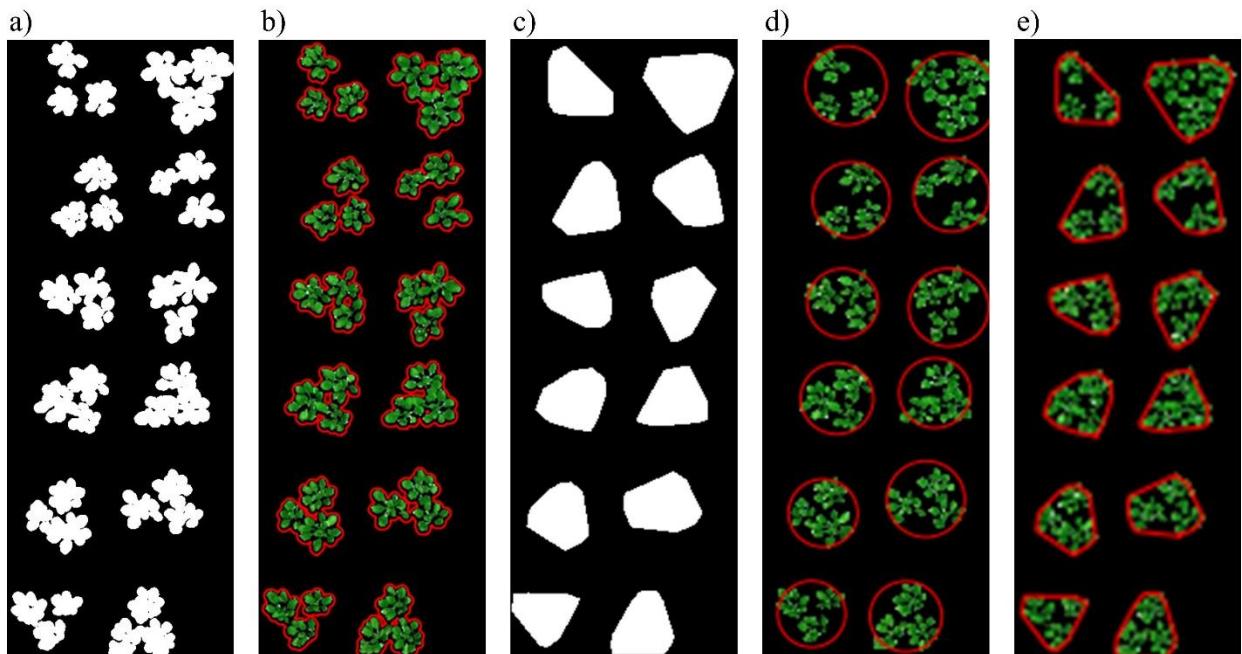
391 3.3.5 *Morphological features*

392 Utilizing the ROI mask (Figure S33a), several growth phenotypes could be quantified reliably for
 393 computing the size variation of a plant over time. Measuring the area of an overhead projection of

394 the plant called “**Projected Leaf Area**” by totaled the green pixels (Figure S33b) to compute the
395 size of the plant. Measure “**Leaf Perimeter**” by calculating the length of the red contour (Figure
396 S33b). “**Leaf Convex Hull**” expresses the extracted leaf region that is enclosed by a 2D convex
397 hull (Jansen *et al.*, 2009), and convex area enveloping all pixels that belong to the plant with a
398 convex polygon (Zhou *et al.*, 2012) (Figure S33c). “**Plant Diameter**” corresponds to the diameter
399 of the maximal circle cover all plant leaves (Figure S33d). “**Stockiness**” is based on the ratio
400 between the leaf projected area and leaf perimeter Eq. (1) (De Vylder *et al.*, 2012).

401
$$S = (4 \times \pi \times A) / P^2 \quad (1)$$

402 Where S is Stockiness, A is a leaf projected area, and P is leaf perimeter. “**Compactness**”
403 expresses the ratio between the leaf projected area and the area of the convex hull of enclosing the
404 rosette (Arvidsson *et al.*, 2011). This is represented in (Figure S33e), where the compactness
405 corresponds to the green area over the area enclosed by the red contour.

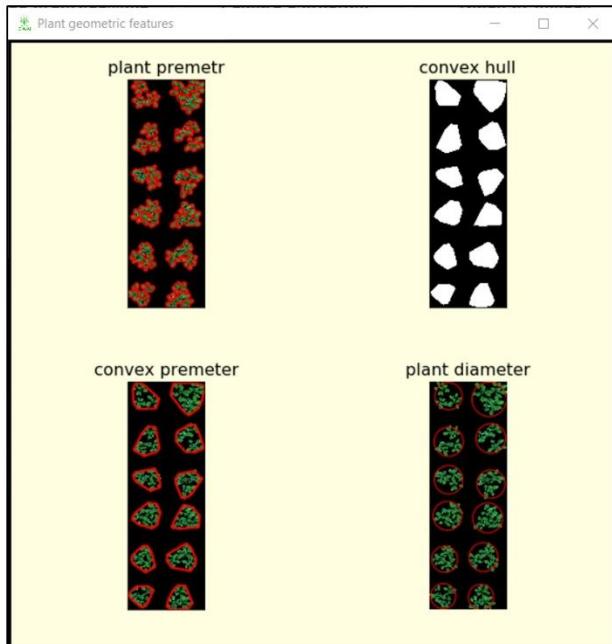


406

407 **Figure S33.** Steps of measuring multiple growth traits. a) ROI masks for every plant
 408 b) Contours generated to
 409 outline the leaf region, and green pixels enclosed by the contours are totaled for computing the size of the projected
 410 leaf area. c) Convex area created for every plant. d) plant diameter according to the diameter of the red circle. e)
 calculate compactness based on the ratio between the plant projected area and convex hull.

411

Click the “**Geometric features**” button, different morphological features will be calculated
 412 and saved as an excel file automatically. The new sub-window will display some of these features
 413 (Figure S34).



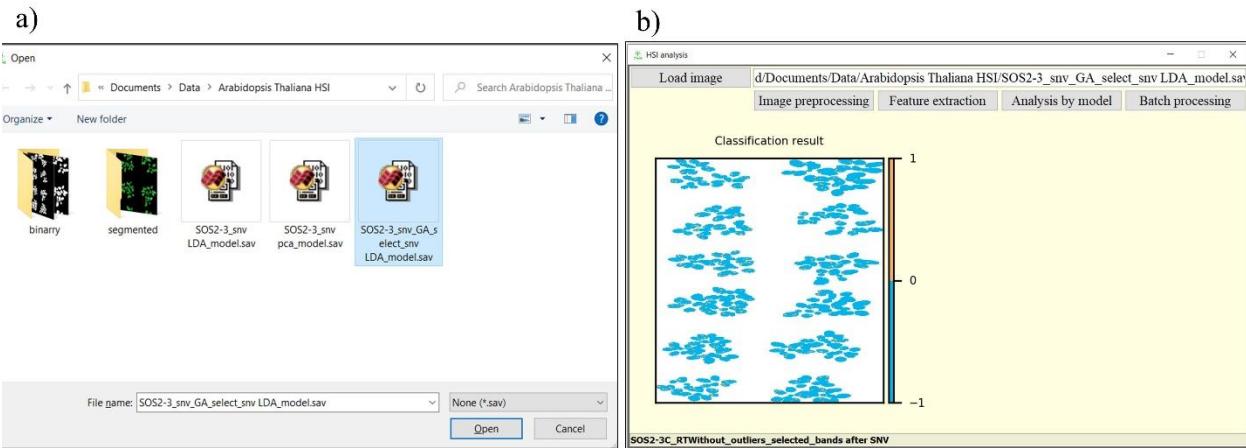
414

Figure S34. Calculate different morphological features, a) select number of sub-images, and b) display

416 morphological features.

417 3.4 Analysis by model

418 To detect the value/class of every pixel in hypercube click “**Analysis by model**” button then import
419 pre-trained model (trained by this software or python see (Data analysis)) (Figure S35a), after that
420 the software will use it and plot the distribution map (Figure S35b).



421

422 **Figure S35.** Analysis hypercube by pre-trained model a) import model and b) classification result.

423 *3.5 Batch processing*

424 For processing a batch of images, the parameters and pipeline chosen in the first image will be
 425 applied for all images in the predefined directory. As shown in (Figure S36), users can toggle one
 426 or multiple checkboxes to specify how many procedures will be processed simultaneously and
 427 how many features to extract. The activated checkboxes according to process applied to first image
 428 (if it applied, it will be activated in this step). Instead of reprocess the first image again, the user
 429 can select which procedure he wants to keep by toggle it (Figure S36a, c) and which he wants to
 430 cancel before start batch processing. (Figure S36b, d).

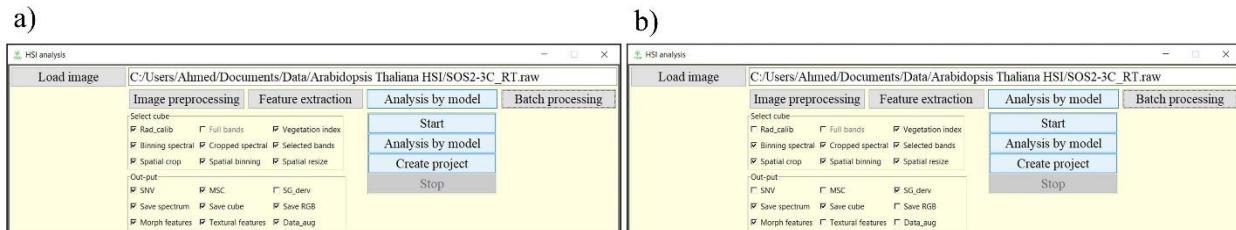
431 From the “**Select cube**” frame toggle one or more checkboxes to select a specific hypercube as
 432 following:

- 433 1) White_calib: calibrate every image,
 434 2) Full bands: calculate spectral features (mean or hypercube) at full wavelength,
 435 3) Vegetation index: calculate vegetation index,

- 436 4) Binning spectral: calculate spectral features (mean or hypercube) after spectral binning,
- 437 5) Cropped spectral: calculate spectral features (mean or hypercube) after deleting
- 438 undesirable bands,
- 439 6) Selected bands: calculate spectral features (mean or hypercube) after selected bands,
- 440 7) Spatial crop: calculate spectral features (mean or hypercube) after split image into sub-
- 441 images; this checkbox can be selected with one of the upper options;
- 442 8) Spatial binning: bin hypercube in spatial dimensions; this checkbox can be selected with
- 443 one of the upper options;
- 444 9) Spatial resize: resize hypercube in spatial dimensions; this checkbox can be selected with
- 445 one of the upper options.

446

447



c)



d)



448

449

Figure S36. Checkboxes have to be selected for processing batch of images (a, b) and the flow chart of

450

processing according to selected checkbox (red is unselect procedures).

451

From the “Out-put” frame toggle one or more checkboxes for saving different features as

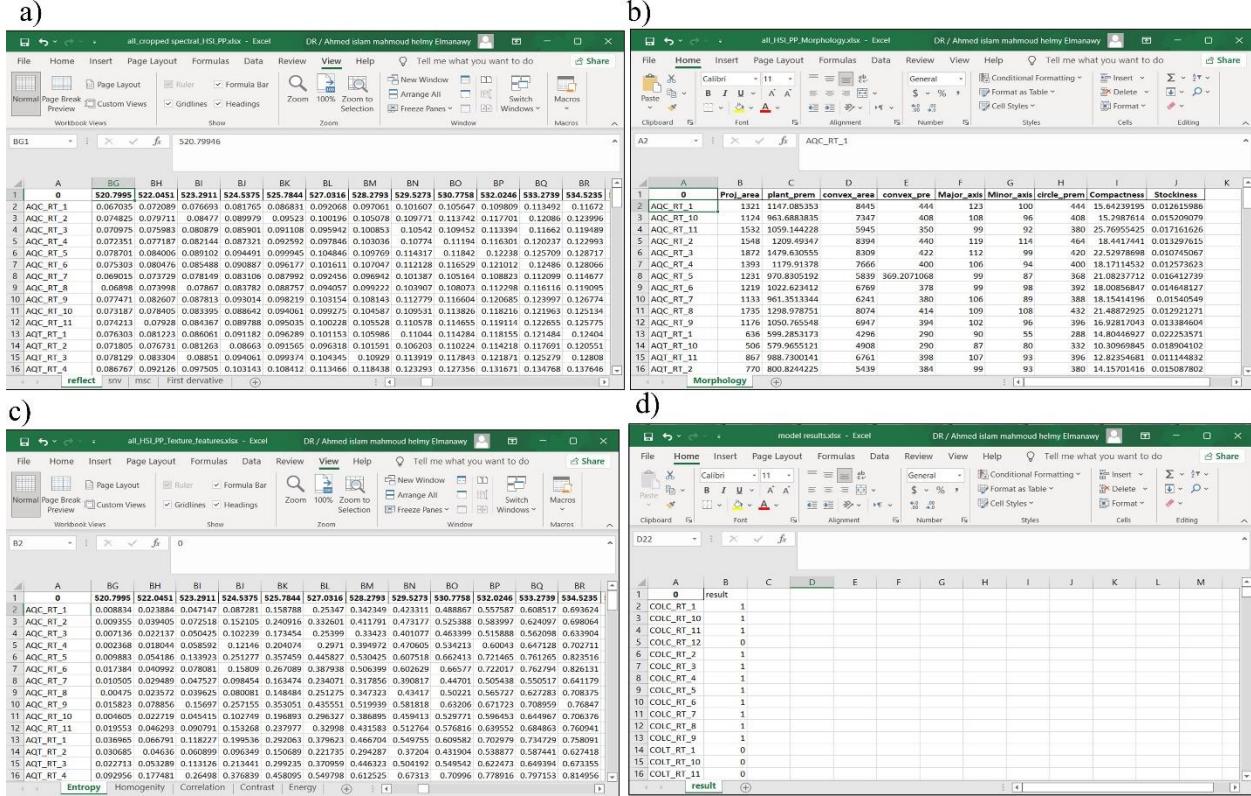
452

following:

453 1) SNV: calculate standard normal variate from new hypercube,
454 2) MSC: calculate multi scatter correction from new hypercube,
455 3) SG_derv: calculate first or second derivative from new hypercube,
456 4) Save spectrum: save mean spectrum and/or SNV, MSC, SG_derv in different tabs in excel
457 file in addition to the image name and wavelength values,
458 5) Save cube: save hypercube as mat file and save metadata for this hypercube as txt file
459 which have the same name as image,
460 6) Save RGB: save segmented RGB image and masked image,
461 7) Morph features: calculate morphological features and save them as an excel file that have
462 the same name as image,
463 8) Texture features: calculate textural features from new hypercube and save it as excel have
464 the same name as image,
465 9) Data_aug: augment new hypercube and save it as mat have the same name as image.

466 During the batch processing, many folders will be created with processed results (e.g.,
467 segmented RGB images, mat files and excel files) to save extracted features. Once toggle one or
468 more of the checkbox aforementioned then click the “**Start**” button, all qualified images (with the
469 same extension as the first image) in a directory will be analyzed. Status bar will show the progress
470 of the software through analysis. To stop the batch of images process for making any modification,
471 click the “**Stop**” button then users can restart it by clicking the “**Start**” button. At this point, users
472 have two options. The first one is saving these features for training models (Figure S37a-c), or

- 473 analyzing these features using a pre-trained model by clicking the “Analysis by model” button to
 474 import this model then the result will saved as excel file (Figure S37d).



The figure displays four Excel spreadsheets (a, b, c, d) showing batch processing results:

- a)** all_cropped_spectral_HSI_PP.xlsx: Contains spectral features like RG, RH, RI, BI, RK, RL, BM, BN, RO, RP, RQ, and BB.
- b)** all_HSI_PP_Morphology.xlsx: Contains morphology features such as Proj_area, plant_prem, convex_area, convex_per, Major_axis, Minor_axis, circle_prem, Compactness, and Skewness.
- c)** all_HSI_PP_Texture_features.xlsx: Contains textural features like Entropy, Homogeneity, Correlation, Contrast, Energy, and others.
- d)** model_results.xlsx: Shows model results for various regions (e.g., AQC_RT_1 to AQC_RT_11, COLC_RT_1 to COLC_RT_11, AQCT_RT_1 to AQCT_RT_4) across columns D through M.

- 475
 476 Figure S37. Batch processing output a) spectral features, b) morphology features, c) textural features and d)
 477 model results.

- 478 When users are satisfied with the parameters and pipelines, it can be saved as project with
 479 default name (HSI-PP_prj.py) (Figure S38) by clicking the “Create project” button, which can
 480 be applied to other images to extract the same features after same preprocessing procedures
 481 (selected in batch processing Figure S36). That project file can be run by this software as will be
 482 mentioned in part 5 (Import project).

```

global Type, analysis_img
def Type():
    Type= '.raw'
    return Type
def analysis_img(image_path):
    HSI_Img = open_hsi_img(image_path)
    HSI_Img=extract_hsi(HSI_Img)
    RGB_Image=open_rgb_img(image_path)
    blur = cv2.GaussianBlur(RGB_Image, (5, 5), 0.75)
    HSV = (cv2.cvtColor(blur, cv2.COLOR_BGR2HSV)).reshape((RGB_Image.shape[0]*RGB_Image.shape[1]),3)
    YCrCb = (cv2.cvtColor(blur, cv2.COLOR_BGR2YCrCb)).reshape((RGB_Image.shape[0]*RGB_Image.shape[1]),3)
    LAB = (cv2.cvtColor(blur, cv2.COLOR_BGR2LAB)).reshape((RGB_Image.shape[0]*RGB_Image.shape[1]),3)
    RGB=blur.reshape((RGB_Image.shape[0]*RGB_Image.shape[1]),3)
    Image=[]

```

483

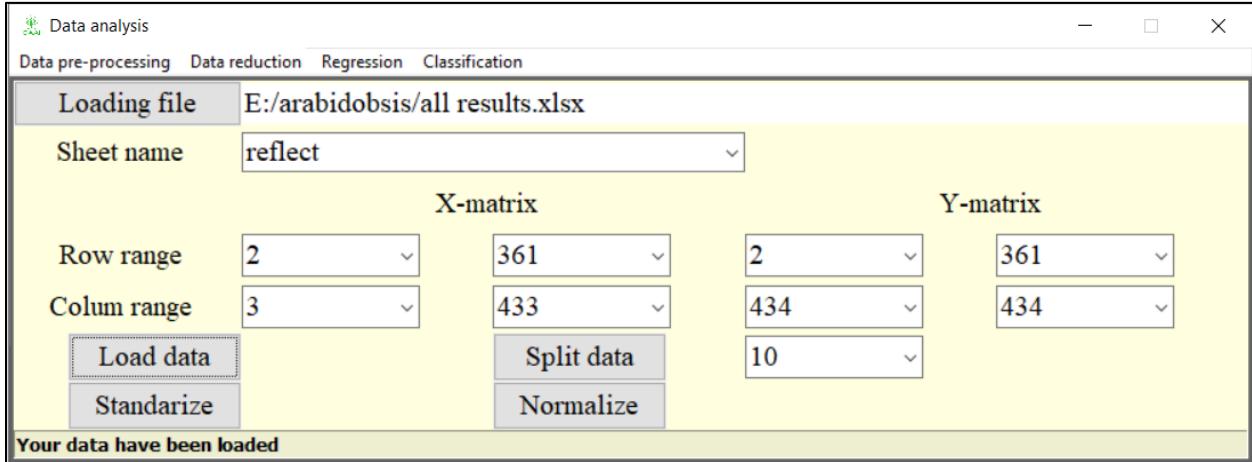
484

Figure S 38. The script of the saved project.

485 4. Data analysis

486 HSI-PP provides different methods for modeling spectral datasets that are collected from
 487 hyperspectral cameras or spectroradiometers (e.g., ASD). A number of appropriate multivariate
 488 chemometric schemes such as multiple linear regression (MLR), partial least square (PLS)
 489 regression, principal component analysis (PCA), linear discriminant analysis (LDA), support
 490 vector machine (SVM), random forest (RF), and multilayer perceptron (MLP) are available in the
 491 software to build calibration models to relate spectral data (X-matrix) with the quantitative data
 492 (Y-matrix). Users can load datasets from excel files which contain the phenotypic traits (e.g.,
 493 average spectrum) (independent variable, X matrix) and the ground truth measurements for
 494 regression models or code number for classification models (dependent variable, Y matrix). Users
 495 can just simply choose the rows and columns that represent every matrix. As shown in (Figure
 496 S39), users select the excel file (.xlsx, .xls, .csv) by clicking the “**Loading file**” button in the “**Data**
 497 **analysis**” window and select direction and files. The phenotypic traits need to be located in
 498 columns and the sample data in rows. From “**Sheet name**” select the sheet containing data. After

499 loading data from the file by clicking the “**Load data**” button, it is ready for analyzing directly or
500 preprocessing. The software has no limitation on the number of phenotyping traits or response
501 variables.



502

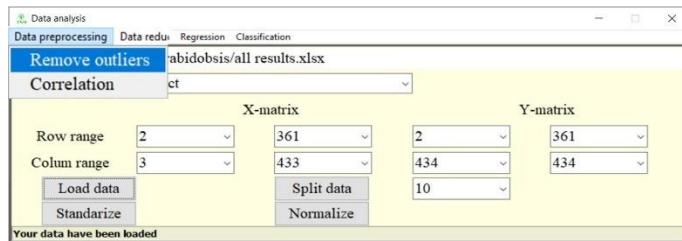
503 **Figure S39.** Data analysis main window.

504 *4.1 Data preprocessing*

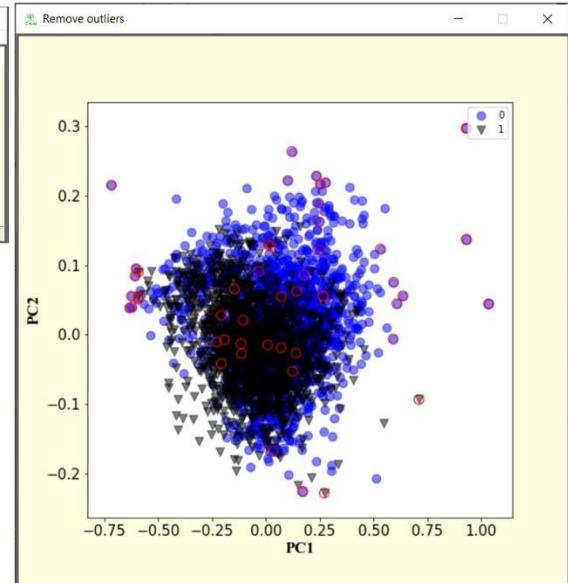
505 The preprocessing of extracted features is needed before data analysis to remove noise. Data
506 preprocessing of HSI-PP consists of normalization, standardization, outlier filtration, and
507 correlated feature removal. Users can normalize or standardize datasets by clicking the
508 “**Normalize**” or the “**Standardize**” button (Figure S39). From the “**Data pre-processing**” menu,
509 users can select between **Remove outliers** to remove outliers’ points in imported data, or
510 **Correlation** to remove correlated features (Figure S40a). At the **Remove outliers** interface, PCA
511 is used for detecting outliers in data and automatically removed (Figure S40b). The reproducibility
512 of extracted features is evaluated by the Pearson correlation coefficient to select uncorrelated
513 features (Figure S40c). The correlation coefficient values are computed over each pair of features

514 before inserting the max coefficient value (Figure S40d).

a)

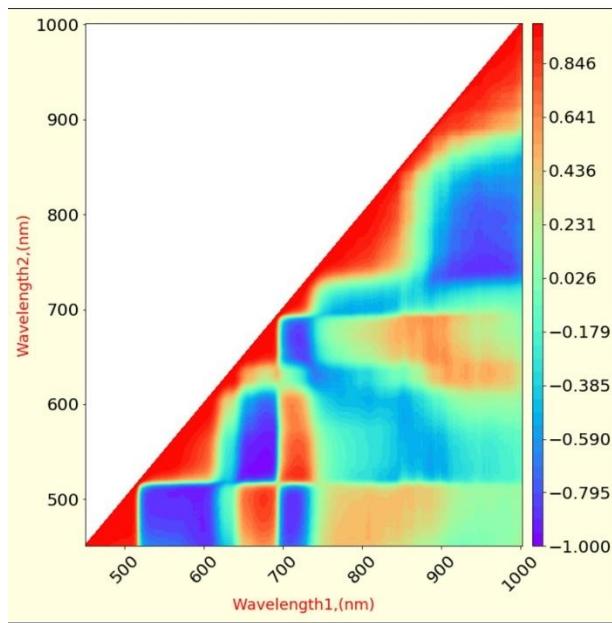


b)

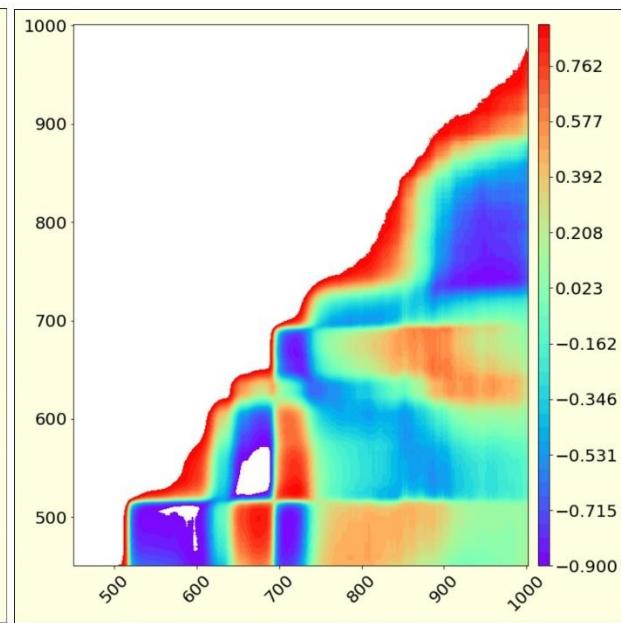


515

c)



d)



516

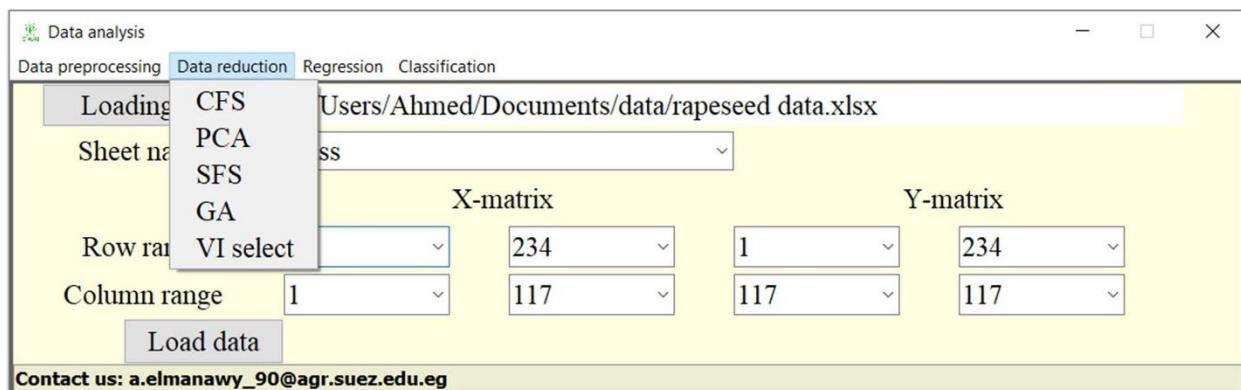
517 **Figure S40.** Data preprocessing. a) select preprocessing procedure, b) remove outliers, c) Pearson correlation coefficient between pair of wavelengths, and d) select wavelength with lower corelated.

519 Users can divide data using the Kennard-Stone algorithm (Kennard and Stone, 1969) into two

520 parts, one for model training and the other for model validation. The Kennard-Stone algorithm
521 allows selecting calibrated samples according to a uniformed sampling procedure, which is aimed
522 at maximizing the minimum distance between pairs of selected samples in the X-matrix. Users can
523 select the percentage of validation data from the drop-down combo (Figure S39) and click the
524 “Split data” button to split the dataset into calibration and validation sets.

525 *4.2 Data reduction*

526 Data reduction is necessary to remove redundant phenotypic information and overcome the
527 problem of model overfitting. HSI-PP contains five statistical methods that are popular for
528 dimensionality reduction and projection of high-dimensional datasets. The “Data reduction” menu
529 contains five reduction methods (CFS, PCA, SFS, GA, and VI-select) (Figure S41).

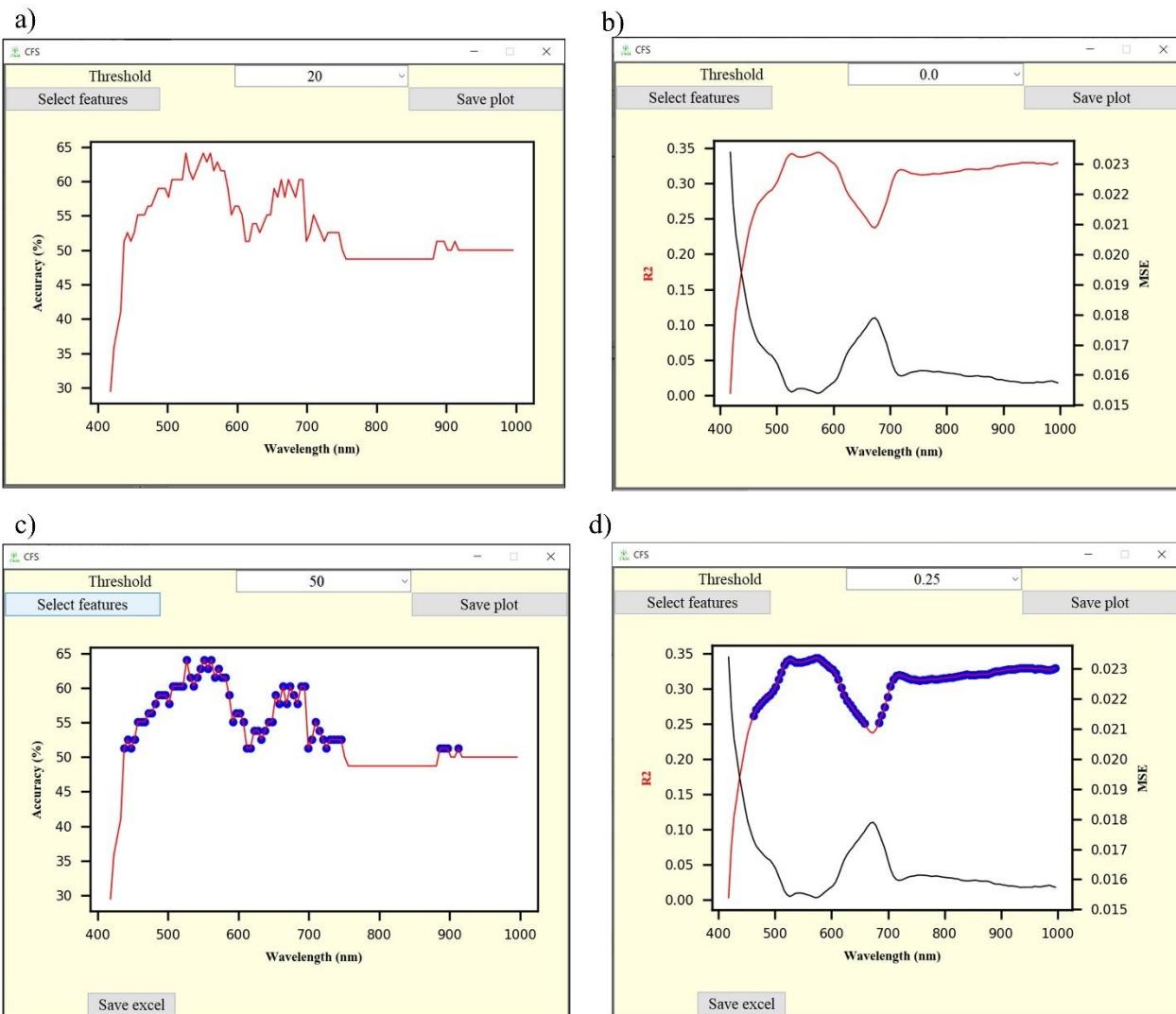


531 Figure S41. Data reduction menu in HSI-PP software.

532 *4.2.1 Correlation-based feature selection (CFS)*

533 CFS is a univariate filter method that measures the Pearson's correlation coefficient between
534 features and the output variable. Once users select “CFS” from the data reduction menu, HSI-PP

535 calculates the Pearson's correlation coefficient between every feature and the output value (for
 536 regression) or classification accuracy, then plot those results (Figure S42 a, b). The user can select
 537 threshold value (Figure S42 c, d), then the features which achieved accuracy higher than threshold
 538 value will be saved as an excel file after pressing the “Save excel” button. That plot can be saved
 539 by clicking the “Save plot” button.



540

541

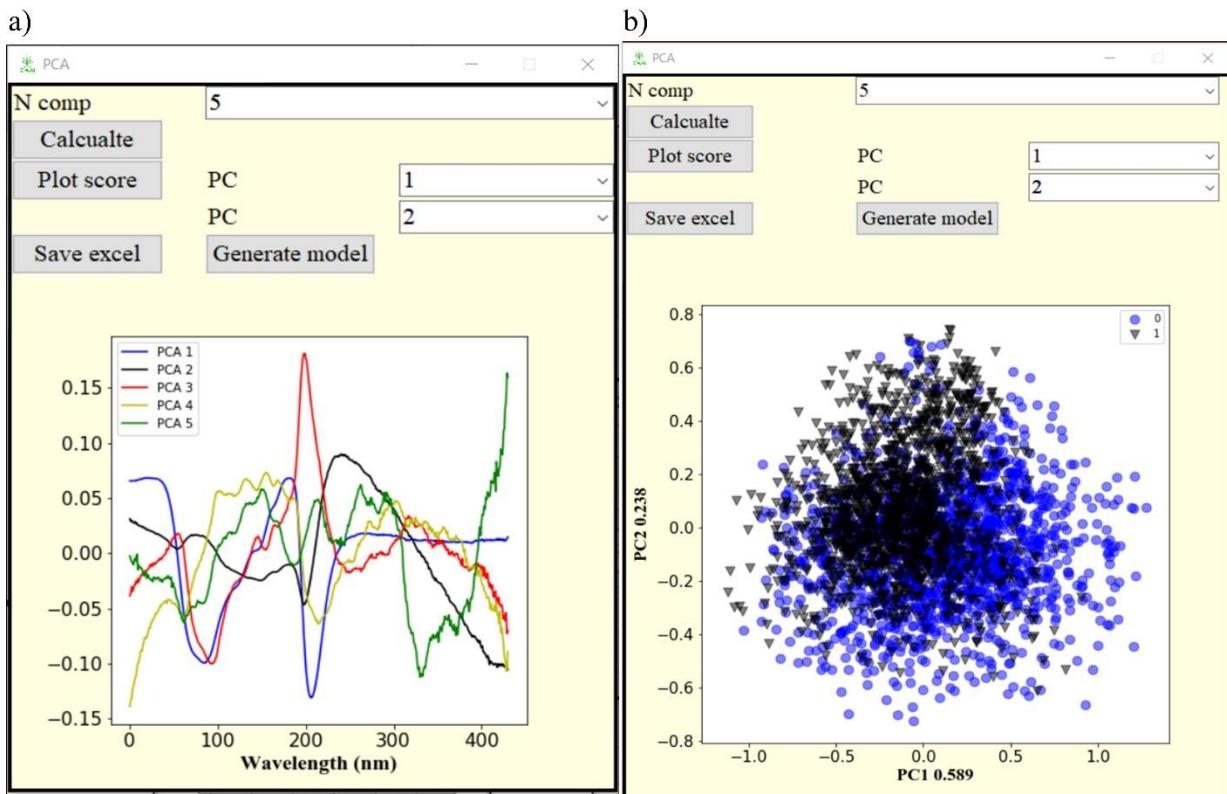
542

543

Figure S42. Correlation-based feature selection (CFS), for classification (a), for regression (b), select wavelength with higher accuracy (c) and select wavelength with higher coefficient (d).

544 4.2.2 Principle component analysis (PCA)

545 PCA is used as a feature extraction and selection method. Feature extraction by transforming
546 datasets into a new orthogonal feature space, where the major part of the data variance can be
547 expressed by the first few principal components (PCs). Feature extraction by selecting the optimal
548 features depends on the values of PCA loading. Select “PCA” from the data reduction menu, then
549 select the number of components. The PCA loadings will be plotted (Figure S43a) and users can
550 plot PC scores by selecting the first and second scores (for example) for the plot (Figure S43b).
551 The scores, loadings, plots and PCA models can be saved as an excel file if needed by clicking the
552 “Save excel” button.



553

554 **Figure S43.** Reduce data by principal component analysis using loading value (a), or score (b).

555 4.2.3 Sequential forward selection (SFS)

556 SFS starts with no features and adds them one by one. SFS adds one that decreases the error the
557 most at each step until any further addition does not significantly decrease the error (Serpico and
558 Bruzzone, 2001). Once users select “SFS” from the data reduction menu, HSI-PP calculates and
559 selects 20 features (default) then plots its performance (Figure S44a). User can then change the
560 feature number (2 to N, where N total number of features) (Figure S44b).

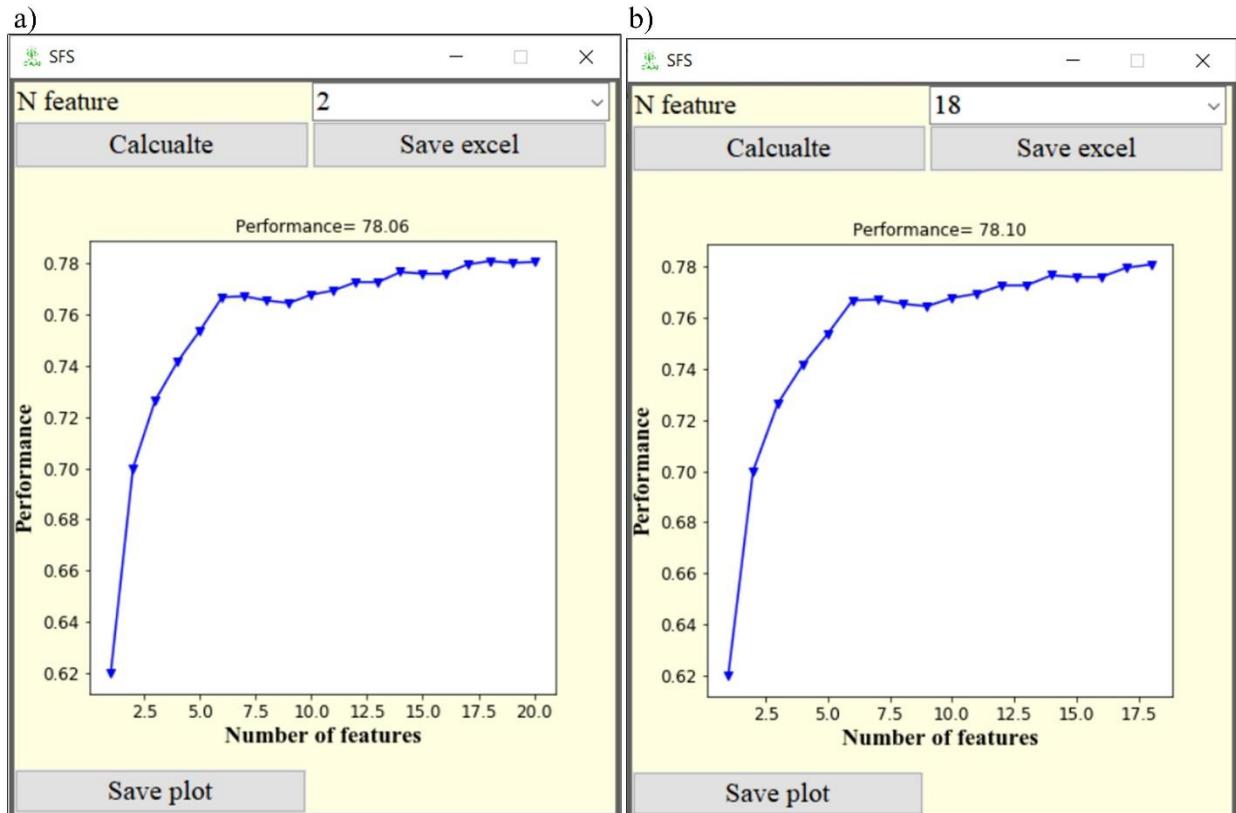
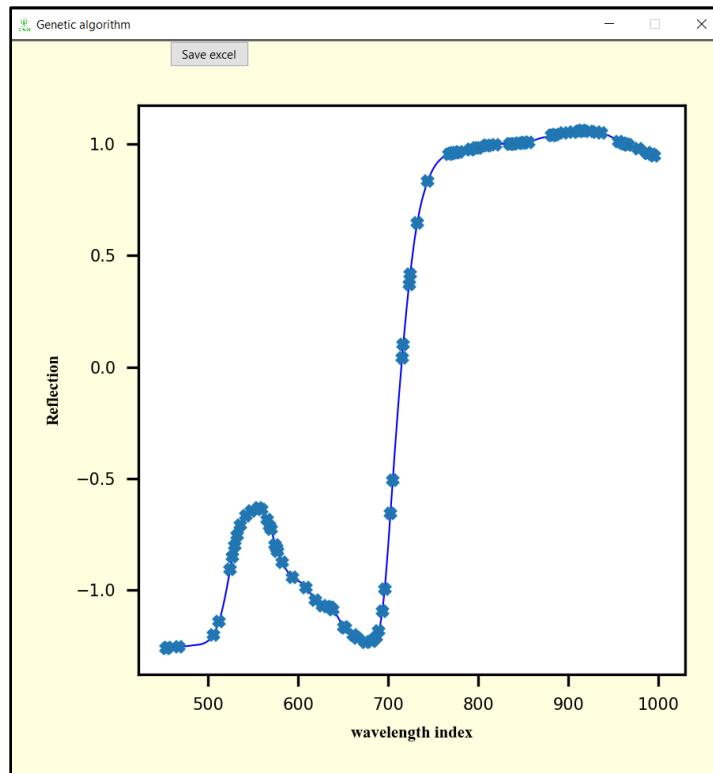


Figure S44. Sequential features selection.

563 4.2.4 Genetic Algorithm (GA)

564 GA is population-based stochastic search optimization techniques inspired by natural selection and

565 natural genetics principles (Goldberg and Holland, 1988). Once select “GA” from the data
566 reduction menu, HSI-PP will operate GA and reduce features to about 20% before plotting the
567 selected phenotypic traits (Figure S45).

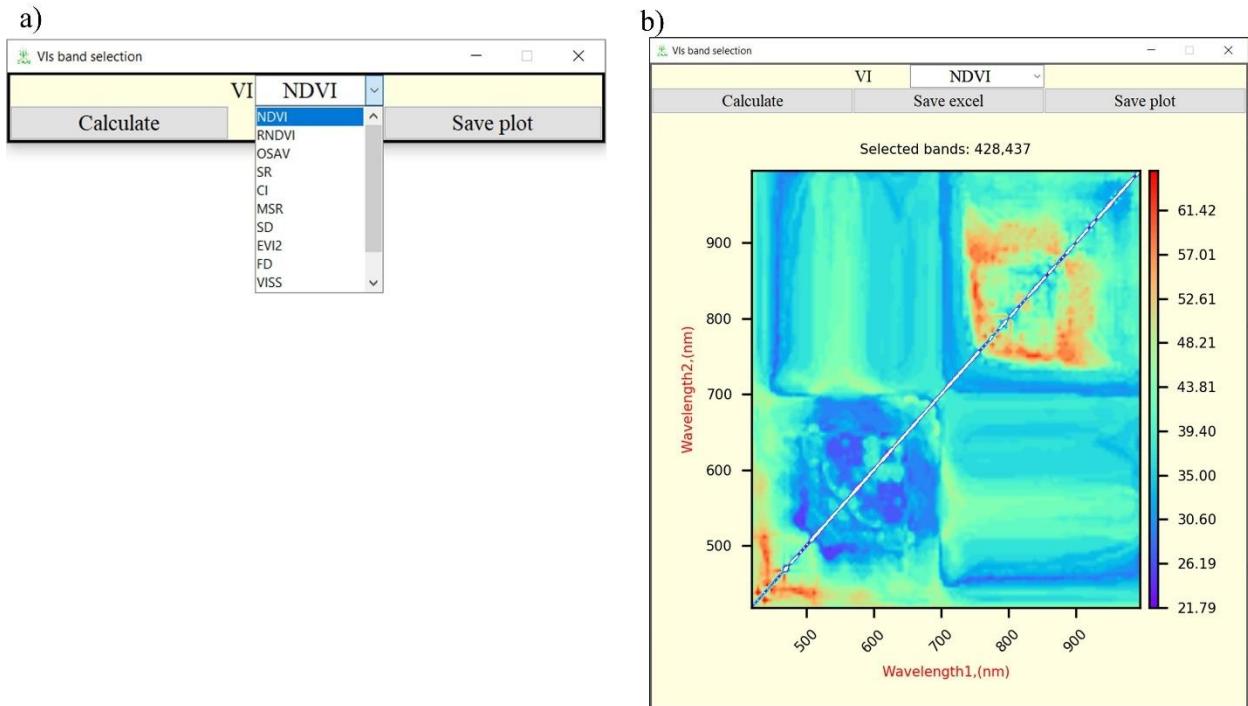


568
569 Figure S45. Genetic algorithm result (the line is mean spectrum and “X” is selected features)

570 4.2.5 Vegetation index band selection (VI select)

571 To further evaluate the appropriate hyperspectral region and select the optimal effective
572 wavelengths, contour plot of r^2 values for the correlation of respond with specific vegetation index
573 (i.e., simple ratio) were involved in HSI-PP. After users select “VI” from the data reduction menu
574 and select the vegetation index from the drop-down box (Figure S46a), HSI-PP plots that contour
575 plot (Figure S46b), and the best two bands will be saved as an excel file after pressing the “Save

576 excel” button. That plot can be saved by clicking the “Save plot” button.



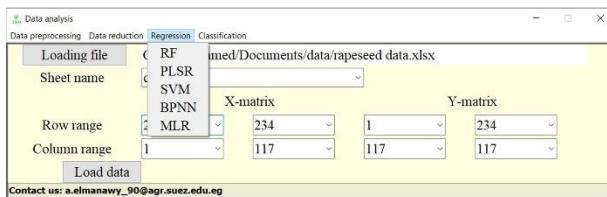
577

578 Figure S46. Vegetation index two bands selection in HSI-PP software, select vegetation index (a), and Contour
579 maps of the coefficients of determination or segment accuracy between NDVI and respond (b).

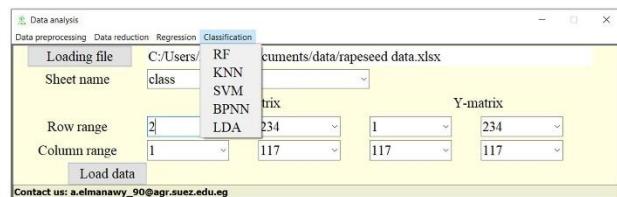
580 4.3 Model generation

581 Creating an appropriate model is the last part of the process of establishing a decision-making
582 system. The algorithms implemented in this software are some of the most used ones in phenotypic
583 traits analysis including, random forest (RF); support vector machine (SVM); multi-layer
584 perceptron (MLP); K-nearest neighbor (KNN); linear discriminant analysis (LDA); multilinear
585 regression (MLR) and partial least square regression (PLSR). That algorithms were implemented
586 in the “Regression” menu for generating regression model regression (Figure S47a) or the
587 “Classification” menu for generating classification model (Figure S47b).

588 a)



589 b)



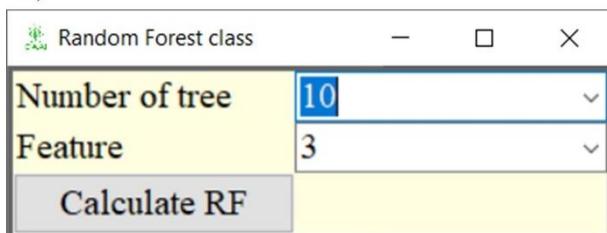
588

589 Figure S47. Model generation menus in HSI-PP software, the regression menu (a), and the classification menu (b).

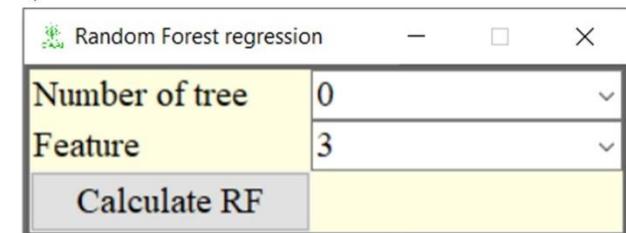
590 4.3.1 Random forests

591 Random forests algorithm is an ensemble method developed by (Breiman, 2001). It is employed
592 to regularize the greedy and heuristics nature of the decision tree training which sometimes causes
593 overfitting. In this software, users can choose the number of trees for model training and the
594 number of features at every tree node (Figure S48). If users set the number of trees at zero, HSI-
595 PP will select the hyperparameters automatically.

a)



b)



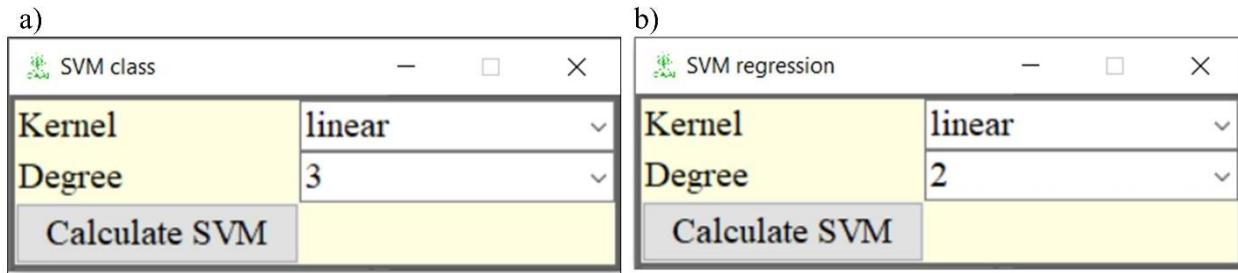
596

597 Figure S48. Random forest hyperparameters options in HSI-PP software a) classification window, and b)
598 regression window.

599 4.3.2 Support vector machines

600 SVM is used for supervised classification and regression providing linear and non-linear
601 discrimination functions (Cortes and Vapnik, 1995). In HSI-PP, users can change between different

602 kernels (linear, polynomial, sigmoid, or radial basis function), select the degree of polynomial
603 kernel, and the software will get the optimal regularized parameters (Figure S49) to achieve higher
604 accuracy for classification/regression model.



605
606 **Figure S49.** Support vector machine hyperparameters options in HSI-PP software a) classification window, and
607 b) regression window.

608 4.3.3 Multi-layer perceptron

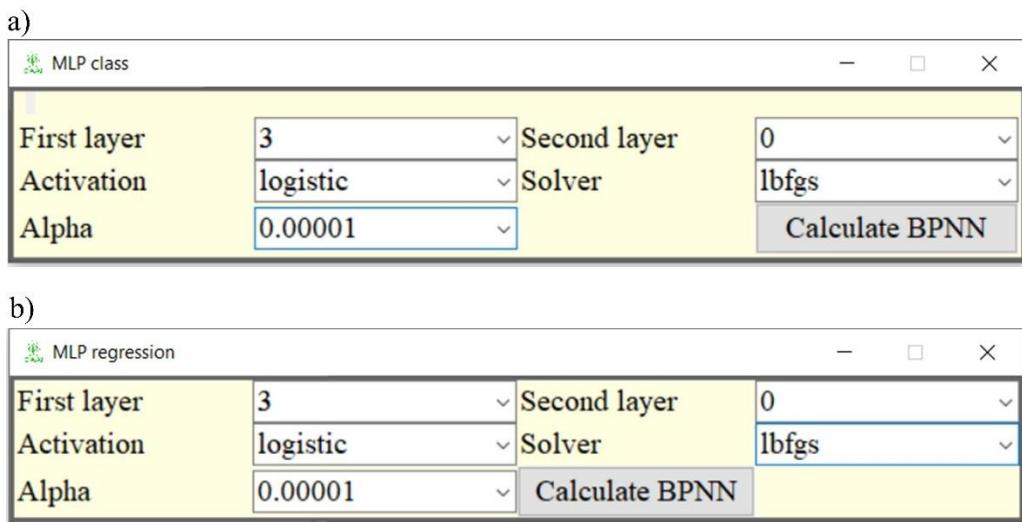
609 Artificial neural networks or neural networks are an established compact model, which is a
610 representation for the analysis of high-dimensional data (Bishop, 1995). The MLP in HSI-PP
611 consists of two hidden layers with the following adjustable hyperparameters: activation function
612 type (linear, logistic sigmoid, hyperbolic tan, and rectified linear unit), number of neurons in
613 hidden layers, weight optimization (Limited-memory BFGS, stochastic gradient descent, and
614 stochastic gradient-based), and mean square error for learning to stop (Figure S50). Users can set
615 the number of neurons in the first layer at zero and the software will select the number of neurons
616 automatically.

617 To improve the predictive capacity of the classification model and reduce data dimensionality,
618 Eq. (2), (Glorfeld, 1996) is implemented in this software to select the most informative wavelength.

619

$$M = \frac{\sum_{j=1}^{n_H} [(|I|_{P_j}/\sum_{k=1}^{n_p} |I|_{P_{j,k}}) |O_j|]}{\sum_{i=1}^{n_p} (\sum_{j=1}^{n_H} [(|I|_{P_{i,j}}/\sum_{k=1}^{n_p} |I|_{P_{i,j,k}}) |O_j|])} \quad (2)$$

620 Where M is the importance measure for the input variable (band), n_p is the number of input
 621 variables, n_H is the number of hidden layer nodes, $|I|_{pj}$ is the absolute value of the hidden layer
 622 weight corresponding to the p^{th} input variable and the j^{th} hidden layer, and $|O_j|$ is the sum of the
 623 absolute value of the output layer weight corresponding to the j^{th} hidden layer. The index M value
 624 is calculated for each input node and then normalized. The higher the M value, the more important
 625 band is for the classification of genotypes.



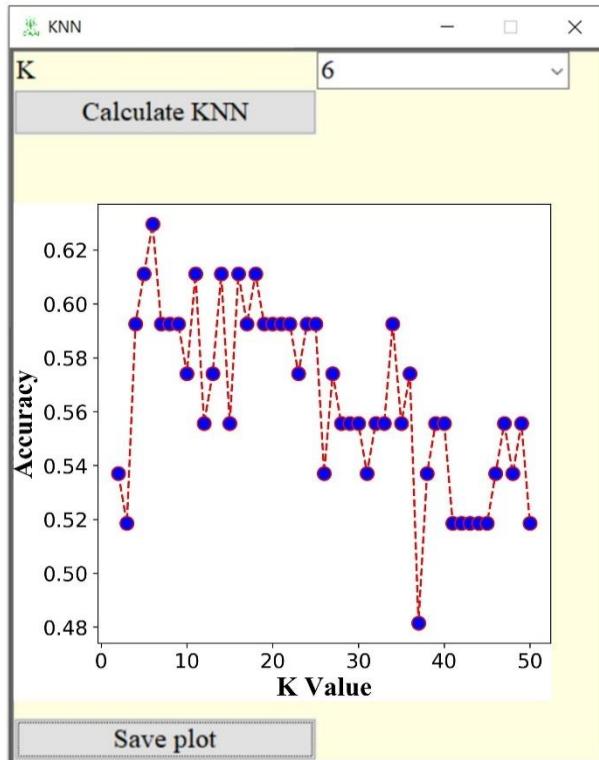
626

627 **Figure S50.** Multilayer perceptron neural network hyperparameters a) classification window, and b) regression
 628 window.

629 *4.3.4 k-nearest neighbor*

630 KNN is one of the simplest methods for pattern classification. It is a non-parametric method for
 631 classifying objects based on the closest training data in feature space (Massart *et al.*, 1998). Once

632 users choose “KNN” from the classification menu, HSI-pp will plot the accuracy with different k
633 values (from 2 to 50) (Figure S51). Users can then select the optimal value to generate the model.

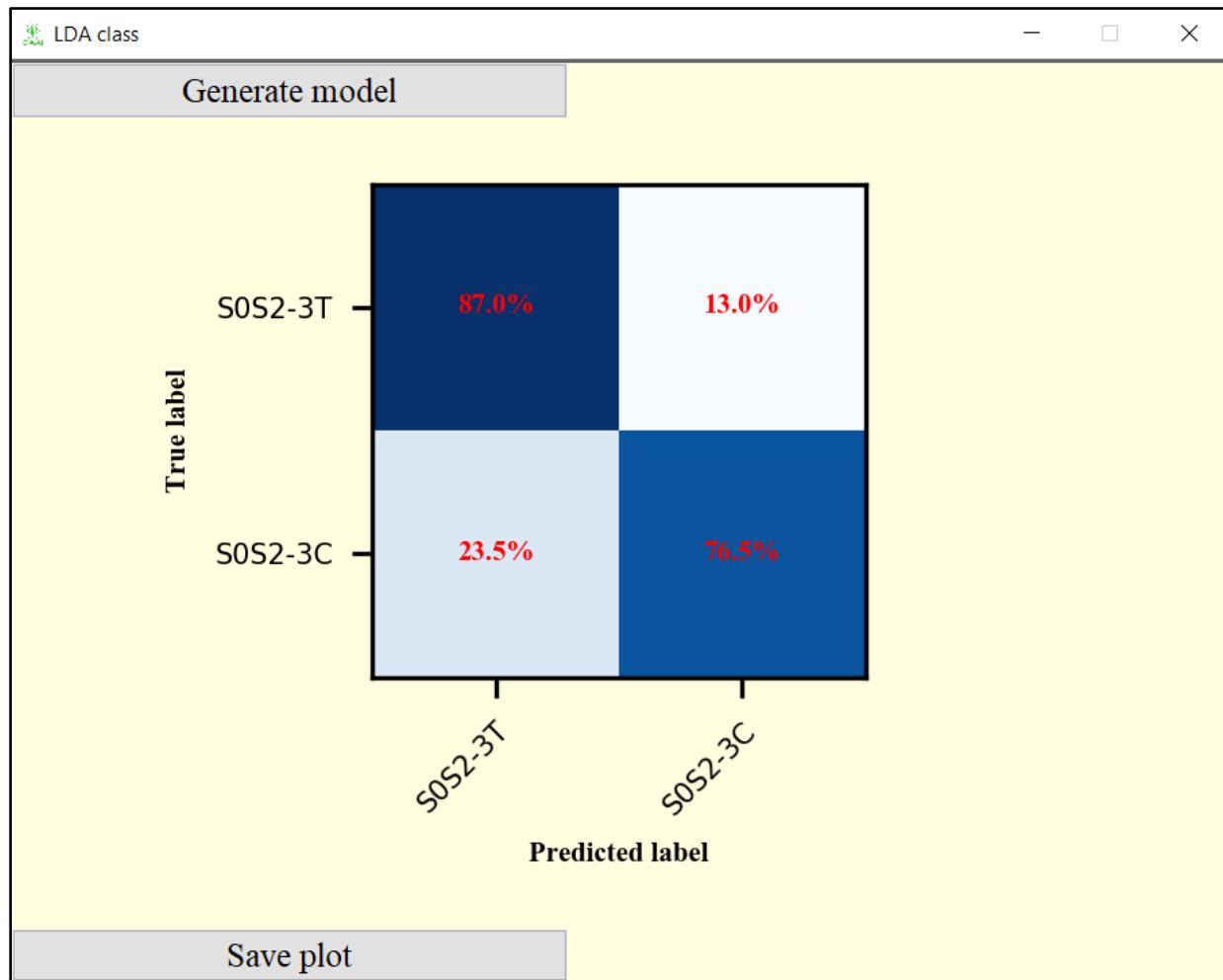


634

635 **Figure S51.** Select the optimal K value for K-nearest neighbors from plot.

636 4.3.5 Linear discriminant analysis

637 LDA is a method for a linear combination of features that separates two or more classes of objects
638 (Martínez and Kak, 2001). Select “LDA” from the classification menu, and the software will
639 conduct LDA and display the results (Figure S52).



640

641

Figure S52. Linear discrimination analysis classification results.

642 *4.3.6 Multiple linear regression*

643 MLR is a conventional tool to model the relationship between a response variable and multiple
 644 predictor variables (Freedman, 2009). Select “MLR” from the regression menu, and the software
 645 will conduct MLR and display the results.

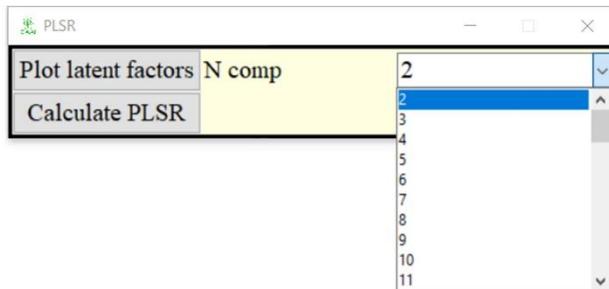
646 *4.3.7 Partial least square regression*

647 After selecting the number of latent variable from the drop-down box (Figure S53a) the software

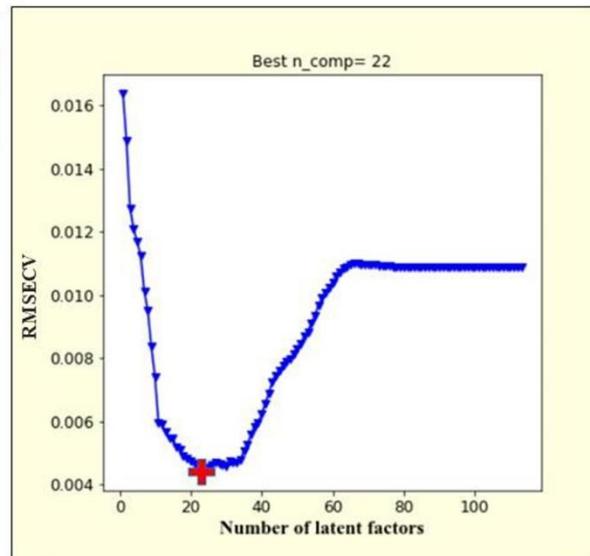
648 will calculate PLSR. In addition, the optimal wavelengths can be identified by the Backward
 649 Variable Selection method for PLS regression (BVSPLS) (Pierna *et al.*, 2009). As it is shown in
 650 (Figure S43a), by clicking the “**Plot latent factor**” button, the software automatically selects
 651 RMSECV (Figure S53b). The optimal number of latent variables (LVs) is determined
 652 automatically at the minimum value of the root mean squared error during leave-one-out full cross-
 653 validation (RMSECV) Eq. (3).

654
$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

a)



b)



655

656 **Figure S53.** Training PLSR model. a) manual select latent factor, and b) plot latent factor and get the optimal
 657 value (red cross).

658 *4.4 Evaluation of generated models*

659 HSI-PP, as an all-in-one package, delivers comprehensive outputs to users. The confusion matrix
 660 gives a full description of errors made by classifiers (Stehman, 1997). In this matrix, the true labels

661 and the predicted labels are displayed. For classifier models, the confusion matrix is displayed with
 662 its evaluation criteria, i.e. precision or positive predictive value (Pr), sensitivity, hit rate, recall,
 663 true positive rate (Se), specificity, true negative rate (Sp), and overall accuracy of the model (Acc)
 664 (Figure S54). Precision represents the ratio of true positives. The recall represents the fraction of
 665 points that have been correctly detected. These measures are based on four different criterion
 666 parameters such as True Positive (TP), True Negative (TN), False Positive (FP), and False
 667 Negative (FN), which will be defined as in the following equations (4-8) (Powers, 2011). If the
 668 result is accepted the user can save the model by clicking “**Generate model**” button in every sub-
 669 window and the model result saved by clicking “**Save plot**” button.

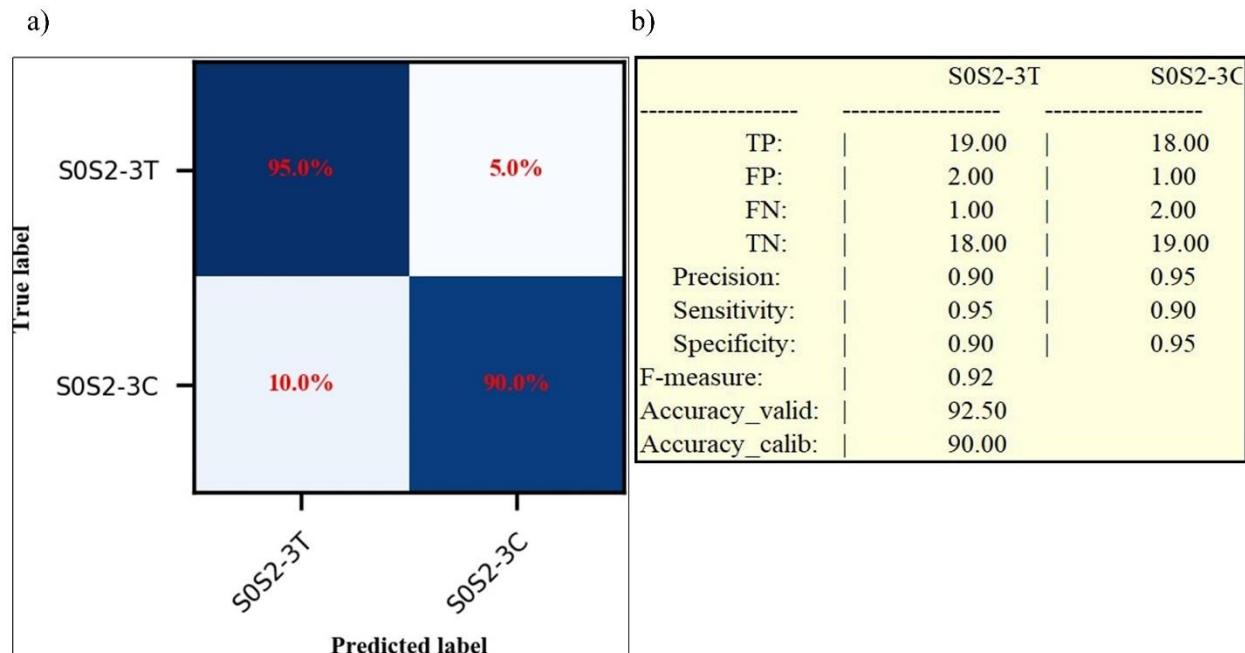
$$670 \quad Acc = \frac{\Sigma TP + \Sigma TN}{\Sigma TP + \Sigma TN + \Sigma FP + \Sigma FN} \times 100 \quad (4)$$

$$671 \quad Pr = \frac{\Sigma TP}{\Sigma TP + \Sigma FP} \times 100 \quad (5)$$

$$672 \quad Se = \frac{\Sigma TP}{\Sigma TP + \Sigma FN} \times 100 \quad (6)$$

$$673 \quad Sp = \frac{\Sigma TN}{\Sigma TN + \Sigma FP} \times 100 \quad (7)$$

$$674 \quad F - measure = \frac{2 * Pr * Se}{(Pr + Se)} \quad (8)$$



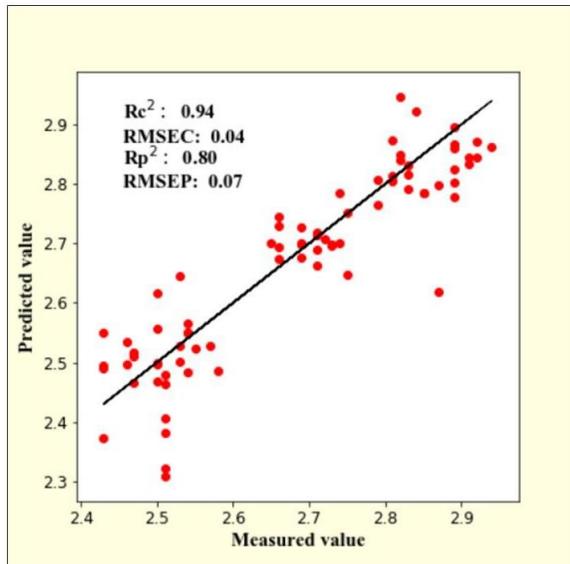
675

676 **Figure S54.** Evaluate classification model using confusion matrix (a), and statistical measurements (b).

677 Similarly, in the case of regression models, target values versus predicted values are plotted in
 678 a graph and corresponding statistical measures, i.e. correlation coefficient R^2 Eq. (9), root mean
 679 square error (RMSE), are calculated (Figure S55).

680
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

681 Where, \hat{y}_i represents the predicted value of the regression model, \bar{y} is the mean of observed
 682 data, y_i is the actual value, and n is a total number of samples.



683

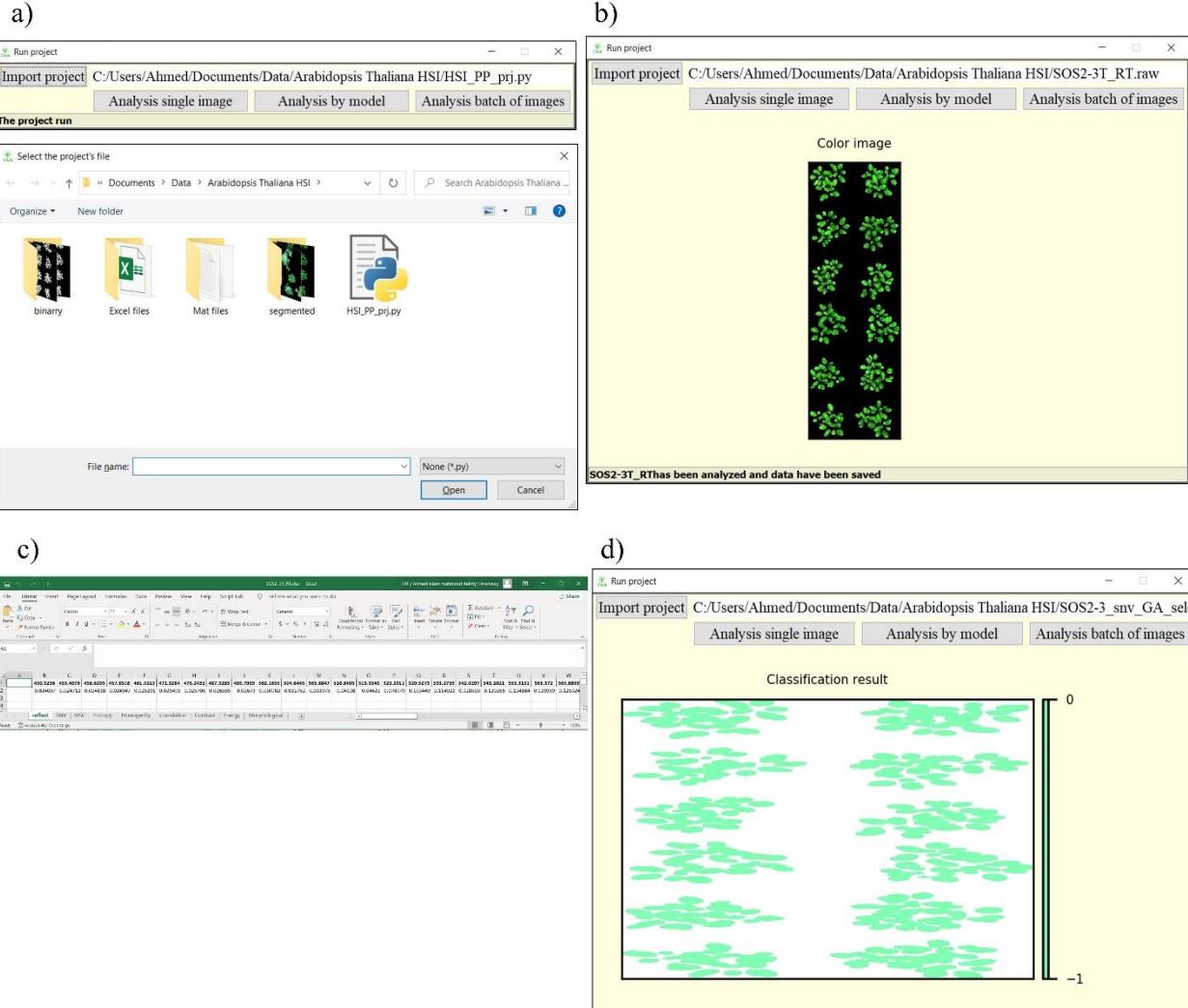
684

Figure S55. Evaluate regression model

685 5. Import project

686 At this step, users can import saved projects to extend existing analysis pipelines and apply the
 687 same parameters into new datasets, which can be applied to a single or batch of images. One can
 688 access the stored models previously developed (pre-trained models written in python) to translate
 689 spectrum values to help users for making a decision. Once the features have been extracted, they
 690 can be analyzed by a pre-trained model to translate these features into valuable information.
 691 Furthermore, users can visualize the spatial distribution of chemical contents which illustrates
 692 hidden information in every pixel of a sample. That will facilitate decision-making in precision
 693 agriculture or functional genomics under different conditions. Select the project file (.py) by
 694 clicking the “**Import project**” button in the “**Run project**” window and select direction and files
 695 (Figure S56a). Once the project imported it is ready for automatically processing and extract

696 features from one hyperspectral image “**Analysis single image**” or batch of images “**Analysis**
 697 **batch of images**” (Figure S56b) and the extracted features will be saved as excel file (Figure S56c)
 698 or analyzing this image by pre-trained model (Figure S56d).



- 705 phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling
706 for robust quantification of genotype effects. *New Phytologist*, **191**, 895-907.
- 707 **Bishop, C. M.** (1995). *Neural networks for pattern recognition*, Oxford university press.
- 708 **Breiman, L.** (2001). Random forests. *Machine learning*, **45**, 5-32.
- 709 **Cortes, C. and Vapnik, V.** (1995). Support-vector networks. *Machine learning*, **20**, 273-297.
- 710 **Cortesi, D.** (2020). PyInstaller Documentation 4.1 ed.
- 711 **De Vylder, J., Vandenbussche, F., Hu, Y., Philips, W. and Van Der Straeten, D.** (2012).
712 Rosette tracker: an open source image analysis tool for automatic quantification of genotype
713 effects. *Plant physiology*, **160**, 1149-1159.
- 714 **Freedman, D. A.** (2009). *Statistical models: theory and practice*, cambridge university press.
- 715 **Glorfeld, L. W.** (1996). A methodology for simplification and interpretation of
716 backpropagation-based neural network models. *Expert Systems with Applications*, **10**, 37-54.
- 717 **Goldberg, D. E. and Holland, J. H.** (1988). Genetic algorithms and machine learning.
718 *Machine Learning*, **3**, 95-99.
- 719 **Jansen, M., Gilmer, F., Biskup, B., Nagel, K. A., Rascher, U., Fischbach, A., Briem, S.,**
720 **Dreissen, G., Tittmann, S. and Braun, S.** (2009). Simultaneous phenotyping of leaf growth and
721 chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in
722 *Arabidopsis thaliana* and other rosette plants. *Functional Plant Biology*, **36**, 902-914.
- 723 **Kennard, R. W. and Stone, L. A.** (1969). Computer aided design of experiments.
724 *Technometrics*, **11**, 137-148.
- 725 **Martínez, A. M. and Kak, A. C.** (2001). Pca versus lda. *IEEE transactions on pattern analysis*
726 *machine intelligence*, **23**, 228-233.
- 727 **Massart, D. L., Vandeginste, B. G., Buydens, L., Lewi, P., Smeyers-Verbeke, J. and Jong, S. D.** (1998). *Handbook of chemometrics and qualimetrics*, Elsevier Science Inc.
- 728 **Meier, B. A.** (2015). *Python GUI programming cookbook*, Packt Publishing Ltd.
- 729 **Pierna, J. a. F., Abbas, O., Baeten, V. and Dardenne, P.** (2009). A Backward Variable
730 Selection method for PLS regression (BVSPLS). *Analytica chimica acta*, **642**, 89-93.
- 731 **Powers, D. M.** (2011). Evaluation: from precision, recall and F-measure to ROC, informedness,
732 markedness and correlation.
- 733 **Rinnan, Å., Van Den Berg, F. and Engelsen, S. B.** (2009). Review of the most common pre-
734 processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, **28**, 1201-
735 1222.
- 736 **Savitzky, A. and Golay, M. J.** (1964). Smoothing and differentiation of data by simplified
737 least squares procedures. *Analytical chemistry*, **36**, 1627-1639.
- 738 **Serpico, S. B. and Bruzzone, L.** (2001). A new search algorithm for feature selection in
739 hyperspectral remote sensing images. *IEEE Transactions on Geoscience Remote Sensing*, **39**,
740 1360-1367.
- 741 **Stehman, S. V.** (1997). Selecting and interpreting measures of thematic classification accuracy.
742 *Remote sensing of Environment*, **62**, 77-89.

744 **Zhou, J., Spallek, T., Faulkner, C. and Robatzek, S.** (2012). CalloseMeasurer: a novel
745 software solution to measure callose deposition and recognise spreading callose patterns. *Plant*
746 *Methods*, **8**, 49.