Opstad Kruse, OM.; Prats Montalbán, JM.; Indahi, UG.; Kvaal, K.; Ferrer Riquelme, AJ.; Futsaether, CM. (2014). Pixel classification methods for identifying and quantifying leaf surface injury from digital images. Computers and Electronics in Agriculture. 108:155-165. doi:10.1016/j.compag.2014.07.010.

Pixel classification methods for identifying and quantifying leaf surface injury from digital images

Ole Mathis Opstad Kruse[a], José Manuel Prats-Montalbán[b], Ulf Geir Indahl[a], Knut Kvaal[a], Alberto Ferrer[b], Cecilia Marie Futsaether[a *]

[a] Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway

[b] Multivariate Statistical Engineering Group, Department of Applied Statistics, Operations Research and Quality, Universidad Politécnica de Valencia, Cno. De Vera s/n, Edificio 7A, 46022, Valencia, Spain

* Corresponding author:

Cecilia Marie **Futsaether**

cecilia.futsaether@nmbu.no

Phone: +47 6496 5436

Department of Mathematical Sciences and Technology

Norwegian University of Life Sciences

P.O. Box 5003

N-1432 Ås

Norway

Ole Mathis Opstad **Kruse,** ole.kruse@nmbu.no

José Manuel **Prats-Montalbán**, jopramon@eio.upv.es

Ulf Geir **Indahl**, ulf.indahl@nmbu.no

Knut **Kvaal**, knut.kvaal@nmbu.no

Alberto **Ferrer**, aferrer@eio.upv.es

1 Abstract

Plants exposed to stress due to pollution, disease or nutrient deficiency often develop visible symptoms on leaves such as spots, colour changes and necrotic regions. Early symptom detection is important for precision agriculture, environmental monitoring using bio-indicators and quality assessment of leafy vegetables. Leaf injury is usually assessed by visual inspection, which is labour-intensive and to a considerable extent subjective. In this study, methods for classifying individual pixels as healthy or injured from images of clover leaves exposed to the air pollutant ozone were tested and compared. RGB images of the leaves were acquired under controlled conditions in a laboratory using a standard digital SLR camera. Different feature vectors were extracted from the images by including different colour and texture (spatial) information. Four approaches to classification were evaluated: (1) Fit to a Pattern Multivariate Image Analysis (FPM) combined with T^2 statistics (FPM-T^2) or (2) Residual Sum of Squares statistics (FPM-RSS), (3) linear discriminant analysis (LDA) and (4) *K*-means clustering. The predicted leaf pixel classifications were trained from and compared to manually segmented images to evaluate classification performance. The LDA classifier outperformed the three other approaches in pixel identification with significantly higher accuracy, precision, true positive rate and F-score and significantly lower false positive rate and computation time. A feature vector of single pixel colour channel intensities was sufficient for capturing the information relevant for pixel identification. Including neighbourhood pixel information in the feature vector did not improve performance, but significantly increased the computation time. The LDA classifier was robust with 95% mean accuracy, 83% mean true positive rate and 2% mean false positive rate, indicating that it has potential for real-time applications.

Key words: classification, feature extraction, Fit to a Pattern Model approach (FPM), linear discriminant analysis (LDA), *K*-means clustering, multivariate image analysis.

2 Introduction

Plants are exposed to a wide variety of stresses including disease, nutrient deficiency, drought and pollution which can affect growth, the diversity of natural vegetation and crop production. These stresses often give rise to visual symptoms on the leaf surfaces such as spots, streaks, colour changes and necrotic regions. Visible lesions may initially be difficult to distinguish from healthy leaf regions and may change in colour, shape and size as the lesion develops. Leaf injury, regardless of the cause, is usually assessed by visual inspection. This procedure relies on human experts and is time-consuming, labour-intensive and to some extent inconsistent (Bock et al., 2010). Digital image analysis has the potential for providing rapid, consistent and non-destructive leaf inspection at reasonable costs. Such systems can be used in crop management for targeted application of fungicides/pesticides/herbicides or fertilizers, quality inspection of leafy agricultural products or environmental monitoring using bio-indicators.

Ground-level ozone pollution is a global air pollution problem resulting in reduced crop yield and quality. Global crop production losses due to ozone pollution for the year 2000 are estimated to be US$11-18 billion (Avnery et al., 2011a) and are projected to reach up to US$35 billion by 2030 (Avnery et al., 2011b). Ozone exposure often results in visible leaf injuries characterized by chlorotic and necrotic spots or regions across the leaf surface and subsequent leaf senescence and abscission (Wilkinson et al., 2012). Clover (*Trifolium*), an important pasture crop, readily develops visible symptoms and has been used as an ozone bio-indicator and to develop critical levels for plant ozone effects within the UN-ECE Convention on Long-Range Transboundary Air Pollution (Karlsson et al., 2003; Karlsson et al., 2009).

Several studies have explored methods based on image processing for classifying whole leaves into categories for purposes such as plant species identification (Cope et al., 2012; Gwo et al., 2013), crop and weed discrimination (Ahmed et al., 2012; Arribas et al., 2011), determination of healthy and diseased plants (Pydipati et al., 2006; Xu et al., 2011) as well as leaf quality grading for the consumer market (Lunadei et al., 2012; Zhang and Zhang, 2011). These studies do not, however, consider classification of individual leaf pixels. Identifying leaf pixels or regions that deviate from normal leaf

78 pixels could reveal disease or injury at an early stage, characterize the severity or stage of the leaf
79 lesion and enable diagnosis of the lesion according to disease type.

80 A common approach to leaf region/pixel classification is often to segment the region of interest,
81 extract relevant features from the selected region and use these features for region classification by
82 some classifier. Huang (2007) developed a classification system to identify leaf lesions from three
83 different orchid diseases. Infected leaf areas were first segmented from the background using an
84 exponential transformation and standard image processing techniques. Texture features derived from
85 the grey level co-occurrence matrix (GLCM) and colour features were used to classify these extracted
86 areas using a back-propagation neural network classifier. Camargo and Smith (2009a) used a
87 segmentation procedure based on the distribution of the intensity histogram of colour transformed leaf
88 images followed by post-processing using morphological operations to remove pixel regions not
89 considered part of the region of interest. The procedure performed well in some cases with around
90 90% identified diseased pixels, but poorly in other cases with around 50-60% identified diseased
91 pixels. Once the diseased region was identified, texture-related features of the region were used as
92 discriminators for support vector machine classification of the type of disease (Camargo and Smith,
93 2009b). Zhang and Meng (2011) used a boosting algorithm to select significant features for
94 segmenting the leaf lesion from the background followed by extraction of zone-based local texture
95 features for classification of the lesion type and obtained classification rates similar to human experts.

96 Another more direct approach is to classify leaf pixels directly without the use of segmentation. Boese
97 et al. (2008) used an unsupervised algorithm to group leaf pixels with similar RGB values into a user
98 given number of classes. These classes were then defined by the user as healthy, diseased or injured
99 leaf areas. Accuracy, precision, recall and computation time of the method were not reported. Sanyal
100 and Patel (2008) used a feature vector that combined colour and 7×7 pixel neighbourhood information
101 extracted from leaf images and a multilayer perceptron classifier to detect two different diseases on
102 rice leaf surfaces. Although an overall pixel classification accuracy of 89% was stated, the precision,
103 recall and computation time of the method were not discussed. Bauer et al. (2011) tested two pixelwise
104 methods, k-nearest neighbour and a Gaussian mixture (GM) model, and one global probabilistic model

105 (conditional random field (CRF) model) for classifying diseased leaf pixels from stereo images of
106 sugar beet leaves infected with two types of fungi. Although the kNN classifier used long computation
107 time and did not achieve the desired classification rate levels, the GM and the CRF models were
108 promising, with the GM model giving classification rates in the range 86%-94%. The best
109 classification rates were obtained when using a feature vector that included colour as well as 4-
110 connected neighbour pixel information. Other neighbourhood relations were not tested. Both GM and
111 CRF models, however, require finding appropriate model parameters, such as the weighting functions
112 and number of and parameters of Gaussian distributions for GM models and parameters of the energy
113 function for CRF models which may be difficult to solve for exactly (Bauer et al., 2011).

114 In this study, we test and compare four approaches for classifying individual leaf pixels directly as
115 healthy or injured from clover leaf RGB images with different degrees of ozone-induced visible
116 injuries. The aim was to determine which combination of feature vector and classifier provided the
117 superior all-round classification performance. Different feature vectors were derived from the images
118 by including different colour and texture information. Three colour spaces, (1) the original RGB
119 colour space, (2) the CIE 1976 L*a*b* colour space (McLaren, 1976) and (3) the CIE 1976 uniform
120 chromaticity scale diagram (UCS) (CIE, 1986), were compared to determine which colour space was
121 best suited to capture leaf injury. Texture characteristics were included by considering two-
122 dimensional square windows of different sizes of neighbour pixel intensity values for each pixel
123 (Bharati et al., 2004; Prats-Montalbán et al., 2011).

124 Four classification approaches were compared in this study: (1) Fit to a Pattern Multivariate Image
125 Analysis (FPM) combined with T^2 statistics (FPM-T^2), (2) Fit to a Pattern Multivariate Image Analysis
126 (FPM) combined with Residual Sum of Squares statistics (FPM-RSS) (Prats-Montalbán, 2005), (3)
127 linear discriminant analysis (LDA) (Fisher, 1936) and (4) the commonly used *K*-means clustering
128 (Jain, 2010). The Fit to a Pattern MIA approach was chosen as it is a general approach for defect
129 detection in random colour textures (Prats-Montalbán, 2005) and has been used successfully in several
130 applications such as identifying diseased areas on citrus fruits (Lopez-Garcia et al., 2010), detecting
131 defects on ceramic tiles and high quality stone surfaces (Lopez et al., 2006; Prats-Montalbán and

Ferrer, 2007) as well as in metallurgic problems (Bharati et al., 2004). It was chosen as leaf ozone lesions can be regarded as defects on a leaf surface. The LDA and *K*-means techniques (Michie et al., 1994; Ripley, 1996) are standard and robust classification methods, applicable also to large datasets, as in the present study, where the training set consisted of several million pixels and the test set of several hundred thousand pixels to be classified. The LDA and *K*-means methods were therefore chosen instead of, for example, a multilayer perceptron classifier as used by Sanyal and Patel (2008) for leaf pixel classification, which due to the nonlinear dependence of the cost function on the unknown parameters, can converge slowly and nonsmoothly. In addition, the LDA and *K*-means techniques do not require considerable heuristic parameter tuning, as do other classifiers such as Gaussian mixture, multilayer perceptron or conditional random field models. The resulting leaf pixel classifications were compared to manually segmented images representing the ground-truth to evaluate classification performance.

3 Materials and methods

3.1 The image datasets

The leaf images originated from studies of leaf injuries induced by ozone in the ozone sensitive species subterranean clover (*Trifolium subterraneum* L.), (see Vollsnes et al., 2009 for experimental details). *T. subterraneum* L. (Svalöf Weibull AB, Svalöf, Sweden) seeds were germinated in trays containing sandy peat soil and transplanted individually into pots containing sandy peat soil after fourteen days. The plants were kept in an environmentally controlled growth room at 20±1°C, >60% relative humidity and with a 16 h light/ 8 h dark cycle. Ozone treatment started when the plants were 31 days old and was applied for six hours during midday (10:00 to 16:00) for three consecutive days. During ozone treatment, plants were positioned randomly in six Perspex ozone exposure chambers (l, w, h = 420×320×400 mm). Plants in three exposure chambers were exposed to charcoal-filtered air supplemented with ozone giving a 70±10 ppb ozone level for six hours per day for three days. This ozone level is realistic for short-term ozone exposure during the growing season in northern Fennoscandia (Manning et al., 2009). The plants in the other three chambers were used as controls and were exposed to six hours per day for three days of charcoal-filtered air containing very little ozone

(~1-3 ppb ozone). Apart from the exposure to ozone, ozone exposed and control plants were treated identically. Ozone was generated by passing pure oxygen through an ozone generator (model GSG 001.2, Sorbios GmbH, Berlin, Germany) and ozone concentrations inside the exposure chambers were recorded by an ozone analyzer (Photometric O3 Analyzer- Model 400, Advanced Pollution Instrumentation Inc., San Diego, CA, USA).

Trifoliate leaves of ozone exposed and control plants were imaged during the mornings and afternoons of the three days of ozone treatment and for two days following treatment. Thereafter, leaves were imaged once per day until harvest eight days after the start of ozone treatment. All leaves were assessed by plant physiologists with extensive experience with ozone-induced foliar injury. Ozone exposed leaves developed necrotic regions and spots characteristic of ozone damage whereas none of the control plants had visible ozone-induced leaf injuries. 24-bits RGB images of the leaves were taken using a Canon EOS 20D SLR camera with a 100 mm lens giving images of 3504 × 2336 pixels at a resolution of 0.021 mm/pixel. Images were taken in the RAW image format and converted to lossless TIFF images. Leaves were illuminated by two 650 W halogen lamps. The camera's shutter time was set to 1/15 s to smooth out variations caused by the AC electric power. The aperture was set to F20. The white balance was set to Auto and checked using a sheet of white paper. Leaves were gently pressed against a glass surface to ensure that they were flat during imaging. The glass surface also ensured homogeneous illumination over the leaf surface as well as homogeneous light conditions between different leaves and at different imaging times.

As illustrated in Fig. 1, a clover leaf consists of three leaflets (i.e. a trifoliate leaf). An area of 300 × 300 pixels was cropped from each leaflet. Thirty six 300 × 300 pixel images (12 representative plants selected by experts from a set of 72 images of ozone exposed leaves × 3 leaflets) with different degrees of injury (see Fig. 1 h – m) formed the first image set used to evaluate pixel classification performance of the four classification approaches. The selected leaves displayed typical ozone damage as well as typical natural background variation of different colour tones and grades. In short, the leaves were typical and representative of ozone damaged clover leaves.

185 Binary masks (Fig. 1e-g) were made for these 36 images where pixels in the image were classified
186 manually as healthy or injured. Images of these leaves taken before ozone exposure were used as
187 comparisons to avoid incorrect assignment of natural background variation as ozone injury. These
188 binary masks were assumed to represent the ground-truth and were used to assess pixel classification
189 performance.

190 A second, separate image set consisting of one hundred whole clover leaves and not only the cropped
191 leaflet areas, was used to test the applicability of the classification approaches to real leaves. Forty
192 four ozone injured and fifty six healthy, control leaf images were included in this set. As no ground-
193 truth masks were made for these ozone injured leaves, this set was used for a more qualitative
194 assessment of the classification approaches.

195 ### 3.2 Image feature extraction

196 Three colour spaces were compared to determine which colour space best captured the colour
197 information required to discriminate between injured and healthy pixels. The original RGB colour
198 space, the CIE 1976 L*a*b* colour space (McLaren, 1976) and the CIE 1976 uniform chromaticity
199 scale diagram (UCS) (CIE, 1986) were used. The CIE 1976 L*a*b* is pragmatically constructed to
200 separate the RGB colour space into visually opponent colours of Red/Green (a) and Yellow/Blue (b)
201 and the achromatic lightness (L). The L signal is the image achromatic greyscale representation, which
202 differs from the widely used (R+G+B)/3 greyscale mean representation. The (UCS) transforms the
203 RGB image into a uniform chromatic scaled image by a pragmatic linear transformation compatible
204 with the human ability to detect equidistant colours.

205 The images were unfolded in a specific manner into a 2-D feature matrix $\boldsymbol{X}$, as described in detail by
206 Bharati et al. (2004) and Prats-Montalbán et al. (2011). The unfolding includes the neighbours of
207 every pixel of interest to capture spatial information. Which neighbouring pixels to include is
208 controlled by a window that can be varied in shape and size (Prats-Montalbán, 2005). Quadratic
209 neighbourhood windows with sides w of an odd number of pixels usually in the range 3 to 49 pixels
210 are often used (Lopez-Garcia et al., 2010; Prats-Montalbán and Ferrer, 2007). In this study, window

sizes $w = 2K + 1$, where $K = 0$, 1 and 2, were used giving windows of 1×1, 3×3 and 5×5. For $K = 0$, only the pixel of interest (single pixel) was included, not the neighbouring pixels. This unfolding was done for each of the colour bands (e.g. R, G, and B in the RGB colour space) to retain colour information and the unfolded matrices of these bands were concatenated horizontally. Thus, for $K = 1$ giving a 3×3 neighbourhood window, the pixel i in row n and column m of the image is given by the row vector:

$$\begin{aligned}\boldsymbol{X}_i = & \, r_{(n-1,\, m-1)}, r_{(n,\, m-1)}, r_{(n+1,\, m-1)}, .., r_{(n+1,\, m+1)}, \\ & g_{(n-1,\, m-1)}, g_{(n,\, m-1)}, g_{(n+1,\, m-1)}, .., g_{(n+1,\, m+1)}, \\ & b_{(n-1,\, m-1)}, b_{(n,\, m-1)}, b_{(n+1,\, m-1)}, .., b_{(n+1,\, m+1)}\end{aligned} \quad (1)$$

as illustrated in Fig. 2. For an $N \times N$ colour image, the unfolding technique gives a colour-spatial feature matrix $\boldsymbol{X}$ of $(N\text{-}w\text{+}1)^2 \times (w^2 \times 3)$ where each row corresponds to the feature row vector $\boldsymbol{X}_i$ of the pixel of interest. The correction term $(-w+1)$ in the number of rows $(N\text{-}w\text{+}1)^2$ was introduced to avoid special handling of image border pixels (i.e. border pixels were not used as pixels of interest for window sizes larger than 1×1) (Prats-Montalbán and Ferrer, 2007). Thus, for the 3×3 neighbourhood window case, the feature matrix dimension for an $N \times N$ colour image is $[(N\text{-}2) \times (N\text{-}2)] \times (9 \times 3)$.

3.3 Pixel classification methods

Four classification methods were tested to evaluate their performance in identifying pixels representing injured areas on the leaf surfaces. Three of these methods were supervised, (1) Fit to a Pattern MIA approach combined with T^2 (FPM-T^2) or (2) RSS statistics (FPM-RSS), and (3) linear discriminant analysis (LDA). The fourth method was unsupervised, (4) *K*-means clustering.

For the first image set of cropped leaflet images, the classification models were validated using a "leave-one-plant-out" approach where each plant was sequentially left out of the model training and instead used to test the classification model accuracy, as illustrated in Fig. 3. Thus, the pixels belonging to one plant (e.g. pixels from the three leaflet images) were used as the test set ($\boldsymbol{X}_{test}$) (see Fig. 3). The pixels from all other 11 plants (i.e. 3×11 images) were compiled into the training matrix $\boldsymbol{X}_{train}$. Thus, when the images were unfolded without using neighbouring pixels (i.e. 1×1 window), the

training matrix $\boldsymbol{X}_{\text{train}}$ consisted of 300 × 300 × 33 = 2 970 000 pixels (i.e. samples), and the test set consisted of 300×300×3 = 270 000 pixels to be classified. The $\boldsymbol{Y}_{\text{test}}$ and $\boldsymbol{Y}_{\text{train}}$ group membership vectors for the corresponding pixels were created using the manually made masks, as illustrated in Fig. 3.

In the case of the second image set of whole leaves, the classification models were trained using all thirty six cropped leaflet images from the first image set as ground-truth masks (and hence the $\boldsymbol{Y}_{\text{train}}$ vector) were available for this set. In addition, as a comparison, ten healthy control leaves not exposed to ozone, giving an additional thirty (10 plants × 3 leaflets) 300 × 300 pixel leaflet images, were added to the training set to test whether including healthy leaves in the model training affected pixel classification. Hence, this second training set consisted of 22 individual leaves, both healthy and ozone injured, from different representative plants, giving a total of 66 leaflet images (or about six million pixels) for training. No $\boldsymbol{Y}_{\text{test}}$ group membership vectors were made for this second image set.

3.3.1 FPM-T^2

In the FPM approach, the pixels in the training set, $\boldsymbol{X}_{\text{train}}$, were used to build a Principal Component Analysis (PCA) model to which the test set $\boldsymbol{X}_{\text{test}}$ was compared (Prats-Montalbán, 2005), as depicted in Fig. 4. First, $\boldsymbol{X}_{\text{train}}$ was scaled to mean centre using the mean of the pixels that according to the ground-truth binary masks were free of defects (i.e. healthy pixels). PCA was then applied on the mean centred $\boldsymbol{X}_{\text{train}}$ matrix to compress the image information into a reduced number of uncorrelated (orthogonal) variables, called principal components (PCs) (Mardia et al., 1979). PCA projects the original variables onto new ones, called latent variables, orthogonal and arranged according to their eigenvalue, giving

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{\mathbf{T}} + \boldsymbol{E} \qquad (2)$$

where $\boldsymbol{T}$ (I×A) and $\boldsymbol{P}$ (J×A) are the scores and loading matrices for A principal components ($A \leq \text{rank}(\boldsymbol{X})$) (Næs et al., 2002), and I and J are the number of rows and columns in $\boldsymbol{X}_{\text{train}}$, respectively. The residual matrix $\boldsymbol{E}$ contains the unexplained data variability from the fitted model.

The FPM approach uses two parameters, the optimal number of principal components for the PCA model and a threshold limit that separates healthy from injured pixels. To determine the number of principal components, the reference eigenspace $\boldsymbol{P}$ was calculated stepwise from one to a maximum of ten principal components (see Fig. 4). For each of these ten models, the T^2 statistics (or *D*-statistics) of every pixel in $\boldsymbol{X}_{\text{train}}$ was calculated using

$$T_i^2 = \sum_{a=1}^{A} \frac{t_{ia}^2}{\sigma_a^2} \quad (3)$$

where t_{ia} is the score value of pixel i in principal component a, and σ_a^2 is the variance in principal component a. The T^2 value summarizes the score values for each pixel and provides a measure of the variation of each pixel inside the model. In short, T_i^2 is the Mahalanobis distance with respect to the model's mean of the pixel neighbourhood projected onto the subspace defined by the A retained principal components (López et al., 2006).

To discriminate between healthy and injured pixels, an optimal threshold T^2 level T_H^2 was found by systematically testing the classification accuracy of equally spaced values of T^2 until the highest accuracy was obtained (Fig. 4). Pixels with T^2 values larger than the threshold limit T_H^2 corresponded to pixels with extreme values and were defined as belonging to injured leaf areas. Pixels with values lower than T_H^2 fit the healthy model and were classified as healthy. The threshold limit T_H^2 and the number of PCs included in the PCA model which provided the highest overall pixel classification accuracy, were selected, giving the model reference eigenspace $\boldsymbol{P^*}$ used to classify the test set.

The mean centred $\boldsymbol{X}_{\text{test}}$ matrix was then projected onto the selected reference eigenspace $\boldsymbol{P^*}$ (see Fig. 4), enabling the score matrix $\boldsymbol{T_{test}}$, the pixel components in the space spanned by the principal components, to be computed:

$$\boldsymbol{T_{test} = X_{test} P^*} \quad (4)$$

The T^2 values of the pixels in the test set ($T_{test,i}^2$) were calculated from $\boldsymbol{T_{test}}$ using Eq. (3). Pixels with T^2 values larger than the chosen threshold limit T_H^2 corresponded to pixels with extreme values, but still maintaining the colour and spatial information correlation structure (Prats-Montalban and Ferrer,

2007), and were classified as belonging to injured pixels. The predicted classification results for all pixels in the test set were collected in $\hat{\boldsymbol{Y}}$, the predicted group membership vector, as follows:

$$\hat{Y}_i = \begin{cases} 0, & T^2_{test,i} \leq T^2_H \\ 1, & T^2_{test,i} > T^2_H \end{cases}. \quad (5)$$

3.3.2 FPM-RSS

The residual matrix $\boldsymbol{E}_{train}$ of the training set was calculated from Eq. (2) as

$$\boldsymbol{E}_{train} = \boldsymbol{X}_{train} - \boldsymbol{T}_{train}\boldsymbol{P}_j^{\boldsymbol{T}} \quad (6)$$

where $\boldsymbol{P}_j$ is the eigenspace for the PCA model with j = 1 to 10 PCs as described in the previous section (see Fig. 4). For each of these ten models, the residual matrix $\boldsymbol{E}_{\text{train}}$ was used to calculate the Residual Sum of Squares (RSS) (also known as Q-statistics) for the FPM method. For each pixel i in the training set:

$$RSS_{\text{i}} = \sum_{\text{a}=1}^{\text{A}} E_{\text{ia}}^2 \quad (7)$$

The RSS values represent the squared Euclidean distance to the subspace defined by the principal components and capture how well the pixels comply with the model (Prats-Montalbán and Ferrer, 2007). The RSS limit, RSS_H, between healthy and injured pixels was found using the same procedure used to determine the threshold limit T_H^2, by systematically testing the classification accuracy of different RSS limits (refer to Fig. 4). Pixels with *RSS*-statistic values larger than the threshold RSS_H corresponded to pixels that did not behave in the same manner as the ones used to create the model, in the sense that there is a breakage in the colour and spatial correlation structure of the model. The model $\boldsymbol{P}^*$ with the number of PCs and RSS limit, RSS_H, giving the highest accuracy was selected for classifying the test set $\boldsymbol{X}_{test}$. The residual matrix $\boldsymbol{E}_{\text{test}}$ for the test set was calculated using Eqs. 4 and 6 as $\boldsymbol{E}_{test} = \boldsymbol{X}_{test} - \boldsymbol{T}_{test}\boldsymbol{P}^{*\boldsymbol{T}}$, from which the RSS-statistics of the pixels in the test set ($RSS_{test,i}$) were calculated using Eq. (7) (see Fig. 4). Pixels with *RSS*-values larger than the threshold RSS_H had extreme values and were classified as injured. The predicted group membership vector $\hat{\boldsymbol{Y}}$ of the pixels in the test set was given as follows:

$$\hat{Y}_i = \begin{cases} 0, & RSS_{test,i} \leq RSS_H \\ 1, & RSS_{test,i} > RSS_H \end{cases}. \quad (8)$$

3.3.3 *K*-means clustering

As shown in Fig. 1, injured pixels were more similar to each other than to uninjured pixels. It may therefore be possible to identify injured and uninjured pixels by clustering the pixels that were most similar. The widely *K*-means clustering algorithm splits the dataset into *K* homogeneous clusters, is fast, usually requiring only a few iterations to converge and can thus be used on large datasets. The number of clusters is an input parameter to the algorithm which assigns the pixels in the image to the *K* predefined clusters by minimizing the sum, over all classes, of the within-clusters sums of pixel-to-clusters-centroid distances (Ripley, 1996).

The *K*-means clustering algorithm was run directly on the $\boldsymbol{X}_{\text{train}}$ matrix starting at a randomly selected data point. The *K*-means algorithm was run for different numbers (2 – 10) of clusters to determine the optimal number required to capture the injured pixels in the image. After each run, the clusters were compared to the manually made binary masks to identify which clusters corresponded most to injured pixels and to estimate the classification accuracy. The run providing the best overall classification result was used to determine the number of required clusters and which clusters corresponded to injured pixels. The centroids $\boldsymbol{\mu}_k$ for these *K* clusters, where $k \in \boldsymbol{K}$ {1,…, *K*} were then used to calculate the Euclidean distance to each pixel $\boldsymbol{x}_i$ in the $\boldsymbol{X}_{\text{test}}$ matrix. The pixels were assigned to the cluster with the closest centroid (Euclidian distance), enabling classification of the pixels $\boldsymbol{x}_i$ in the test set and giving the predicted group membership vector $\hat{\boldsymbol{Y}}$ as follows:

$$\hat{Y}_i = argmin_{k \in \boldsymbol{K}} \|\boldsymbol{x}_i - \boldsymbol{\mu}_k\|^2. \quad (9)$$

3.3.4 Linear discriminant analysis (LDA)

Fisher's Linear Discriminant Analysis (Fisher, 1936; Mardia et al., 1979; Rao, 1948) is a statistical classification method developed by assuming the involved groups to share a joint multinormal covariance structure in the feature space. The robustness of LDA, even when the model assumptions

are violated to some extent, was documented in an extensive comparison study by Michie et al. (1994) of 22 different classification methods, where LDA (referred to as "DISCRIM") was ranked top three among the classification methods most frequently performing best on a large number of different classification problems. With a given number of G groups or classes (G = 2 classes, injured or healthy pixels, in the present study), LDA implicitly identifies a G-1 dimensional subspace of the feature space accounting for all the information relevant for group separation. Robustness for the present pixel classification application using LDA is mainly assured by the large number of available pixels (i.e. 300 × 300 pixels per leaflet image) compared to the small number of features (maximum 75, 3 colour channels × no. pixels in neighbourhood window) characterizing each pixel.

An LDA model was built based on the $\boldsymbol{X}_{\text{train}}$ matrix and the manually created masks $\boldsymbol{Y}_{\text{train}}$ corresponding to these pixels. From the model building we obtained the two class centres, $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$, and the *within groups* covariance matrix $\boldsymbol{\Sigma}$. In LDA, classification of the pixels $\boldsymbol{x}_{\text{i}}$ in the test set $\boldsymbol{X}_{\text{test}}$, yields the predicted group membership vector $\hat{\boldsymbol{Y}}$ by a simple rule comparing Mahalanobis distances between the observed pixel and the class centres:

$$\hat{Y}_i = argmin_{g \in \{0,1\}} (\boldsymbol{x}_i - \boldsymbol{\mu_g})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu_g}). \qquad (10)$$

Contrary to the other classifiers, LDA does not require heuristic parameter tuning/optimization such as selection of model complexity (like the number of latent variables in PCA or PLS models), threshold limit selection (as for the FPM methods) or selection of the number of clusters (as for the *K*-means method).

3.4 Defect maps

By folding the predicted group membership vectors $\hat{\boldsymbol{Y}}$ (Eqs. 5, 8, 9, 10) back into binary 2-D matrices, we obtain the so-called defect maps. Defect maps are appropriate for visualization and evaluation of the pixels classified as injured or healthy by comparison to both the corresponding original leaflet images (Fig. 1) and corresponding binary masks (subjective ground-truth).

3.5 Performance measures

Comparison between the predicted group membership vector $\hat{\boldsymbol{Y}}$ for each classification (Eqs. 5, 8, 9, 10) and the ground-truth given in $\boldsymbol{Y}_{\text{test}}$ enabled calculation of the following:

True positives, *tp*, representing the number of injured pixels correctly classified as injured,

True negatives, *tn*, representing the number of healthy pixels correctly classified as healthy,

False positives, *fp*, representing the number of healthy pixels misclassified as injured,

False negatives, *fn*, representing the number of injured pixels misclassified as healthy.

These calculations were used to determine the following five performance measures:

$$true\ positive\ rate\ (Recall) = \frac{tp}{tp+fn}\ ,\ \left(\frac{positives\ (injured)\ correctly\ classified}{total\ positives\ (injured)}\right)$$

$$false\ positive\ rate = \frac{fp}{fp+tn}\ ,\ \left(\frac{negatives\ (healthy)\ incorrectly\ classified}{total\ negatives\ (healthy)}\right)$$

$$Precision = \frac{tp}{tp+fp}\ ,\ \left(\frac{positives\ (injured)\ correctly\ classified}{number\ classified\ as\ positives\ (injured)}\right) \quad (11)$$

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}\ ,\ (fraction\ correctly\ classified)$$

$$F\text{-}score = 2\ \times \frac{Precision \times Recall}{Precision+Recall}$$

These are measures of how well the different methods perform in classifying injured pixels from the healthy background. Because neither Precision nor Recall is a good stand-alone parameter for performance we also included the (associated) F-score (Lopez-Garcia et al., 2010).

The computation timing results required by each combination of feature vector and classification method were obtained using a single threaded code in a standard computer (HP EliteBook 8460p, 8 GB RAM, Intel i5, 2.5 GHz dual core processor) and Microsoft Windows 7 (Service Pack 1) operating system. The computation time for feature extraction and pixel classification by one classifier for all 12 plants (all 36 cropped leaflet images of the first image set) was recorded, and divided by 12 to give the calculation time required for one plant (i.e. three leaflet images).

3.6 Statistical Analysis

A coincidence matrix with the number of true positives (*tp*), false positive (*fp*), false negative (*fn*) and true negative (*tn*) was obtained for each classification. The fractions of each of these values (sum = 1) were used for statistical analyses in a two-step procedure to quantify classifier comparisons. After verifying the normality of the percentages, ANOVA (Montgomery, 2005) analyses with a significance level of 5% were conducted on the performance data of the four classifiers (LDA, FPM-T^2, FPM-RSS and *K*-means clustering) enabling between classifier comparisons. The performance within each classifier was evaluated by analysing the effects of the feature matrices based on the different colour spaces and pixel neighbourhoods on the classification results. Since all data were derived from a Design of Experiments, interactions were also analysed. Response variables studied were the five performance measures (Eq. 11).

3.7 Calculations

All calculations were performed in MATLAB® (v. 8.01, R2013a, The MathWorks, Inc., Natick, MA, USA) in combination with the MATLAB compatible PLS-Toolbox® (v. 7.0.3, Eigenvector Research Inc., Wenatchee, WA, USA).

Statistical analyses were conducted in the MATLAB Statistics Toolbox v. 8.2.

4 Results and discussion

The results listed in Table 1 show that the LDA method significantly outperformed the three other classification methods in pixel identification with significantly higher accuracy, precision, true positive rate and F-score as well as significantly lower false positive rate and computation time. *K*-means performed significantly better than the two Fit to a Pattern MIA approaches, which had similar performances, apart from a significantly higher precision for the FPM-T^2 compared to the FPM-RSS. The RSS statistic generally captures more noise thereby hindering clear detection of defective areas. Given the iterative nature of the *K*-means algorithm, the *K*-means approach was the slowest of the methods.

As can be seen in Table 2, the choice of feature vector did not significantly affect the classification performance of any of the four methods in our study, implying that changing the colour space or increasing the number of neighbours included in the feature vector was not critical. Thus, it was sufficient to build the classifiers based on 3-dimensional pixel feature vectors in the original untransformed RGB colour space. Not including the neighbourhood pixel information means ignorance of the spatial (texture) information, leaving us to conclude that the information discriminating between healthy and injured pixels was primarily present in the pixel colours rather than in the size, shape or texture of the lesions. The large variation in the size of the injured areas from small spots, a few pixels in size, to large continuous areas larger than the 5×5 neighbourhood window may be the reason why inclusion of neighbourhood information was ineffective. In other applications, larger window sizes closer to the defect size may provide improved performance of the classifier(s) (López et al., 2006; Prats-Montalbán, 2005; Prats-Montalbán and Ferrer, 2007). Not requiring neighbourhood information means less complexity and corresponding computational savings. We experienced that by expanding the neighbourhood window size from 1×1 (single pixel) to 3×3 and 5×5 pixels, the computation time increased by 1.7 & 3.5 times, respectively, for the LDA classifier, 7 & 590 times, respectively, for the FPM classifiers, and 7 & 20 times, respectively, for the *K*-means classifier.

The significantly higher true positive rate of LDA indicates that this method correctly identified more of the injured pixels than the others (Table 1). A true positive rate of about 80% indicates, however, that 20% of the injured pixels were falsely identified as healthy (false negatives). The true positive rates of the three other classifiers were at the level 50-60%, indicating that they failed to identify a considerably larger fraction of the injured pixels (40-50%). LDA also provided a 2% false positive rate, thereby misclassifying very few healthy pixels as injured and correctly identifying the majority (98%) of healthy pixels. In comparison, the FPM approaches misclassified about 10% of the healthy pixels as injured whereas the *K*-means method had a false positive rate similar to LDA (Table 1). Consequently, it is fair to say that the LDA and *K*-means classifiers outperformed the FPM methods with respect to precision. It should be noted that most of the leaves had a low fraction of injured pixels

(27 of the 36 leaves had less than 20 % injured pixels in the manually created masks). Therefore, the number of misclassified pixels was low relative to the total number of pixels in the images. This results in the high overall accuracy of all methods (≥85%) and the lower precision and F-score (Table 1).

Figure 5 gives an overview of the classifier accuracy on each leaflet image. The accuracy of the LDA method appears as fairly stable, with only one leaflet image where the pixel classification accuracy was less than 85% (Fig. 5a, plant nr. 6, leaflet nr. 1). The performance of the other methods, particularly of the FPM approaches, was more variable, with high accuracy for some images and low for others.

Figure 6 gives an overview of the defect maps of selected leaflets made by comparing the manually created masks (subjective ground-truth) and the predicted pixel classification, colour coded to show false positives in red and false negatives in blue. As depicted by the large, blue coloured areas, the *K*-means and FPM classifiers failed to identify a large fraction of the injured pixels (false negatives). A conspicuous trend, particularly for the LDA classifier, was that the misclassified healthy pixels (false positive, red in Fig. 6) were mainly located along the borders between injured and healthy areas. This effect is especially notable in the middle column of Fig. 6, which corresponds to the leaflet classified with less than 85% accuracy by the LDA classifier (see Fig. 5a, plant nr. 6, leaflet nr. 1). These borders were difficult to set manually since the visible signs of injury faded towards healthy areas. It is therefore possible that the ground-truth mask did not correctly include the complete set of injured pixels. Hence, a considerable amount of the apparently false positives may be subject to manual reclassification after a possible second revision of the image, resulting in an adjustment of the precision rates in the classifications.

Figure 7 shows the application of the four classification approaches to seven representative whole clover leaves. Based on the findings that the choice of feature vector was not critical (Table 2), a three dimensional feature vector using the original untransformed RGB colour space and no neighbourhood pixel information was used. For the LDA and two FPM methods, the pixel classification obtained for

all 100 whole leaves was identical for both training sets, with the same pixels identified as injured and healthy. Thus, including ten healthy leaves in the model training in addition to the 36 ozone injured cropped leaflet images of the first image set did not alter the pixel classification. For the *K*-means method, only a slight difference of 0.002% was obtained in pixel classification when using the two training sets. Thus, the 36 original leaflets of the first image set contained the necessary pixels representative of both the injured and healthy classes. The FPM-RSS approach failed to identify healthy pixels, classifying over 92% of the pixels as injured, even for healthy control leaves where 99±1% (mean ± standard deviation, n=56 leaves) of the pixels were misclassified as injured (Fig. 7, top three rows). The FPM-T^2 method performed reasonably well on ozone injured leaves (Fig. 7, bottom four rows), identifying injured pixels, but with a tendency of assigning natural variation such as leaf veins as injury. The FPM-T^2 approach had difficult with healthy control leaves, misclassifying 60±30% (mean ± standard deviation, n=56 leaves) of the pixels as injured. Thus, despite good performance in other applications, the FPM approaches were not optimal for this application. As seen in Fig. 7 (top three rows), LDA and especially the *K*-means method performed well on healthy control leaves, respectively misclassifying 3±2% and 0.2±0.2% (mean ± standard deviation, n=56 leaves) of the pixels as injured. Closer inspection of the leaves revealed that these misclassifications were caused by light olive coloured spots and areas and leaf hairs which appeared as light coloured streaks in the image. As ozone damaged areas usually are light brown instead of deep green, the colour of these natural features resembled ozone damage more closely than healthy regions, and were hence misclassified. This effect was particularly evident for the LDA results of the control leaf of the third row of Fig. 7. The LDA and *K*-means methods captured the ozone injury patterns of the ozone damaged leaves (Fig. 7, bottom four rows) with LDA predicting 5-10% more injury than the *K*-means classifier for all 44 ozone exposed leaves. Closer inspection of the leaves revealed that the *K*-means classifier did not identify some injured areas, thereby underestimating injury, as also seen in Fig. 6 (blue, false negatives). Both methods correctly identify natural variations such as leaf veins and dark patches as healthy pixels.

5 Conclusion

Different feature vectors and classification methods were compared to determine a robust and accurate approach for pixelwise identification of leaf surface injury from RGB images acquired using a standard digital SLR camera. Four classifiers and feature vectors including different colour and spatial information were evaluated.

The LDA approach provided the best overall pixelwise classification performance with significantly higher accuracy, precision, recall and F-score than the other approaches. The LDA approach provided a high mean accuracy of 95% and was robust with relatively little variation in classification performance when tested on leaves with different degrees of injury and natural background variation. The LDA approach also performed well on whole leaves, capturing the injury patterns in damaged leaves and handling the natural surface variations in control leaves. The LDA approach was also the most computationally efficient and simplest method to implement requiring no heuristic tuning of model parameters. By inspection of the Mahalanobis distances calculated by LDA, it is also straight forward to design heuristic diagnostics for outlier detection, and thereby "weed out" future pixel vectors not sufficiently represented in the training data used for model building.

The simplest feature vector requiring only colour information proved to be sufficient for leaf pixel classification. Inclusion of spatial data did not improve classification performance, but increased the computation time significantly. Transporting the MATLAB code to a faster programming language such as C, optimizing the code and using multithread code will reduce the timing costs considerably such that an LDA classification program could be included in real-time applications in, for example, precision agriculture and quality inspection.

In conclusion, we have seen that the LDA classifier combined with a colour feature vector can provide rapid and accurate pixelwise identification of injury from images of leaf surfaces.

Acknowledgements

We thank Associate Professor Aud Berglen Eriksen and Dr. Ane V. Vollsnes at the Dept. of Biology, University of Oslo, for the collaboration on the ozone exposure experiments.

We would also like to thank former Master student Hilde Landrø at the Norwegian University of Life Sciences, for her help with the ozone exposure experiments and for taking the images used in this study.

References

Ahmed, F., Al-Mamun, H.A., Bari, A., Hossain, E., Kwan, P., 2012. Classification of crops and weeds from digital images: A support vector machine approach. Crop Protection 40, 98-104.

Arribas, J.I., Sanchez-Ferrero, G.V., Ruiz-Ruiz, G., Gomez-Gil, J., 2011. Leaf classification in sunflower crops by computer vision and neural networks. Comput. Electron. Agric. 78, 9-18.

Avnery, S., Mauzerall, D.L., Liu, J.F., Horowitz, L.W., 2011a. Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage. Atmos. Environ. 45, 2284-2296.

Avnery, S., Mauzerall, D.L., Liu, J.F., Horowitz, L.W., 2011b. Global crop yield reductions due to surface ozone exposure: 2. Year 2030 potential crop production losses and economic damage under two scenarios of O-3 pollution. Atmos. Environ. 45, 2297-2309.

Bauer, S.D., Korc, F., Forstner, W., 2011. The potential of automatic methods of classification to identify leaf diseases from multispectral images. Precision Agriculture 12, 361-377.

Bharati, M.H., Liu, J.J., MacGregor, J.F., 2004. Image texture analysis: methods and comparisons. Chemometrics Intell. Lab. Syst. 72, 57-71.

Bock, C.H., Poole, G.H., Parker, P.E., Gottwald, T.R., 2010. Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. Critical Reviews in Plant Sciences 29, 59-107.

Boese, B.L., Clinton, P.J., Dennis, D., Golden, R.C., Kim, B., 2008. Digital image analysis of Zostera marina leaf injury. Aquatic Botany 88, 87-90.

Camargo, A., Smith, J.S., 2009a. An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosyst. Eng. 102, 9-21.

Camargo, A., Smith, J.S., 2009b. Image pattern classification for the identification of disease causing agents in plants. Comput. Electron. Agric. 66, 121-125.

539 CIE, 1986. Colorimetry, second edtition, Technical Report CIE 15.2-1986, Second ed.
540 Cope, J.S., Corney, D., Clark, J.Y., Remagnino, P., Wilkin, P., 2012. Plant species identification using
541 digital morphometrics: A review. Expert Syst. Appl. 39, 7562-7573.
542 Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenic 7, 179-
543 188.
544 Gwo, C.Y., Wei, C.H., Li, Y., 2013. Rotary matching of edge features for leaf recognition. Comput.
545 Electron. Agric. 91, 124-134.
546 Huang, K.Y., 2007. Application of artificial neural network for detecting Phalaenopsis seedling
547 diseases using color and texture features. Comput. Electron. Agric. 57, 3-11.
548 Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognition Letters 31, 651-666.
549 Karlsson, G.P., Karlsson, P.E., Danielsson, H., Pleijel, H., 2003. Clover as a tool for bioindication of
550 phytotoxic ozone - 5 years of experience from southern Sweden - consequences for the short-term
551 critical levels. Science of the Total Environment 301, 205-213.
552 Karlsson, P.E., Pleijel, H., Danielsson, H., Karlsson, G.P., Piikki, K., Uddling, J., 2009. Evidence for
553 Impacts of Near-ambient Ozone Concentrations on Vegetation in Southern Sweden. Ambio 38, 425-
554 431.
555 Lopez-Garcia, F., Andreu-Garcia, G., Blasco, J., Aleixos, N., Valiente, J.M., 2010. Automatic
556 detection of skin defects in citrus fruits using a multivariate image analysis approach. Comput.
557 Electron. Agric. 71, 189-197.
558 López, F., Prats, J.M., Ferrer, A., Valiente, J.M., 2006. Defect detection in random colour textures
559 using the MIA T-2 defect maps, In: Campilho, A., Kamel, M. (Eds.), Image Analysis and Recognition,
560 Pt 2. Springer-Verlag Berlin, Berlin, pp. 752-763.
561 Lunadei, L., Diezma, B., Lleo, L., Ruiz-Garcia, L., Cantalapiedra, S., Ruiz-Altisent, M., 2012.
562 Monitoring of fresh-cut spinach leaves through a multispectral vision system. Postharvest Biology and
563 Technology 63, 74-84.
564 Manninen, S., Huttunen, S., Tømmervik, H., Hole, L.R., Solberg S., 2009. Northern plants and ozone.
565 Ambio 38, 406-412.
566 Mardia, K.V., Kent, J.K., Bibby, J.M., 1979. Multivariate Analysis. Academic Press, London.

McLaren, K., 1976. The Development of CIE 1976 (L*a*b*) Uniform Colour Space and Colour-difference Formula. J. Soc. Dye. Colour 92, 338-341.

Michie, D., Spiegelhalter, D.J., Taylor, C.C., 1994. Machine Learning, Neural and Statistical Classification. Ellis Horwood Ltd, Chichester.

Montgomery, D.C., 2005. Design and Analysis of Experiments. 6th ed. John Wiley & Sons, Inc., Tempe, Arizona.

Næs, T., Isaksson, T., Fearn, T., Davies, T., 2002. A User-Friendly Guide to Multivariate Calibration and Classification. NIR Publications, Chichester, UK.

Prats-Montalbán, J.M., 2005. Control Estadísitico de Procesos mediante Análisis Multivariante de Imágenes. Tesis Doctoral. Universidad Politécnica de Valencia.

Prats-Montalbán, J.M., de Juan, A., Ferrer, A., 2011. Multivariate image analysis: A review with applications. Chemometrics Intell. Lab. Syst. 107, 1-23.

Prats-Montalbán, J.M., Ferrer, A., 2007. Integration of colour and textural information in multivariate image analysis: defect detection and classification issues. J. Chemometr. 21, 10-23.

Pydipati, R., Burks, T.F., Lee, W.S., 2006. Identification of citrus disease using color texture features and discriminant analysis. Comput. Electron. Agric. 52, 49-59.

Rao, C.R., 1948. The Utilization of Multiple Measurements in Problems of Biological Classification. Journal of the Royal Statistical Society Series B-Statistical Methodology 10, 159-203.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.

Sanyal, P., Patel, S.C., 2008. Pattern recognition method to detect two diseases in rice plants. Imaging Science Journal 56, 319-325.

Vollsnes, A.V., Eriksen, A.B., Otterholt, E., Kvaal, K., Oxaal, U., Futsaether, C.M., 2009. Visible foliar injury and infrared imaging show that daylength affects short-term recovery after ozone stress in Trifolium subterraneum. Journal of Experimental Botany 60, 3677-3686.

Wilkinson, S., Mills, G., Illidge, R., Davies, W.J., 2012. How is ozone pollution reducing our food supply? Journal of Experimental Botany 63, 527-536.

Xu, G.L., Zhang, F.L., Shah, S.G., Ye, Y.Q., Mao, H.P., 2011. Use of leaf color images to identify nitrogen and potassium deficient tomatoes. Pattern Recognition Letters 32, 1584-1590.

595 Zhang, F., Zhang, X.H., 2011. Classification and Quality Evaluation of Tobacco Leaves Based on
596 Image Processing and Fuzzy Comprehensive Evaluation. Sensors 11, 2369-2384.
597 Zhang, M., Meng, Q.G., 2011. Automatic citrus canker detection from leaf images captured in field.
598 Pattern Recognition Letters 32, 2036-2046.

599

600

Table captions

Table 1. The pixel classification performance of the four classifiers LDA, *K*-means, FPM-T^2 and FPM-RSS. A feature vector with three features corresponding to the *r*, *g* and *b*-values from the RGB colour space for a single pixel was used (i.e. $\boldsymbol{X}_i = r_{(n,m)}, g_{(n,m),} b_{(n,m).}$ for an arbitrary pixel *i* in row *n* and column *m* of the image, neighbourhood window size 1 × 1). Mean values ± standard deviation for *n* = 36 cropped leaflet images of the first image set are shown. Each timing corresponds to feature extraction and pixel classification for one plant (i.e. three leaflet images, mean ± standard deviation for three runs).

Method	Accuracy	Precision	True positive rate (recall)	False positive rate	F-score	Timing (s)
LDA	0.95±0.05	0.8±0.2	0.83±0.12	0.02±0.02	0.75±0.15	3.03±0.09
K-means	0.93±0.05	0.8±0.2	0.5±0.3	0.03±0.05	0.6±0.3	470±50
FPM-T^2	0.86±0.11	0.5±0.3	0.6±0.3	0.11±0.12	0.5±0.3	23.6±1.1
FPM-RSS	0.85±0.12	0.2±0.2	0.5±0.5	0.12±0.12	0.2±0.2	23.96±1.1

610 **Table 2.** The pixel classification accuracy of the four classifiers LDA, *K*-means, FPM-T^2 and FPM-
611 RSS using feature vectors constructed using different neighbourhood window sizes and image colour
612 spaces. Mean values ± standard deviation for $n = 36$ cropped leaflet images of the first image set are
613 shown.

Neighbourhood window size	Colour space	LDA	*K*-means	FPM-T^2	FPM-RSS
1×1 window	RGB	0.95 ±0.05	0.93 ±0.05	0.86 ±0.11	0.85 ±0.12
	L*a*b	0.95 ±0.05	0.92 ±0.06	0.85 ±0.10	0.88 ±0.08
	USC	0.95 ±0.04	0.93 ±0.05	0.86 ±0.10	0.87 ±0.09
3×3 window	RGB	0.95 ±0.05	0.93 ±0.05	0.85 ±0.12	0.88 ±0.09
	L*a*b*	0.95 ±0.05	0.92 ±0.06	0.85 ±0.10	0.89 ±0.07
	USC	0.95 ±0.04	0.93 ±0.05	0.86 ±0.11	0.87 ±0.10
5×5 window	RGB	0.95 ±0.04	0.93 ±0.05	0.85 ±0.12	0.89 ±0.09
	L*a*b*	0.95 ±0.04	0.91 ±0.06	0.84 ±0.11	0.89 ±0.07
	USC	0.95 ±0.04	0.93 ±0.05	0.85 ±0.12	0.87 ±0.09

614

615

Figure captions

Figure 1. Original leaf (a) where the quadratic 300 × 300 pixel areas used in the analysis of this image are shown. The corresponding leaflet images are shown in (b – d), and the manually created binary masks are shown in (e – g). Images cropped from ozone-exposed leaflets where the degree of ozone-induced injury increases from slight (h) to extensive (m) are shown.

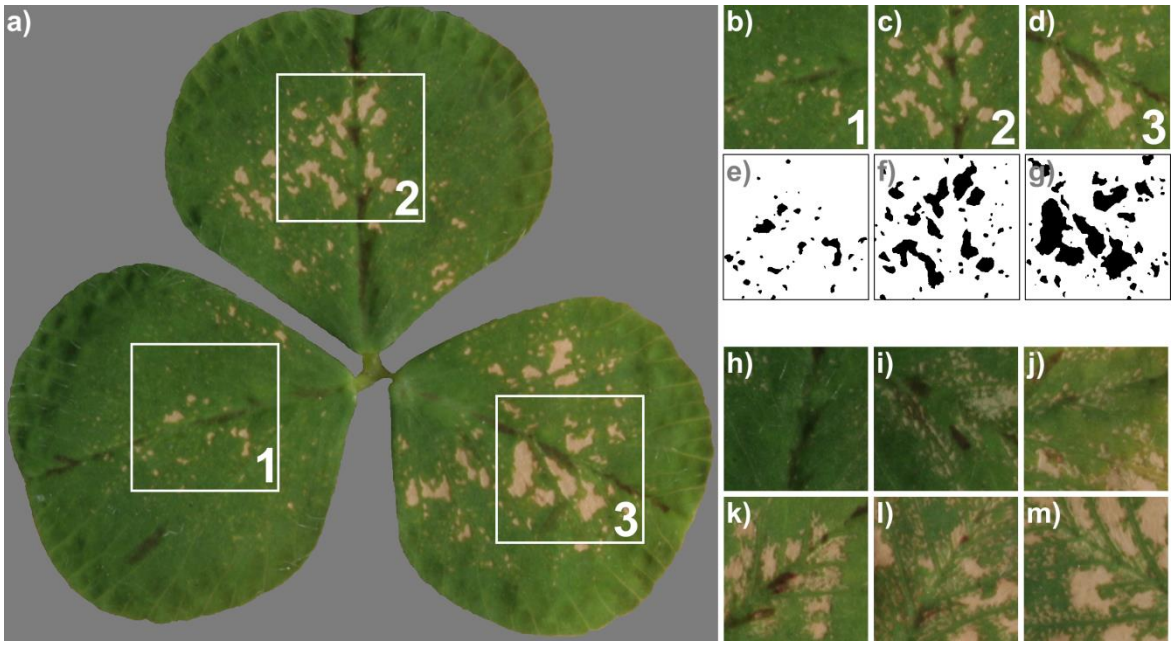


Figure 2. a) A 3×3 neighbourhood window where the centre pixel is the pixel *i* of interest (dark grey) and the surrounding pixels are the eight neighbours (light grey). The arrows show the direction of the unfolding starting with the left column. b) Diagram of the resulting colour-spatial feature matrix ***X*** given by Eq. 1, for an *N*×*N* RGB image, where each row represents the feature vector for a given pixel *i*.

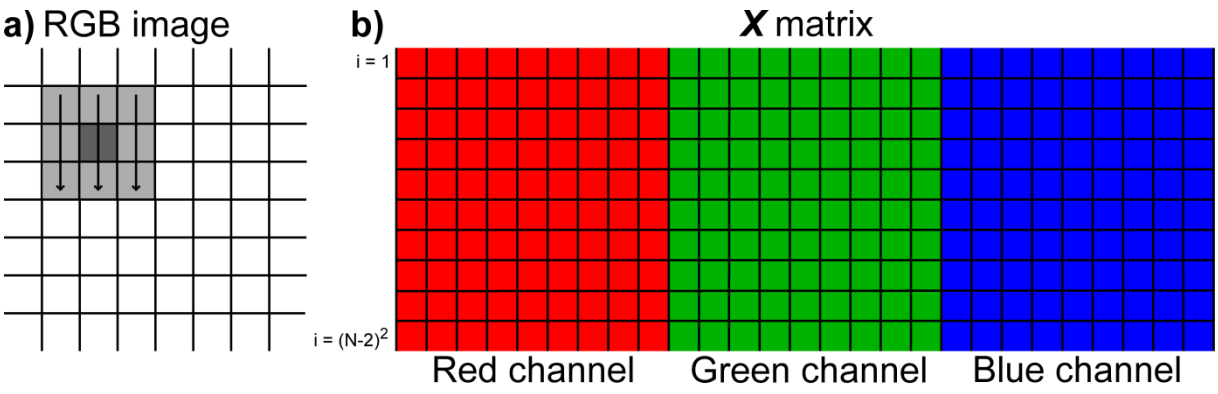


Figure 3. Overview of the classification procedure showing the colour space transformation and unfolding of the input leaflet images to generate the feature training $\boldsymbol{X}_{\text{train}}$ and test $\boldsymbol{X}_{\text{test}}$ matrices and the unfolding of the corresponding manually segmented binary masks to generate the $\boldsymbol{Y}_{\text{train}}$ and $\boldsymbol{Y}_{\text{test}}$ group membership vectors. The predicted group membership vector $\hat{\boldsymbol{Y}}$ was compared to the ground-truth given in $\boldsymbol{Y}_{\text{test}}$ to assess the performance of the classifier for pixel classification.

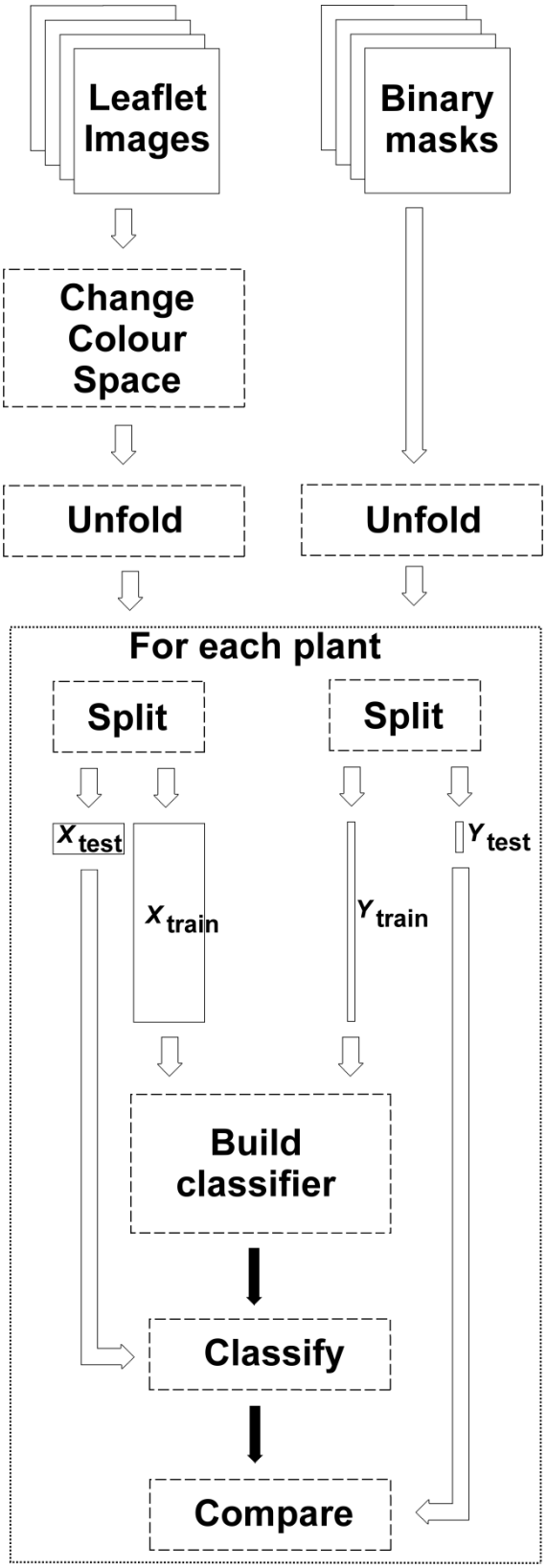
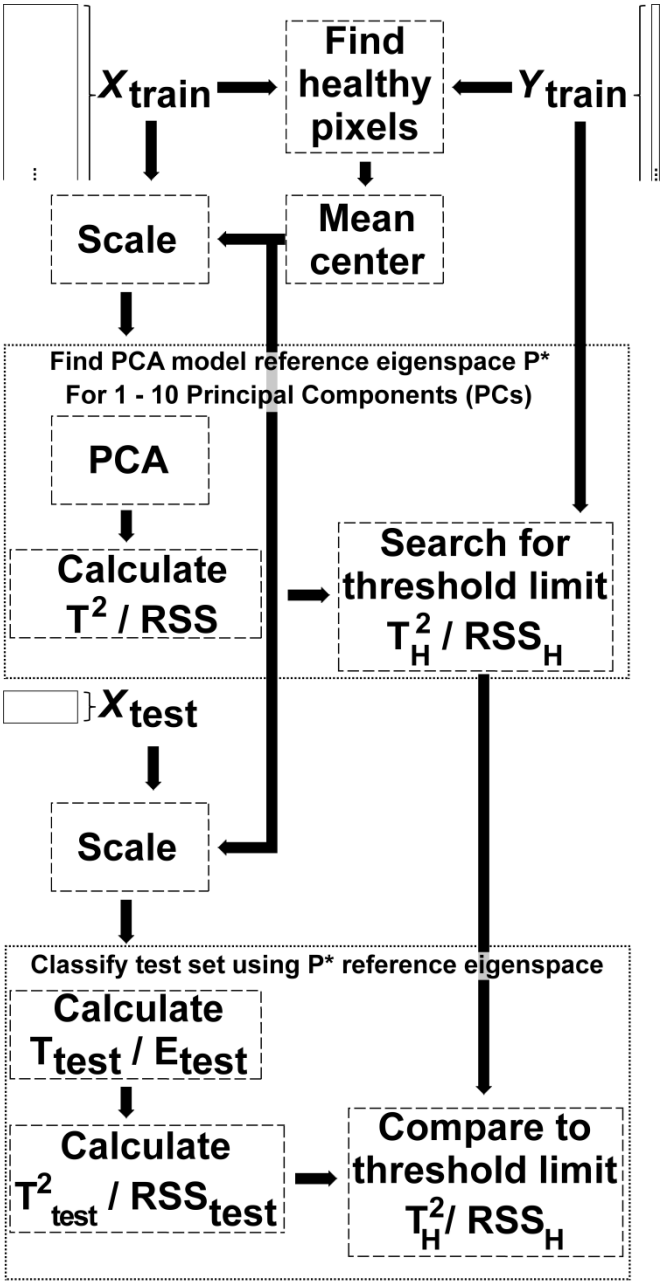
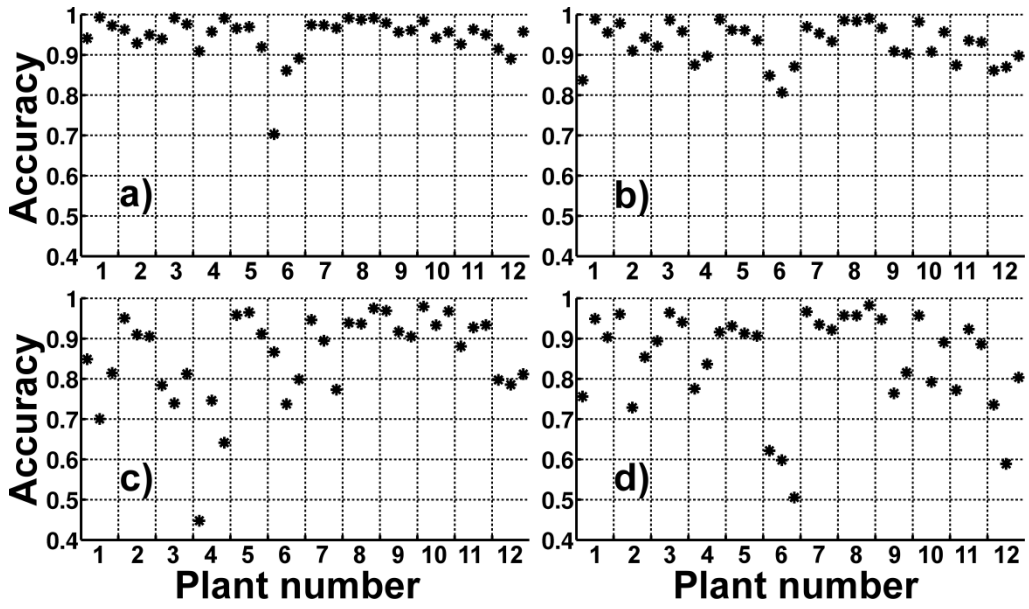


Figure 4. Flow chart of the Fit to a Pattern Multivariate Image Analysis classification approach. The feature matrix $\boldsymbol{X}_{train}$ used for training was mean centred using the mean of the pixels defined as healthy. A PCA model was built and used to classify the pixels in the test set $\boldsymbol{X}_{test}$, giving the predicted group membership vector $\hat{\boldsymbol{Y}}$.



644 **Figure 5**. Pixel classification accuracy of the classifiers (a) LDA, (b) *K*-means, (c) FPM-T^2 and (d)
645 FPM-RSS for all 36 cropped leaflet images. For a perfect classifier, the accuracy (*) for each leaflet
646 would be equal to 1. A feature vector with three features corresponding to the *r*, *g* and *b*-intensity
647 values of the RGB colour space for the pixel of interest was used (no neighbourhood pixel information
648 included). The numbering 1 to 12 on the abscissa corresponds to the examined 12 clover leaves
649 collected from 12 different plants. Each clover leaf of a given plant (numbered 1 to 12) consisted of
650 three individual leaflets giving a total of 3 × 12 images.

651



652

Figure 6. Original RGB images (top row) for selected leaflets and corresponding defect maps (subsequent rows) for the four classifiers LDA, *K*-means, FPM-T^2 and FPM-RSS calculated using a feature vector with three features corresponding to the *r*, *g* and *b*- intensity values of the RGB colour space for the pixel of interest (no neighbourhood pixel information included). Each column represents one leaflet image, numbered as plant number and leaflet number, for correspondence with the abscissa in Fig. 5 (i.e. Pl. 1 - 1 corresponds to plant nr. 1, leaflet nr. 1). Leaflets Pl. 4 - 1 and Pl. 11 - 2 were slightly injured, Pl. 1 - 1 moderately injured and Pl. 6 - 1 and Pl. 6 - 2 extensively injured. In the defect maps (rows two to five), white areas correspond to correctly classified injured pixels (true positives), black areas correspond to correctly classified healthy pixels (true negatives), red areas show healthy pixels wrongly classified as injured (false positives) and blue areas show injured pixels not detected by the classifier (false negatives).

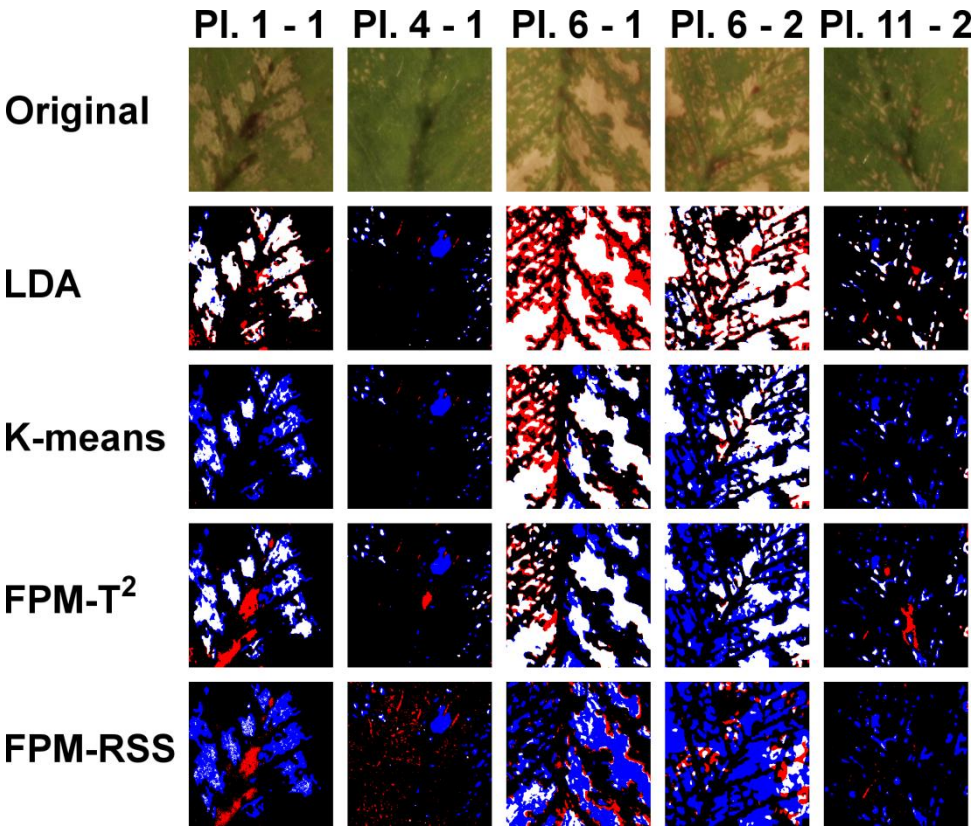


Figure 7. Pixel classification for seven representative whole clover leaves using the four classifiers LDA, *K*-means, FPM-T^2 and FPM-RSS calculated using a feature vector with three features corresponding to the *r*, *g* and *b*- intensity values of the RGB colour space for the pixel of interest (no neighbourhood pixel information included). Each row gives the prediction results for one clover leaf. The original RGB images are shown in the first column and the predictions of the classifiers are shown in the subsequent columns, with pixels predicted as healthy in black and injured pixels in grey. The first three rows show healthy control leaves and rows four to seven show leaves with increasing ozone injury.