Automatic Assesment of Phenotypes in lettuce plants by using Chlorophyl Fluorescence Kinetics and Machine Learning

Xanthoula Eirini Pantazi, Dimitrios Moshou, Dimitrios Kateris, Dimitrios Kasampalis and Pavlos Tsouvaltis, Aristotle University of Thessaloniki, School of Agriculture, Department of Hydraulics, Soil Science and Agriculture Engineering, Laboratory of Agricultural Engineering and Department of Horticulture, P.0 275 54 124, Thessaloniki, Greece.

Abstract

Agriculture aims at increasing production and provision of high quality products to the market. Most of the times, quality is strongly correlated the variety or the hybrid. Specifically, lettuce qualitative characteristics and nutrients appear to vary strongly in different varieties and hybrids. In the current research, lettuce plants were harvested at baby, immature and mature stage in 46, 60 and 70 days of growth, respectively. Then, the parameters of chlorophyll fluorescence were determined in two middle leaves of 3 plants of each hybrid at different harvesting stages by using chlorophyll fluorescence kinetics. The measurements revealed significant differences between varieties and hybrids. The fluorescence parameters were utilized as inputs for training different models of different novelty detection methods aiming at the identification of the phenotype of different varieties and hybrids. It is already known that novelty detection can be easily combined with machine learning techniques so as to detect abnormal events. This system was capable of diagnosing a new fault that did not appear in the training data set. For change detection, a normality description (baseline condition) was constructed. As a result, deviations were detected from this description of the normal domain (as new varieties and hybrids). In this paper, an active learning method based on novelty detection in the form of one-class classifiers for the iterative detection of different phenotypes of lettuce plants based on fluorescence sensing is proposed. This method learns to distinguish between different hybrids of lettuce plants based on their fluorescence parameter differences. The proposed active learning method uses one-class classification to detect phenotypic deviations between hybrids as outliers which are then augmented in a baseline multi-class classifier which is subsequently ready for novelty detection in case new phenotypes appear. Then, the procedure continues for the next hybrid and the proposed scheme can learn and augment new phenotypes indefinitely. It was shown that the identification of different lettuce phenotypes corresponding to hybrids is precise by the proposed Active Learning Method due to non-linearity problem which is due to the heterogeneity of the fluorescence kinetics parameters.

Keywords: horticulture, lettuce plants, novelty detection, postharvest quality, phenotypes

1 Introduction

An important issue for vegetable production concerns their harvesting stage which is highly associated with their nutrient content. The content of certain nutrients increases with age, due to senescence. In the same paper, it is concluded that the same trend appeared in the
content of potassium, magnesium, manganese, iron, and zinc, which decreased progressively during the seven-day harvest period. Currently, there is no objective method to characterize the senescence of lettuce plants and therefore there is no existent standard according to which the growth stage can be characterized except of the size of the lettuce head. But between different hybrids the lettuce head size can vary so it is not reliable criterion regarding the determination of the growth stage. For this reason a more objective criterion is sought regarding the determination of the growth stage as related to the level of senescence. A frequently applied technique to determine lettuce maturity concerns optical remote sensing either spectroscopic or by using fluorescence. Chlorophyll fluorescence has been routinely used for many years to monitor the photosynthetic performance of plants non-invasively. Possible specific applications of chlorophyll fluorescence include the screening of plants for tolerance to environmental stresses and for improvements in glasshouse production and post-harvest handling of crop. It is already known that when a dark-adapted leaf is exposed to light, large changes in chlorophyll fluorescence occur.

The rapid changes in fluorescence that occur during the rapid induction to a peak have long been attractive for detecting differences in photosynthetic performance between plants that are due phenotypic differences due to maturation. The light energy absorbed by plants is converted into chemical energy (photosynthesis), heat and fluorescence. The necessity of active learning in the case of phenotype sensing emanates from the limitations of real-time crop monitoring due to the sole reliance on one crop measurement to indicate maturity level. The design presented here is oriented towards a real-time maturity stage phenotyping management system. The proposed system is based on One Class Classifiers that can provide identification of maturity level that is present using as input previously unseen fluorescence features. The resulting Active Learning classifier is capable of learning progressively so as to recognize different maturity levels in lettuce plants by adding new classes of unforeseen phenotypic stages based on One Class Classification against the previously accumulated knowledge. This is achieved by using in an iterative way, One Class Classifiers with recognized class augmentation.

2 Materials and methods

One of the main objectives in agriculture is to identify ways in which chlorophyll fluorescence may be used effectively to improve plant selection processes and rapidly evaluate plant performance in agricultural and horticultural crop improvement programmes. Specifically, in the case of lettuce qualitative characteristics and nutrients appear to vary strongly in different development stages. In the current research, lettuce plants belonging to the hybrids Mastamar, Atoll and Starfighter of Batavia type as well as hybrids of Bacio, Picos CLX and Picos FM of Romana type (Fig 1.a) were cultivated in a heated greenhouse constructed from glass during the period of 15/10-27/12/2012. In 46, 60 and 70 days of growth, the plants were harvested at baby, immature and mature stage.

![Figure 1. (a) Hybrids Batavia type (Atoll, Mastamar and Starfighter) and Romana Type (Bacio, Picos CLX, Picos FM). (b) FluorPen FP 100-MAX-LM.](image)

Then, the parameters of chlorophyll fluorescence were determined in two middle leaves of 3 plants of each hybrid at different harvest stage by using chlorophyll fluorescence kinetics. The measurements revealed significant differences between harvesting stages by utilizing
FluorPen FP 100-MAX-LM of SCI (Fig. 1. (b)) which is capable of measuring chlorophyll fluorescence kinetics through the OJIP method which concerns the fluorescence transient (Strasser et al, 2000).

The fluorescence parameters were utilized as inputs for training different models of supervised Self Organizing Maps (SOMs) aiming at the prediction of harvesting stage. Fluorescence kinetics have been already used for the detection of mealiness in apples (Moshou et al, 2003) by using self-organizing maps to discriminate between different mealiness severity levels. Multisensor fusion of fluorescence kinetics and hyperspectral imaging has been applied for the detection of plant diseases (Moshou et al., 2012). Self-Organizing Maps have been used for the monitoring and classification of pea varieties (Pisum sativum) according to their degree of resistance against drought stress (Maldonado-Rodriguez et al., 2003).

2.1 Fluorescence parameters

Many fluorescence parameters have already been proposed. The rapid changes in fluorescence that occur during the induction of photosynthesis when a dark-adapted leaf is exposed to light, have long been attractive for detecting differences in photosynthetic performance between plants (Fig. 2).

![Figure 2. Kautsky Curves resulting from the fluorometer](image)

With this instrument, the fluorescence is excited by ultra-bright light emitting diodes (LED) with a peak wavelength of 650 nm. Chlorophyll fluorescent signals were detected using a photocell after passing through a high-pass filter (50% transmission at 720 nm). The recording time during the experiments was 1 s with a resolution of 10 μs during the first 2 μs and after that with a resolution of 1 μs, resulting in 1200 values per measurement. The lettuce plants were not dark-adapted and were held under normal artificial lighting before measuring. The induction curves are shown in Fig. 2.

2.2 Data analysis

2.2.1 Fluorescence parameters of the FP100

The fluorometer calculates automatically certain geometric parameters of the Kautsky curves. The obtained parameters are shown in Table 1. The fluorescence data were normalized by dividing by $F_0$ (or $F_{50\text{ms}}$), the fluorescence value measured after 50 ms. This value corresponds to the initial fluorescence, $F_0$ and is the fluorescence intensity when the electron acceptors are in their oxidized form. By normalization, differences between the measurements due to factors, which are not characteristic of the internal properties of the sample as for example fluctuations in excitation light intensity caused by the diminishing power of the battery, and variations in extent of the excited surface due to variations in the curvature of the surface and in homogeneities in the tissue (e.g. lenticels), were minimized. Fluorescence was measured with high time resolution, resulting in a large number of data points for each measurement. Therefore the use of the automatically calculated parameters is preferable due to their physiological significance (Strasser et al., 2000).
<table>
<thead>
<tr>
<th>Formula abbreviation</th>
<th>Formula explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Bckg</strong></td>
<td>Background</td>
</tr>
<tr>
<td><strong>2. F₀</strong></td>
<td>$F_{50\mu s}$, fluorescence intensity at 50 $\mu$s</td>
</tr>
<tr>
<td><strong>3. F_j</strong></td>
<td>fluorescence intensity at J-step (at 2 $\mu$s)</td>
</tr>
<tr>
<td><strong>4. F_i</strong></td>
<td>fluorescence intensity at i-step (at 60 $\mu$s)</td>
</tr>
<tr>
<td><strong>5. F_M</strong></td>
<td>maximal fluorescence intensity</td>
</tr>
<tr>
<td><strong>6. F_V</strong></td>
<td>$F_M - F_0$ (maximal variable fluorescence)</td>
</tr>
<tr>
<td><strong>7. V_j</strong></td>
<td>$(F_j - F_0) / (F_M - F_0)$</td>
</tr>
<tr>
<td><strong>8. V_i</strong></td>
<td>$(F_i - F_0) / (F_M - F_0)$</td>
</tr>
<tr>
<td><strong>9. F_M / F_0</strong></td>
<td></td>
</tr>
<tr>
<td><strong>10. F_V / F_0</strong></td>
<td></td>
</tr>
<tr>
<td><strong>11. F_V / F_M</strong></td>
<td></td>
</tr>
<tr>
<td><strong>12. M₀ or (dV/dt)₀</strong></td>
<td>$TR_0 / RC - ET_0 / RC = 4 (F_{300} - F_0) / (F_M - F_0)$</td>
</tr>
<tr>
<td><strong>13. Area</strong></td>
<td>Area between fluorescence curve and $F_M$ (background subtracted)</td>
</tr>
<tr>
<td><strong>14. Fix Area</strong></td>
<td>Area below the fluorescence curve between $F_{40\mu s}$ and $F_{1s}$ (background subtracted)</td>
</tr>
<tr>
<td><strong>15. S_m</strong></td>
<td>Area / $(F_M - F_0)$ (multiple turn-over)</td>
</tr>
<tr>
<td><strong>16. Sₘₚ</strong></td>
<td>the smallest SM turn-over (single turn-over)</td>
</tr>
<tr>
<td><strong>17. N</strong></td>
<td>$S_M$, $M₀$, $(1 / V_j)$ turn-over number $Q_A$</td>
</tr>
</tbody>
</table>
2.3 One Class Classification

One-class classification utilizes data that belongs to the target class while does not consider outliers in the calibration of the one class classifier. The limit demarcating the border between the target class and any data not belonging to it, has to be calculated from data stemming from the target class only; The main task concerns the definition of a boundary surrounding the target class (to classify as many as possible of the target examples correctly, while simultaneously minimizing the prospect of accepting outlier examples). Fig 3 depicts a target domain $X_T$ where there are two errors which are defined as $E_I$ related to falsely rejected target examples and $E_{II}$ associated to falsely accepted outlier objects. The circular area symbolizes the target domain of the selected one class classifier. Utilizing an even outlier distribution also leads to the assumption that when $E_{II}$ becomes smaller, the data description that has minimal volume is obtained., $E_I$ combined with the volume of the description can be minimized to get a good data description instead of minimizing both $E_I$ and $E_{II}$.

![Figure 3. Domains of target dataset and one-class classifier.](image)

2.4 One Class SOM

A one class SOM (OCSOM) is calibrated using normal operation data. Subsequently, the feature vector that corresponds to a new measurement is examined in order to assess its similarity to the weight
vectors of every other map unit. If the smallest distance exceeds a predetermined threshold, it is assumed that the process belongs to a novel situation. This result emanates from the assumption that quantization errors exceeding a certain value are associated with the operation points that are external to the region that has been covered by the training data. Hence the situation is novel and raising the possibility of novelty detection. Depending on the magnitude of deviation from the normal operation state, a degradation index can be calculated. The one-class SOM (OCSOM) constructs a model from healthy plants signatures and subsequently classifies new data according to its deviation from the healthy training data. During novelty recognition, novel examples from plant signatures of not definable health state are used to formulate the input to the network while the SOM algorithm selects the best matching unit. In Saunders and Gero (2001) if the quantization error that results from the comparison between the new exemplar data ($x_{NEW}^{i}$) and best matching unit (bmu) is larger than a pre-specified threshold ($d$) then the example is considered as novel. Eq. 1 represents the minimum distance for bmu and against the threshold.

$$
\min_{i \in M} \left( \sum_{j=0}^{M-1} (x_{NEW}^{i} - m_{j})^2 \right) > d
$$

Where M denotes the SOM grid of neurons and $m_{j}$ denotes the bmu.

There are various heuristics to determine a threshold based on the usefulness of the threshold and the specific structure of the data set. A simple way to define a threshold ($d$) depends on the similarity between the SOM centroid vectors and target training vectors that have selected them as best matching units which determines the quantization error. These distances have to be estimated according to Eq. 2:

$$
distances = \min_{i \in M} \left( \sum_{k=0}^{N-1} (x_{TARGET}^{k} - m_{i})^2 \right)
$$

The threshold is estimated by using the steps which are presented below:

1. Sort distances calculated according to eq.3 in ascending order
2. Define the fraction of targets not allowed to be misclassified as outliers (e.g. 95%)
3. Calculate the target set vector distances that will be retained by applying the fraction of step 2
4. Calculate the threshold as the mean value of the largest distance retained in step 3 and the next larger one

### 2.5 Active Learning

The detailed proposed active learning scheme had the following steps:

1) The initial training set was based on feature vectors consisting of 27 features (fluorescence parameters) originating from different phenotypic states related to maturity level of lettuce plants. From each of the lettuce maturity stages a total number of 210 vectors were used that corresponded to 60 baby lettuce plants, to 60 immature lettuce plants and 90 mature lettuce plants. The trained one-class SOM was then tested with equal numbers of test vectors from all of three stages. The success criterion was the ability to be able to classify an unknown maturity stage to be an outlier when in comparison with the baby lettuce plants. Baby lettuce plants were considered to be the baseline set.

2) The initial target set was subsequently augmented with outlier values that came from the newly detected outlier maturity stage. The resulting set was then considered to be the new baseline set. The same procedure of outlier detection was then repeated but this time the criterion of success was the ability to be able to classify any new maturity stage to be an outlier in comparison with the just augmented baseline set which includes the majority of baseline stages but also the just incorporated maturity stages (which are baby, immature and mature).

3) In the case a new sample is belonging to a class which was already included as part of the target set, then the detection of outliers continues execution internally per class in the baseline set and then the sample is being classified in one of existent subclasses belonging to the baseline set.

4) Steps 2 and 3 were repeated, more specific, detection of outliers and augmentation were executed for not yet known data which could belong to an already existing maturity stage category or it might concern some new unclassified phenotypes. In the presented application the repetition of steps was terminated following the creation of 3 classes that corresponded to baby, immature and mature lettuce plant phenotypes.
It has to be noted that the total procedure followed from steps 1 to 4 does not require any external intervention but is based solely on detection of outliers and subsequent augmentation steps which are performed automatically.

3 Results and Discussion

The Active Learning procedure described in previous section was implemented by using OCSOM. The number of samples was 60, 60 and 90 per class which meant that a total of 210 samples were available for implementing the Active Learning procedure. The OCSOM was tested with different sizes between 5x5 and 30x30 units with a step of 5 per dimension. The best results were obtained from 15x15 OCSOM while larger networks gave the same results with 15x15, so this can be considered the best result with a less complex network. The arrangement of units was rectangular. The training was implemented in a batch mode. The threshold was set at 5% which meant that the tolerance to outliers was set at 0.05 in the novelty detection algorithm. The results for OCSOM are shown in the Table 2.

Table 2. The results of Active Learning using OCSOM augmentation

<table>
<thead>
<tr>
<th>Real</th>
<th>Estimated by augmented OCSOM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baby lettuce plant</td>
</tr>
<tr>
<td>Baby lettuce plant</td>
<td>95</td>
</tr>
<tr>
<td>Immature lettuce plant</td>
<td>0</td>
</tr>
<tr>
<td>Mature lettuce plant</td>
<td>0</td>
</tr>
</tbody>
</table>

The fluorescence features are able to provide a high performance of discrimination between different phenotypic stages due to the different behavior of certain fluorescence parameters that differ significantly between baby, immature and mature lettuce plants. More specifically the initial slope of the fluorescence curve (parameter 25 in Table 1) should be higher in baby lettuce plants making it easier to differentiate the corresponding class while the maximum yield of primary photochemistry (parameter 18 in Table 1) should be higher in mature lettuce plants.

4 Conclusions

In this paper, an active learning method based on novelty detection in the form of one-class classifiers for the iterative detection of different phenotypes of lettuce plants based on fluorescence sensing is proposed. This method learns to distinguish between different hybrids of lettuce plants based on their fluorescence parameter differences. The fluorescence features are able to provide a high performance of discrimination between different phenotypic stages due to the different behavior of certain fluorescence parameters that differ significantly between baby, immature and mature lettuce plants. The performance of discrimination reaches 95% which can be characterized as very promising percentage for future work on different specialty crops and quality based sorting in industrial applications and traceability at different stages of the production chain.

5 References

Moshou D., Gravalos I., Kateris D. Pantazi X.E. (2012). Water stress detection based on optical multisensor fusion with a least squares support vector machine classifier. Oral and Pro-