Computers and Electronics in Agriculture 151 (2018) 311–318

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Automated grapevine cultivar classification based on machine learning using leaf morpho-colorimetry, fractal dimension and near-infrared spectroscopy parameters

S. Fuentes[a,*], E. Hernández-Montes[b], J.M. Escalona[b], J. Bota[b], C. Gonzalez Viejo[a], C. Poblete-Echeverría[c], E. Tongson[a], H. Medrano[b]

[a] Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville 3010, Victoria, Australia
[b] Institute for Agro-environment and Water Use Efficiency Research (INAGEA) [University of Balearic Islands (UIB)-National Institute of Agricultural Research (INIA)-Balaeric Islands Government (CAIB)], Research Group in Plant Biology under Mediterranean Conditions, 07012 Palma de Mallorca, Spain
[c] Department of Viticulture and Oenology, Faculty of AgriSciences, Stellenbosch University, Matieland 7602, South Africa

ARTICLE INFO

Keywords:
Leaf morpho-colorimetry
Computer vision
Artificial neural networks
Pattern recognition

ABSTRACT

The application of computer vision algorithms and chemometric fingerprinting using near-infrared spectrometry (NIR) of plant leaves, offers enhanced capabilities for ampelography by providing more accurate methods to discriminate leaves based on morphological parameters, and chemometrics, respectively. This paper showed that machine learning algorithms based on morpho-colorimetric parameters and NIR analysis separately, were able to automatically classify leaves of 16 grapevine cultivars. The artificial neural network (ANN) model developed with morpho-colorimetric parameters as inputs (Model 1), and 16 cultivars as targets, rendered an accuracy of 94% to classify leaves for all cultivars studied. The ANN model obtained with the NIR spectra per leaf as inputs (Model 2), and the real classification as targets, rendered 92% accuracy. The automatic extraction of morpho-colorimetric data, NIR chemical fingerprinting and machine learning modelling rendered rapid, accurate and non-destructive methods for cultivar classification, which can aid management practices.

1. Introduction

Grapevine leaves of different cultivars vary in chemical composition and morphology such as shape, dimension, color and edge shape. These differences in morphometric characteristics have been acquired as evolutionary traits corresponding to specific gene expressions and their interaction with the environment to which each cultivar has been adapted to (Nicotra et al., 2011; Vlad et al., 2014). Therefore, every leaf morphology and chemical parameter is unique for all cultivars, which allows their classification through a series of different observations or measurements for identification purposes (Chitwood et al., 2014). Besides the typical morphometric measurements such as area, petiole size, perimeter, eccentricity and edge shape, the fractal dimension has been considered as a robust classification parameter. In a study on Sangiovese grapevine genotypes (2001b), the fractal dimension and the box-counting method provided a more objective approach in the classification of these genotypes as it can easily capture the complexity of the leaf structure. This remedy the misclassifications commonly occurring in the conventional methods due to the existence of heterogeneous leaves, homonyms and synonyms.

Ampelography is the characterization and classification of grapevine cultivars using either color or shape of leaves, or their fruits or berries, with early origins in France (Rendu, 1857). The first systematic attempt of classification using ampelography methods was proposed after the second world war (Galet, 1968). In the early days of this technique, experts, whom were familiar with all the morphological characteristics of leaves and berries from different cultivars, were able to identify and name them; however, this method is considered very subjective and inaccurate (Backhaus et al., 2010). Nowadays, there have been a few improvements in the classification of grapevine cultivars using more objective methods, mainly through chemical, spectroradiometry, and genetic fingerprinting techniques (García-Muñoz et al., 2012). However, these techniques require high level of specialized skills, specific and expensive instrumentation, and tedious laboratory work, which renders their generalized practical use difficult and inaccessible for routine grapevine cultivar classification. With the advancement in digital photography and image analysis algorithms, it is possible to obtain semi-automatically a series of morphological parameters that have been already applied for grapevine leaves (Alessandri et al., 1996; Bodor et al., 2013; Fourie, 2012) and from Arabidopsis

* Corresponding author.
E-mail address: sfuentes@unimelb.edu.au (S. Fuentes).

https://doi.org/10.1016/j.compag.2018.06.035
Received 17 February 2018; Received in revised form 3 April 2018; Accepted 13 June 2018

leaves (Backhaus et al., 2010). Nevertheless, these semi- and automatic methods have not yet applied color parameters in their algorithms as part of their output variables as other authors such as Orru et al. (2012) have included for grapevine seeds classification. Furthermore, none of these techniques offered an automated classification, rather they are only descriptive, showing different cultivars or clones clustered and separated using multivariate statistical methods such as principal component analysis (PCA) and cluster analysis.

Feature extraction from objects using image analysis algorithms or morphometrics, have been widely used for different plant species based on different parameters used for morphometric analysis (Backhaus et al., 2010) such as leaf shape, elliptic Fourier descriptors, contour signatures, landmark and linear signatures, shape features, polygon fitting and fractal dimensions, venation extraction and analysis, leaf margin analysis and leaf texture analysis (Cope et al., 2012). Furthermore, color extraction from leaves has been used for different applications such as disease identification in grapevine leaves (Meunkaewjinda et al., 2008). Studies based on leaf features and color extraction (morpho-colorimetrics) can be descriptive through multivariate data analysis assessing the main features that separate different leaves either by shape, contour or color descriptors. However, there is an increasing interest in the separation of leaves automatically using machine learning algorithms either for identification, classification, or for detection of biotic and abiotic stresses (Backes et al., 2012; Bruno et al., 2008; Meunkaewjinda et al., 2008; Pandolfi et al., 2009a; Rossatto et al., 2011; Wu et al., 2007).

Leaves can also be analyzed using reflectance of light spectra from the leaf surface and detected using near-infrared spectroscopy (NIR). The NIR technique has been used as a chemical fingerprint to characterize leaves, and to obtain proxy data that can be robustly correlated to different compounds and nutrients (De Bei et al., 2017; Susan et al., 1998), and water status (De Bei et al., 2011; Santos and Kaye, 2009).

This paper describes the automated extraction of morpho-colorimetric and fractal dimension (FD) features from scanned mature leaves of 16 different grapevine cultivars by using image analysis and computer vision algorithms through a customized code written in Matlab® ver. R2017b (Mathworks Inc., Natick, MA. USA). NIR was also used to obtain the chemical fingerprinting of the leaves to compare the accuracy of classification based on both the morpho-colorimetric data and the chemical fingerprinting. Principal component analysis (PCA) was executed to visualize the classification of the leaves based on the parameters measured. Finally, pattern recognition models using artificial neural networks (ANN) were developed to automatically classify each cultivar using morpho-colorimetry and chemical fingerprinting.

2. Materials and methods

2.1. Site and cultivars description for the cultivar classification

This study was carried out in 2014, in an experimental vineyard located in Palma de Mallorca (39°35′N, 2°39′E) (Balearic Islands, Spain). In this region, the climate is Mediterranean with hot and dry summers and precipitations during autumn and winter. The soil physical and chemical properties from the trial site presents a loamy texture with alkaline pH due to the high concentration of active limestone and carbonates, which is typically found in soils from Mallorca.

All the plant samples were 10-year old vines grafted in 99-Richter rootstock. The plants were spaced at 2.5 × 1.0 m and trained on bilateral Royat Cordon system. All vines were uniformly pruned to 12 nodes per vine. A total of 16 different grapevine cultivars were used, from which three mature leaves from each of three different plants for each cultivar (Table 1) were selected, scanned and processed to obtain all the automated measurements. Leaf samples were collected from the field, stored in plastic bags and transported in a cooler with ice to avoid dehydration. Fully expanded and mature leaves were collected consistently from the fifth position from each shoot (counting from the tip) to uniform physiological maturity of leaves for modelling purposes.

Table 1
Grapevine cultivar names from the cultivar classification trial in Spain and abbreviations used for the multivariate data analysis.

Number	Cultivar	Abbreviation
1	Chenin blanc	Che
2	Chardonnay	Chy
3	Cabernet Sauvignon	CS
4	Escursac	Esc
5	Grenache	Gar
6	Gewurztraminer	Ge
7	Macabeo	Mac
8	Malvasia	Mal
9	Mando	Man
10	Merlot	Mer
11	Monastrell	Mon
12	Parellada	Par
13	Riesling	Rie
14	Sauvignon Blanc	SB
15	Shiraz	Shz
16	Tempranillo	Tem

2.2. Morpho-colorimetric analysis

Leaves were scanned using a Hewlett Packard Scanjet G3010 (Hewlett-Packard Software Company, Palo Alto, CA. USA) scanner (Fig. 1). All images were analyzed using a customized code written in Matlab® ver. R2017b (Mathworks Inc., Natick, MA. USA). Initial calibration for the scanner was performed by analyzing images with black squares as references of known dimensions to relate pixel count in the x and y coordinates to dimensions and area using metric units (cm and cm^2). As described by Fuentes et al (2012), this code (after scanner calibration) was able to automatically analyze leaf images to extract morphometric, color parameters, and fractal dimension of each sample. All the parameters obtained from the customized code are described in Table 2.

2.3. Fractal dimension analysis

The complexity of leaf shapes based on the irregular or fragmented pattern of shapes can be described using fractal analysis to characterize complex biological structures (Borkowski, 1999; Bruno et al., 2008). The fractal dimension (FD) from scanned leaves was obtained using the box-counting method (Foroutan-pour et al., 1999; Liebovitch and Toth, 1989). The FD is calculated from the morpho-colorimetric analysis previously described once the original greyscale image of single leaves are automatically identified and cropped to create binary images (Fig. 1-right; Fig. 2b). The edge detection algorithm to obtain morphological features creates an image containing only the edge of the leaves (Fig. 2a and c). Then the edge is divided into boxes of fixed length (d) and a number of boxes containing part of an edge [N(d)] (Fig. 2d). The log[N(d)] versus log(d) can then be computed and plotted to obtain an averaged value of FD corresponding to the most linear portion of the curve.

2.4. Near-infrared spectrometry

Same leaves from all cultivars used for morpho-colorimetric assessment were measured using a HR2000+ high-resolution spectrometer connected with an optic fiber QP600-2-VIS/NIR BX (Ocean optics, Dunedin, FL, USA). This device was able to measure the absorbance of the components which can be detected at a wavelength range of 200–1050 nm (n = 2049). All leaves were measured at five different spots, always starting from the top right lobe down and finishing in the top left lobe from each leaf, in order to get an average of their absorbance values per leaf (Fig. 1). Averaged values per leaf were