Distinction of “Sangiovese” clones and grapevine varieties using Elliptic Fourier Analysis (EFA), neural networks and fractal analysis

S. Mancuso, M. Boselli, E. Masi
Dipartimento di Ortoflorofrutticoltura, Università degli Studi di Firenze, Viale delle Idee, 30, 50019 Sesto Fiorentino (FI), Italy.

Key words: back propagation neural network, Elliptic Fourier analysis, fractal geometry, grapevine identification.

Abstract: Parameters coming from Elliptic Fourier Analysis (EFA) and fractal analysis of the leaf were used as input in a backpropagation neural network to distinguish between ten accessions of grapevine (eight clones of ‘Sangiovese’ and two varieties, ‘Abrostine’ and ‘Pugnitello’). The results enabled us to identify all the accessions studied. Cluster analysis of the backpropagation outputs revealed the existence of a uniform group for the ‘Tebano’ clones supporting the hypothesis that ‘SG-12-T’ and ‘SG-14-T’ could be mutations originated from the same genotype. The use of neural networks coupled to new tools in the shape analysis of leaves opens interesting prospects for ampelography.

1. Introduction

Morphological leaf characters and quantitative measurements of anatomical elements of the leaf, i.e. angles, area, teeth number, petiole length, have been extensively utilised in ampelographic research (OIV-IBPGR-UPOV charts, 1983; Galet, 1985). However, the origin of the grapevine varieties, their heterogeneity and the frequent cases of homonimy and synonymy, often resulted in doubtful classification. It is thus important to define good shape measures that can be effectively applied to leaf shapes, so they can be compared and analysed by meaningful and objective criteria. One approach that researchers have proposed for describing biological shapes is the fractal based measure of digitally acquired images.

Unlike many other shape analysis methods, EFA (Elliptic Fourier Analysis) can describe complex shapes, does not require mathematically determined centroid, does not require points on the outline to be equally spaced, and can include simple normalizations for size, position, orientation, and starting position of the trace. A further property of Fourier methods is the ability to invert the transformation and reconstruct an outline from a set of Fourier coefficients. Hence, for example, an “average” shape can be reconstructed from the mean coefficients of a large number of outlines (Ferson et al., 1985). Recently, because of the many excellent features, EFA has been utilised in several scientific fields, from botany to ampelography (Diaz et al., 1991; Mancuso, 1999 a).

Fractal geometry based analysis has received increasing attention as a number of studies have shown fractal based measures to be useful for characterising complex biological structures in human anatomy (Caldwell et al., 1990; Glenny and Robertson, 1991; Glenny et al., 1991) or in many fields of engineering and science (Logan and Wilkinson, 1990; Moghaddam, 1991; Avnir et al., 1992; Cox and Wang, 1993; Anderson et al., 1996; Smith et al., 1996). Fractal scaling (Fig. 1) is evident in natural objects...
from the micro-scale to the macro-scale, e.g. the human body contains many structures with fractal characteristics. Since investigators have found that non-fractal objects were the exception, rather than the rule in many natural systems (Avnir et al., 1992), fractal geometry has been utilised to characterise many complex biological structures of the human anatomy or also other natural objects such as leaves (Mancuso, 1999 b).

Artificial Neural Networks (ANN) are powerful computational tools that “learn” with training examples and have the capability for extrapolating their “knowledge” to new situations in problems of classification, modelling, mapping and association types. Neural network models have received increased attention in the recent years. Aimed at achieving human-like performance in tasks of the cognitive science domain, these models are composed of highly interconnected mesh of non-linear computing elements, whose structure is drawn from our current knowledge of biological neural systems. Several neural network learning algorithms have been developed in the past years. In these algorithms a set of rules defines the evolution process undertaken by the synaptic connections of the networks, thus allowing them to learn how to perform specific tasks. Excellent and comprehensive reviews have been written on ANNs (Bishop, 1995; Ripley, 1996), which explain the main principles involved and cover some of the numerous applications.

In horticulture the applications of ANNs are just at the beginning despite their skill and speed in the recognition of patterns in complex, non-linear data, such as those derived from many experimental areas of horticulture. Recently, however, neural networks have been successfully applied to the identification of grapevine (Mancuso et al., 1998; Mancuso, 2001), olive (Mancuso and Nicese, 1999; 2001) and chestnut varieties (Mancuso et al., 1999). In this study, the possibility of using EFA and fractal dimension to describe the leaves of different Sangiovese clones were tested. The parameters coming from such analysis were utilised as input for a backpropagation neural network built to distinguish among the different accessions.

2. Materials and Methods

The study was carried out on seven registered ‘Sangiovese’ clones (“SG-12-T”, “SG-14-T”, “AP1”, ‘Peccioli 1’, “R4”, “R24” and “SS-F9-A5-48”), one presumed clone of the same variety (“SG-19-T”), one coloured variety (‘Abrostine’) and one variety in the process of being registered in the national catalogue of vine varieties (‘Pugnitello’). These varieties and clones were grown in the grapevine germplasm collection of the Department of Horticulture of the University of Florence. The material was chosen because it is currently the object of research using taxonomic criteria based on charts of the descriptive traits or utilising distinction systems based on DNA analysis (Sensi et al., 1996). The analysis procedure employed is summarised in figure 2.

From five plants of clonal original of each genotype, 50 adult leaves were removed as samples. The leaf images were acquired at 300 x 300 d.p.i. (16 million colors) by way of an optical scanner. The leaf margin (1500 equidistant points) was therefore obtained through image analysis. Elliptic Fourier Analysis was carried out according to the method described by Mancuso (1999 a).

The fractal dimension of the leaf was estimated using the “box-counting method”. Implementation of this method was described in detail by Mancuso (1999 b). The elliptic Fourier coefficients and the parameters from fractal analysis were treated as input in a backpropagation neural network generated according to the method described in Mancuso and Nicese (1999).

The outputs obtained by the neural network were utilised to measure the dissimilarities or distances between the various accessions. The Euclidean quadratic distances and dendogram were obtained by way of UPMGA cluster analysis using Statistic software, version 4.0.
3. Results and Discussion

The results obtained with application of artificial neural networks to the descriptive data from Fourier analysis and fractal analysis of leaves are reported in figure 3. All the accessions utilised were correctly classified by the network. The phyllometric parameters of leaves used in the recognition phase for each accession always gave greater output values (nearer to 1) in the correct class.
Some accessions, such as ‘Abrostine’ or ‘Pugnitello’, gave extremely defined output values by virtue of their ‘distance’ from the group of Sangiovese clones. It should be underlined that the use of Fourier analysis, fractal analysis and neural networks in series demonstrated a very elevated discriminating capacity able to correctly identify each examined clone.

The output values obtained by the neural network were analysed utilising the matrix of distances reported in Table 1. The Euclidean quadratic distances presented variable values from 18.1 for registered clones noted in Table 1. The Euclidean quadratic distances presented variable values from 18.1 for registered clones ‘SG-14-T’ and ‘SG-12-T’ and 147.3 for the Pugnitello and Abrostine species. The UPMGA cluster analysis of the matrix of distances separated clones ‘SG-14-T’, ‘SG-12-T’ and ‘SG-19-T’ from the other Sangiovese genotypes. The high degree of nearness between the Tebano clones ‘SG-14-T’, ‘SG-12-T’ could suggest a common origin by mutation from a single plant. Among all the studied genotypes, Abrostine and Pugnitello seem to be the most distant. The high degree of divergence evidences the different origin of these genotypes with respect to the Sangiovese clones (Fig. 4).

The obtained results show that backpropagation neural networks, together with evolved methods of shape analysis, have made it possible to differentiate with remarkable confidence a series of Sangiovese clones using simple phyllometric parameters.

The method points out the possibility of comparing and distinguishing, rapidly and objectively, a theoretically unlimited number of genetic entities based on quantitative taxonomic characteristics. The potential use of this method, therefore, seems quite wide in that it could be applied to both the classification of material present in varietal collections, verifying possible cases of homonymy and synonymy, and the creation of databanks able to identify the processes of evolution of morphologic traits in various phases of the ontogenetic cycle. In addition, the method could be applied to genetic improvement by clonal selection for early and rapid distinction of the biotypes for study.

### Table 1 - Euclidean distance matrix

<table>
<thead>
<tr>
<th></th>
<th>T.19</th>
<th>T.14</th>
<th>Pugn.</th>
<th>Pec. 1</th>
<th>AP 1</th>
<th>Abr.</th>
<th>R 24</th>
<th>R 4</th>
<th>T.12</th>
<th>F 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG-19-T</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG-14-T</td>
<td>32.6</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pugnietello</td>
<td>113.8</td>
<td>95.8</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peccioli 1</td>
<td>79.8</td>
<td>56.4</td>
<td>140.1</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP 1</td>
<td>76.9</td>
<td>55.2</td>
<td>136.7</td>
<td>91.3</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abrostine</td>
<td>81.7</td>
<td>60.3</td>
<td>147.3</td>
<td>113.2</td>
<td>108.2</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R 24</td>
<td>74.1</td>
<td>55.3</td>
<td>134.0</td>
<td>97.4</td>
<td>97.1</td>
<td>106.5</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R 4</td>
<td>58.2</td>
<td>39.8</td>
<td>114.8</td>
<td>85.9</td>
<td>82.0</td>
<td>91.3</td>
<td>72.2</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG-12-T</td>
<td>43.0</td>
<td>18.1</td>
<td>106.3</td>
<td>65.8</td>
<td>65.2</td>
<td>75.5</td>
<td>65.0</td>
<td>49.7</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>SS-F9-A5-48</td>
<td>62.9</td>
<td>43.9</td>
<td>123.8</td>
<td>89.6</td>
<td>85.7</td>
<td>95.1</td>
<td>79.8</td>
<td>67.4</td>
<td>53.6</td>
<td>0.0</td>
</tr>
</tbody>
</table>

![Fig. 4 - Dendogram of ten grapevine accessions generated by UPMGA cluster analysis of the distance values shown in Table 1.](image-url)


