Vitis 37 (1), 27–32 (1998)

Application of an artificial neural network (ANN) for the identification of grapevine genotypes

by

S. Mancuso, P. L. Pisani, R. Bandinelli and E. Rinaldelli

Dipartimento di Ortoflorofrutticoltura Università di Firenze, Italia

Summary: Neural networks were employed to distinguish between 15 accessions of "coloured" (fruit gives intense red colour to the wine) grapevines found in some viticultural zones of Tuscany. Our results enabled us to distinguish, with considerable certainty, between 9 accessions and to denote three pairs of synonyms. The use of neural networks opens interesting prospects for ampelography; its advantages over traditional ampelographic methods are demonstrated.

Key words: ampelography, artificial neural networks, cultivar identification, Vitis vinifera L.

Introduction

The definition and the identification of varieties are of considerable scientific and practical importance in modern viticulture and ampelography due to the high number of varieties (many of ancient origin and genetically heterogeneous) and the frequent cases of homonyms and synonyms.

An important contribution to the solution of ampelographic problems has been made by the OIV-IBPGR-UPOV charts (1983), not only for the completeness of the descriptors but also because, having been internationally adopted, they offer the possibility of common terminology. Recently, statistical methods such as multivariate analysis and discriminant functions (Costacurta *et al.* 1996; Silvestroni *et al.* 1996) have been proposed in the area of ampelographic data processing. Other interesting perspectives come from the analysis of isoenzymes (Subden *et al.* 1987; Benin *et al.* 1988) and of chemical compounds of phenolic nature (such as anthocyanins) having taxonomic value (Eder *et al.* 1994). Finally, DNA marker technologies (simple sequence repeat polymorphism, SSRP; randomly amplified polymorphic DNA, RAPDs; inverse sequence-tagged repeat, ISTR; amplified fragment length polymorphism, AFLP) have proven to be useful tools for characterization of varieties (Bowers *et al.* 1993; Collins and Symons 1993; Thomas and Scott 1993; Thomas *et al.* 1994; Xu *et al.* 1995; Sensi *et al.* 1996). These analytical techniques, however, are time and money consuming and require special laboratory equipment. Therefore, the possibility of using artificial neural network (ANN) for varietal distinction was examined.

ANN has been applied in many fields of experimental science: in physics for the identification of elemental particles (Dawn 1994), in medicine for diagnostic purposes (Fujita *et al.* 1992; Spreckelsen *et al.* 1994; Ster and Dobnikar 1996) and in botany for the taxonomic identification of plankton (Simpson 1992). Moreover, the use of artificial neural networks has been shown to be extremely efficient in the field of handwriting, voice and human face recognition, and to be particularly adapted to the resolution of problems which require the discrimination of different shapes (Hertz *et al.* 1991).

To our knowledge, specific studies about the use of neural networks for varietal recognition of plant species have not been effectuated to date; therefore, it seemed interesting to verify the possibility of applying neural networks for varietal identification of grapevines.

Materials and methods

Plant material: The study was carried out with 15 accessions of "coloured" grapevines (Tab. 1) from the grapevine germplasm collection of the Department of Horticulture of the University of Florence, which were recently the subject of ampelographic studies (Bandinelli *et al.* 1993) and characterization by DNA marker technology (Sensi *et al.* 1996). The vine genotypes were selected because they offered the possibility of validation of the neural network technique. At the time of veraison, samples of 10 leaves located between the 7th and 11th shoot nodes were removed (Alleweldt and Dettweiler 1985). Thirty-one parameters, previously used by Galet (1985); Schneider and Zeppa (1988) and Alessandri *et. al.* (1996), were determined for each leaf (Fig. 1, Tab. 2) by way of a computerised system (AmpeloCADs) which allowed the integration of digitisation and filing of data and its statistical analysis (Alessandri *et al.* 1996).

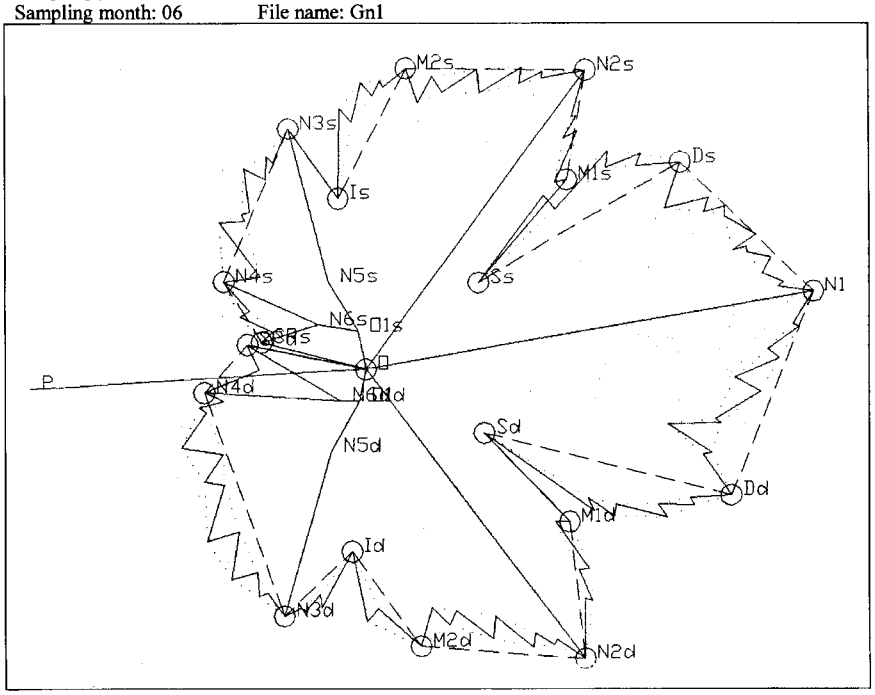
Table 1

List of the grapevine accessions used for neural network technique

Abrostine	Grané
Abrusco	Granoir
Colorino americano	Morone
Colorino di Lucca	Negratino
Colorino di Pisa	Nereto
Colorino di Valdarno	Paspo rosso
Giacché	Tinturié
Grand noir	

Correspondence to: Dr. S. Mancuso, Dipartimento di Ortoflorofrutticoltura, Università degli Studi di Firenze, Via G. Donizetti, 6, I-50144 Firenze, Italy. Fax: +39-55-331497. E-mail: electro@cesit1.unifi.it

Cultivar: Grand noir
Sampling year: 96
Sampling month: 06
Sampling day: 23
Drawing name: Gn1
File name: Gn1



Alpha (N2 O N1) Pi x	0.29465	R6	0.37779
Beta (N O N2) P1 x	0.30617	RSP	0.03219
Tau (N4 O N3) Pi x	0.28916	Area mm^2	22700.57213
Gamma (N6 O1 N5) Pi x	0.31099	Perimeter mm	1348.94455
Delta (Ds N1 Dd) Pi x	0.61956	Amp. Area mm^2	21954.86267
LxW mm^2	36216.17526	Amp. Per. mm	932.76335
Width mm	184.90026	RPf	9.38291
Length mm	195.86871	RAf	1.09831
PO mm	106.42412	RPAf	8.54307
ON1 mm	143.76618	RPfPam	8.54307
SP-index mm	4.62795	TB-index	0.07011
RLW	1.05932	TN-estimate n.	47.00000
R1	0.74026	TH-index	0.04060
R2	0.80478	STS-index	0.05567
R3	0.55984	Digitiz. Time	0.0507
R4	0.36379	PDT x100	0.00409
R5	0.69344		

Fig. 1: AmpeloCADs printout. The entire foliar margin has been digitised, including the foliar teeth. The foliar shape and the reference length have been drawn and the ampelographic points have been labeled.

What is a neural network? The most simple definition of a neural network, more properly referred to as an "artificial" neural network, is provided by the inventor of one of the first neurocomputers, HECHT-NIELSEN (1989). He defines a neural network as "...a computing system made up of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs". Although the mathematics involved with neural networking is not a trivial matter (for review: BISHOP 1995), it is possible to easily gain at least an operational understanding of their structure and function.

Many theoretical models ("paradigms"), dating as far back as the 1950's, have been developed in an effort to mimic a brain's cerebral activity. Most have had limited real-world application potential. The backpropagation paradigm, however, is an extremely effective learning tool that can be applied to a wide variety of problems. Backpropagation related paradigms require supervised training. This mean they must be taught using a set of training data where known solutions are supplied.

Neural networks are typically organized in layers which are made-up of a number of interconnected "nodes" containing an "activation function". Patterns are presented to the network via the "input layer", which communicates to one or more "hidden layers" where the actual processing is done via a system of weighted "connections". The hidden layers are linked to an "output layer" where the answer is output as shown in Fig. 2. During the "training phase", the network's response at the output layer is compared to a supplied set of known answers (training targets). The errors are determined and backpropagated through the network in an attempt to improve the network's response. The nodal weight factors are adjusted by amounts determined by the training algorithm. The iterative procedure of processing inputs through the network, determining the errors and backpropagating the errors through the network to adjust the weights, constitutes the learning process. One training iteration is complete when all supplied training cases have been processed through the network. Iteration continues until the network's response error is kept to a minimum. The network can then be tested on new data and if this proves successful it can be used to predict the output for a given set of input values (for more detailed reviews see HERTZ *et al.* 1991; HINTON 1992).

Table 2

Leaf parameters considered by neural network

Alpha	Ampelographic angle N_2- O-N_1 (radiants/Pi)
Beta	Ampelographic angle N_3- O-N_2 (radiants/Pi)
Tau	Ampelographic angle N_4- O-N_3 (radiants/Pi)
Gamma	Ampelographic angle N_6- O-N_5 (radiants/Pi)
Delta	Ampelographic angle Ds-N_1-Dd (radiants/Pi)
W	Foliar width (mm)
L	Foliar length (mm)
L x W	Length x width (mm^2)
PO	Distance P-O (mm)
ON_1	Distance O-N_1 (mm)
SP-index	Distance SPs-SPd (mm)
RLW	Length / width
R_1	PO / ON_1
R_2	ON_2 / ON_1
R_3	ON_3 / ON_1
R_4	ON_4 / ON_1
R_5	OI / ON_3
R_6	OS / ON_2
RSP	SP-index / ON_1
Area	Area of digitised foliar surface (mm^2)
Perimeter	Length of digitised foliar perimeter (mm)
Amp. area	Ampelographic polygon area (mm^2)
Amp. perimeter	Length of ampelographic polygon perimeter (mm)
RPf	Perimeter / ON_1
RAf	Area / ON_1
RPAf	(Perimeter / Area) x ON_1
RPfPam	Perimeter / Ampelographic perimeter
TN-estimate	Teeth number estimate[1)]
TB-index	Tooth base index[1)]
TH-index	Tooth height index[1)]
STS-index	Sum of tooth sides index[1)]

[1)] see ALESSANDRI *et al.* 1996

M e t h o d s : In total, data from 450 leaves (30 per vine accessions) were utilised. The learning phase for the back- propagation artificial neural network was achieved on a P120 personal computer, by using the phyllometric data. The network (Fig. 2) was designed using a total of 31 inputs represented by the phyllometric parameters (see Tab. 2) and 15 outputs represented by the accessions under examination. Output values were 1 or 0 (true or false).

In order to optimise the neural network activity, the number of "hidden neurons" and the number of iterations was modified. Minimum error was reached with a network composed of 40 hidden neurons, positioned on three levels (15x10x15) with a hybrid activation function, composed of a sigmoidal function for the first and third hidden layers and a gaussian function for the second hidden layer. The learning phase was carried out for 70,000 iterations, at the end of which the RMS (Root Mean Squared error) was 0.065. This

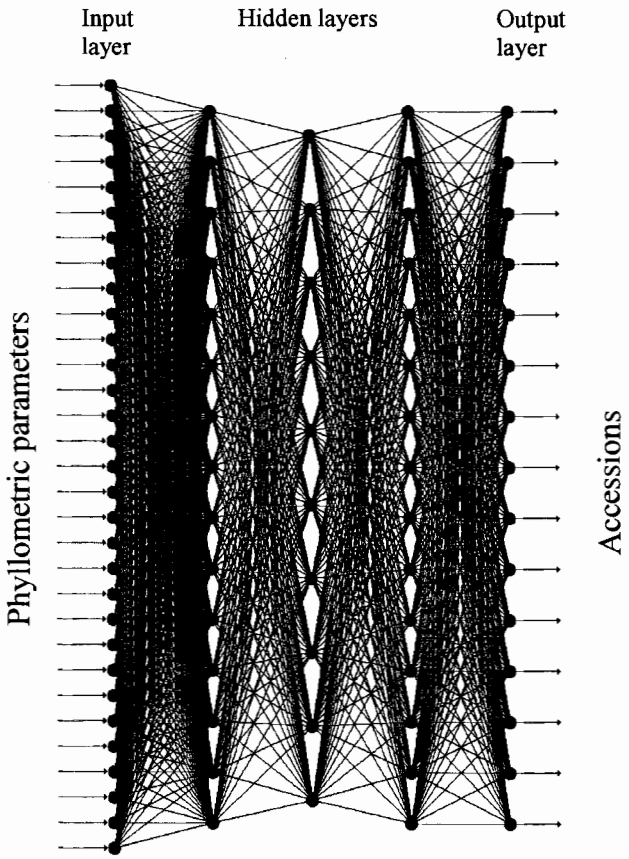


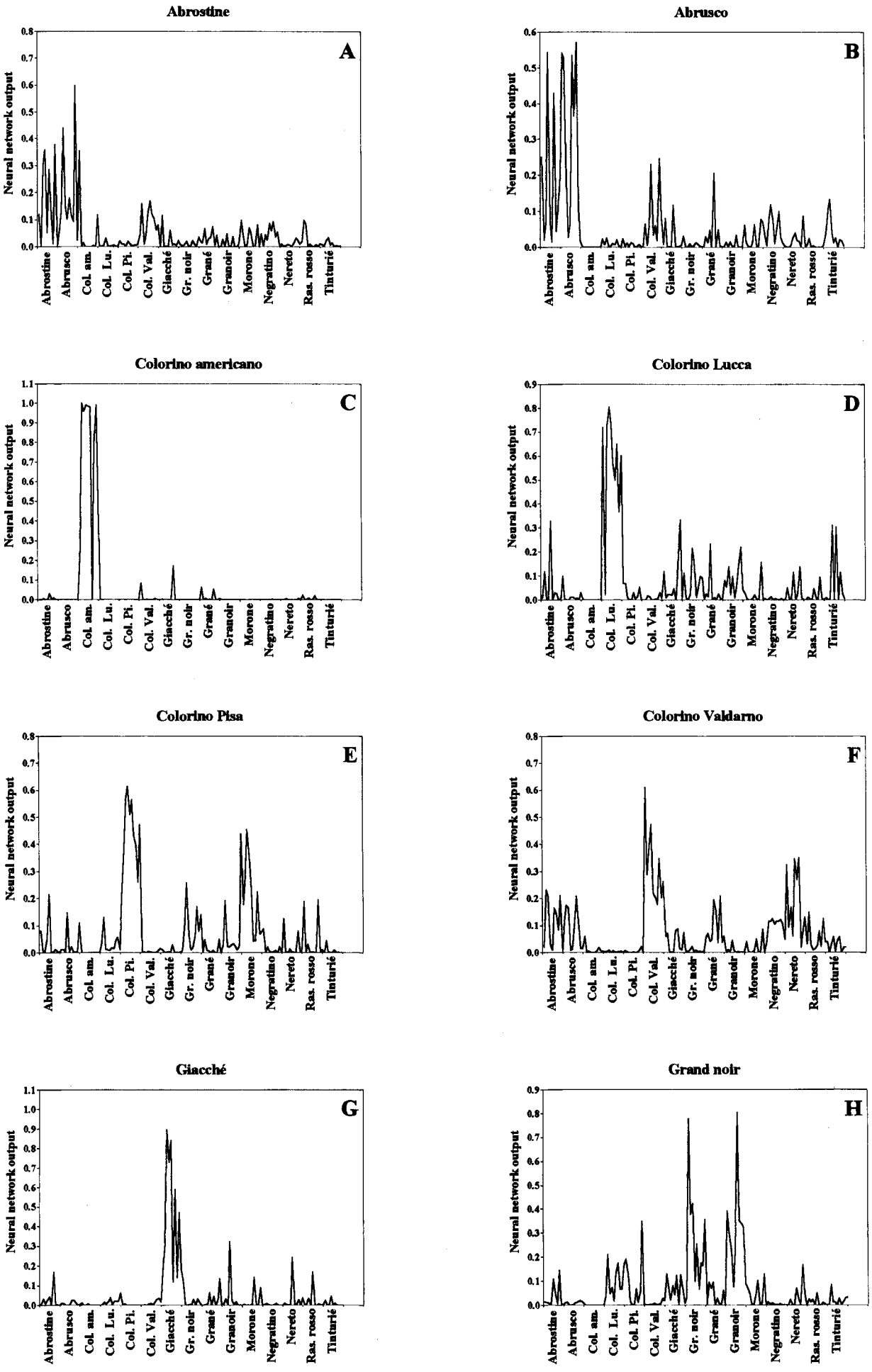
Fig. 2: Diagram of the structure of the artificial neural network.

parameter indicates the effectiveness of the ANN; the smaller the value, the closer the network is to modelling the input data. - The ANN was tested by presenting them with inputs (phyllometric parameters) for which the output was known, so that the predicted and actual outputs could be compared. These data had not been used previously to train the network.

Results and Discussion

Fig. 3 contains the results of the recognition phase of the neural network. Each graph illustrates the network output for input represented by the phyllometric parameters of 10 leaves of a given variety. For example, in diagram 3 A output values for the recognition test effectuated on the accession Abrostine are reported. Each figure presents in the abscissa the names of the 15 studied accessions. The name of the unknown accession is given by the accession in abscissa, which present the highest output values (i.e. the nearest to 1).

By examining the graphs in Fig. 3, it can be noted that the artificial neural network succeeded in identifying the unknown variety with certainty, except in the case of Grand noir (Fig. 3 H) and Granoir (Fig. 3 L), which is obviously a case of synonymy. A considerable similarity between Abrusco (Fig. 3 B) and Abrostine (Fig. 3 A) also appears. It is assumed that these are two clones of the same vine variety; the two different denominations are therefore considered to be synonyms. Similarity was found between Colorino from the province of Pisa (Fig. 3 E) and Morone (Fig. 3 M). Negratino (Fig. 3 N) and Nereto (Fig. 3 O) despite a similar denomination showed different response.


Abrostine
A
Abrusco
B
Colorino americano
C
Colorino Lucca
D
Colorino Pisa
E
Colorino Valdarno
F
Giacché
G
Grand noir
H
Neural network output
Abrostine
Abrusco
Col. am.
Col. Lu.
Col. Pi.
Col. Val.
Giacché
Gr. noir
Grané
Granoir
Morone
Negratino
Nereto
Ras. rosso
Tinturié

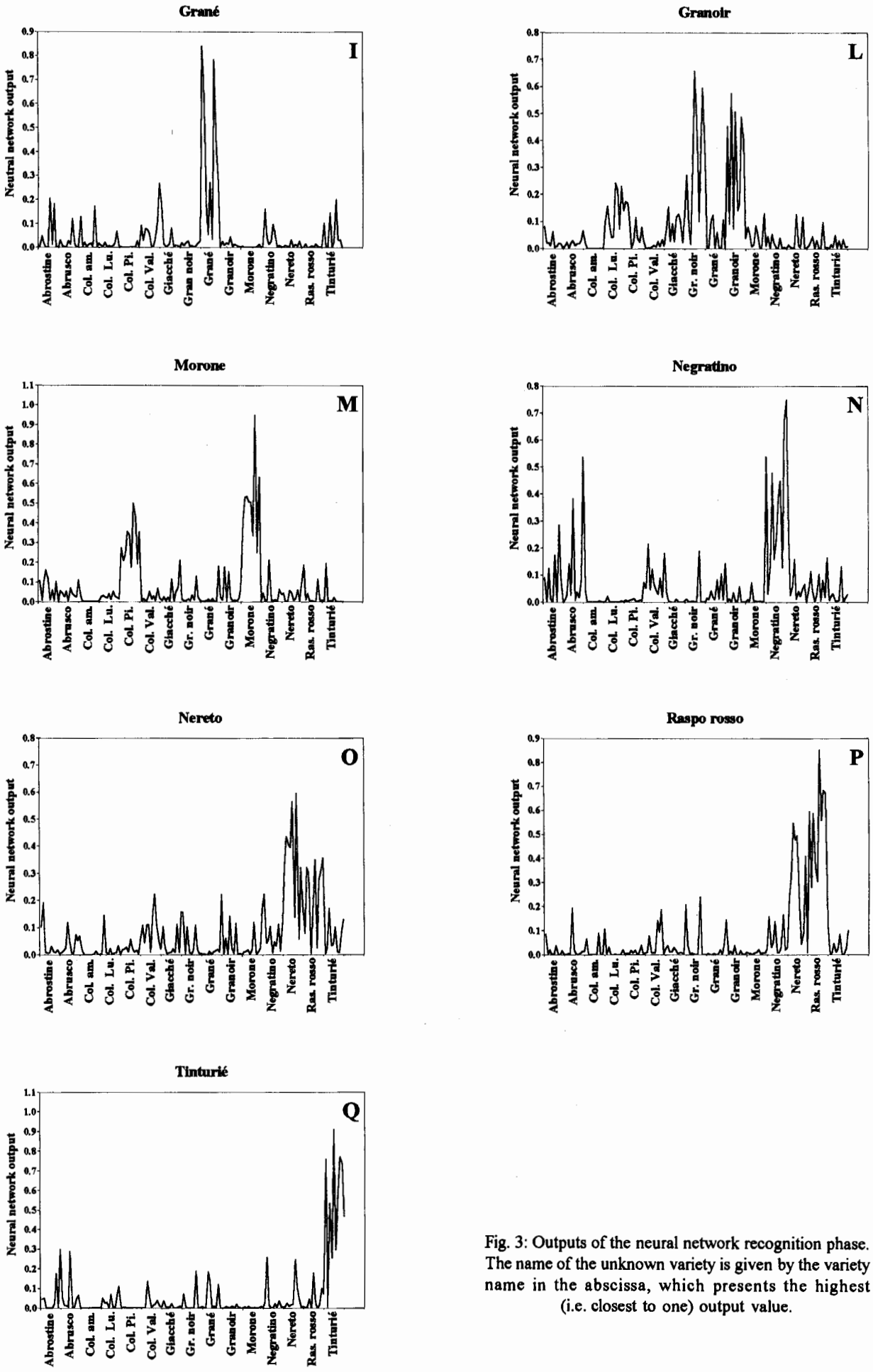


Fig. 3: Outputs of the neural network recognition phase. The name of the unknown variety is given by the variety name in the abscissa, which presents the highest (i.e. closest to one) output value.

Colorino americano (Fig. 3 C), Colorino del Valdarno (Fig. 3 F) and Colorino from the province of Lucca (Fig. 3 D) and the other vine accessions examined, Giacché (Fig. 3 G), Grané (Fig. 3 I), Raspo rosso (Fig. 3 P) and Tinturié (Fig. 3 Q) are clearly differentiated. In agreement with the finding of SENSI *et al.* (1996), the response of the accession Colorino americano was very different from the others, a hypothesis, suggested from the name, could be that this genotype is a hybrid of *Vitis labrusca*.

The results obtained are interesting, above all, because the use of the neural network allowed distinction of the vine accessions, in agreement with the results of studies made on the same genetic material, with ampelographic or molecular marker methods (SENSI *et al.* 1996). Moreover, it seems that the neural network system demonstrates a greater possibility of distinction compared to the method based on ampelographic charts; the two accessions of Colorino found in the provinces of Pisa and Lucca were not distinct with certainty by the ampelographic chart method while the neural network identified a clear difference.

Conclusions

The results present some aspects to be considered. (1) The applied method offers the possibility to compare rapidly and objectively a theoretically unlimited number of genetic entities on the basis of quantitative taxonomic traits. Therefore, on a theoretical level, the creation of an international data base for varieties could by hypothesised. Moreover, it would seem possible, once the data base of ampelographic parameters for a determined varietal assortment is created, to ascertain if other genetic entities correspond or are similar to one or more of the entities already present in the data base. (2) The experimental method can be further improved by increasing the number of ampelographic parameters to include quantifiable morphologic and physiological traits. (3) It should be pointed out that the use of neural networks requires instrumentation which is only limited to personal computers of medium potential and that the time required is only that of digitising of the phyllometric parameters (approximately one minute per leaf); this time could be considerably reduced by a low-cost automated system.

Acknowledgements

The authors wish to thank the direction of the "Agricola San Felice" (Castelnuovo Berardenga, Siena, Italy) for their technical collaboration and management of the experimental field. Work carried out within a research project financed by A.R.S.I.A. (Agenzia Regionale per lo Sviluppo e l'InnovazioneAgricola e Forestale of the Regione Toscana).

References

ALESSANDRI, S.; VIGNOZZI, N.; VIGNINI, A. M.; 1996: AmpeloCADs (ampelographic computer-aided digitizing system): An integrated system to digitize, file, and process biometrical data from *Vitis* spp. leaves. Amer. J. Enol. Viticult. **47**, 257-267.

ALLEWELDT, G.; DETTWEILER, E.; 1985: Ampelographic studies to characterize varietes. Proc. IV Simposio internazionale di genetica della vite, Verona, Italy, April 1985.

BANDINELLI, R.; PISANI, P. L.; VIGNOZZI, N.; 1993: Caratterizzazione ampelografica di alcuni vitigni autoctoni toscani. Meet. Proc. La ricerca sperimentale in corso per la viticoltura in Toscana, S. Felice, Italy, May 1993, 131-163.

BENIN, M.; GASQUEZ, J.; MAHFOUDI, A.; BESSIS, R.; 1988: Caractérisation biochimique des cépages de *Vitis vinifera* L. par électrophorèse d'isoenzymes foliaries: Essai de classification de variétés. Vitis **27**, 157-172.

BISHOP, C. M.; 1995: Neural Networks for Pattern Recognition. Oxford Univ. Press, Oxford.

BOWERS, J. E.; BANDMAN, E. B.; MEREDITH, C. P.; 1993: DNA fingerprint characterization of new polymorphic cultivars. Amer. J. Enol. Viticult. **44**, 266-274.

COLLINS, G. G.; SYMONS, R. H.; 1993: Polymorphism in grapevine DNA detected by the RAPD PCR technique. Plant Mol. Biol. Reptr. **11**, 105-112.

COSTACURTA, A.; CALÓ, A.; CARRARO, R.; GIUST, M.; ANTONIAZZI, M.; LAZZARO, G.; 1996: Metodologie computerizzate per la caratterizzazione di vitigni (I. contributo). Riv. Viticolt. Enol. **1**, 27-34.

DAWN, T.; 1994: Neural computing makes its mark in science. Scientific Computing **3**, 25-30.

EDER, R.; WENDELIN, S.; BARNA, J.; 1994: Classification of red wine cultivars by means of anthocyanin analysis. First report: use of multivariate statistical methods for the differentiation of grape samples. Mitt. Klosterneuburg **44**, 201-212.

FUJITA, H.; KATAFUCHI, T.; UEHARA, T.; NISHIMURA, T.; 1992: Neural network approach for the computer-aided diagnosis of coronary artery diseases in nuclear medicine. In: IJCNN, International Joint Conference on Neural Networks, 215-220, Baltimore, IEEE.

GALET, P.; 1985: Précis d'Ampelographie Pratique. Imprimerie Charles Dehan, Montpellier, France.

HECHT-NIELSEN, R.; 1989: Neural Network Primer: Part I. AI Expert (February), 4-51.

HERTZ, J.; KROGH, A.; PALMER, R.; 1991: Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City, California.

HINTON, G. E.; 1992: How neural networks learn from experience. Scientific American. **267**, (September), 144-151.

OIV-IBPGR-UPOV; 1983: Code de caractères descriptifs des variétés et espèces de *Vitis*. Office Intern. Vigne Vin, Paris.

SCHNEIDER, A.; ZEPPA, G.; 1988: Biometria in ampelografia: l'uso di una tavoletta grafica per effettuare rapidamente misure fillometriche. Vignevini **15** (9), 37-40.

SENSI, E.; VIGNANI, R.; ROHDE, W.; BIRICOLTI, S.; 1996: Characterization of genetic biodiversity with *Vitis vinifera* L. Sangiovese and Colorino genotypes by AFLP and ISTR DNA marker technology. Vitis **35**, 183-188.

SILVESTRONI, O.; INTRIERI, C.; DI DOMIZIO, N.; 1996: Impiego di metodi fillometrici per la caratterizzazione di alcuni vitigni dell'Emilia Romagna. Riv. Viticolt. Enol. **49** (1), 17-26.

SIMPSON, R.; 1992: Biological pattern recognition by neural networks. Marine Ecology Progress Series **79**, 303-308.

SPRECKELSEN, S. VON; GRUMSTUP, P.; JOHNSEN, J.; HANSEN, L. K; 1994: Modeling of glaucoma induced changes in the retina and neural net assisted diagnosis. In: VLONTZOS J. *et al.* (Eds.): Neural Networks for Signal Processing IV, 490-498. Institute of Electrical and Electronics Engineers Inc., New York, NY.

STER, B.; DOBNIKAR, A.; 1996: Neural networks in medical diagnosis: Comparison with other methods. In: BULSARI A. B. *et al.* (Eds.): Solving Engineering Problems with Neural Networks, 427-430. Systeemitekniikan seura ry, Turku.

SUBDEN, R. E.; KRIZUS, A.; LOUGHEED, S. C.; CAREY, K.; 1987: Isozyme characterization of *Vitis* species and some cultivars. Amer. J. Enol. Viticult. **38**, 176-181.

THOMAS, M. R.; CAIN, P.; SCOTT, N. S.; 1994: DNA typing of grapevine: a universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol. Biol. **25**, 939-949.

- -; SCOTT, N. S.; 1993: Microsatellite repeats in grapevine reveal DNA polymorphism when analysised as sequence-tagged sites (STSs). Theoret. Appl. Genet. **86**, 985-990.

XU, H.; WILSON, D. J.; ARULSEKAR, S.; BAKALINSKY, A. T.; 1995: Sequence-specific polymerase chain reaction markers derived from randomly amplified polymorphic DNA markers for fingerprinting grape (*Vitis*) rootstocks. J. Amer. Soc. Hort. Sci. **120**, 714-720.

Received September 23, 1997