Food Chemistry 112 (2009) 1046–1052

Contents lists available at ScienceDirect

Food Chemistry

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

Analytical Methods

Classification of Slovak white wines using artificial neural networks and discriminant techniques

Dasa Kruzlicova [a], Jan Mocak [a,b,*], Branko Balla [a], Jan Petka [c], Marta Farkova [d], Josef Havel [d]

[a] Institute of Analytical Chemistry, Faculty of Chemical and Food Technology, Slovak University of Technology, Radlinskeho 9, SK-81237 Bratislava, Slovakia
[b] Department of Chemistry, Faculty of Natural Sciences, University of Ss. Cyril and Methodius, Nam. J. Herdu 2, SK-91701 Trnava, Slovakia
[c] Food Research Institute, Priemyselna 4, SK-82475, Bratislava
[d] Department of Analytical Chemistry, Faculty of Science, Masaryk University, Kotlarska 2, CZ-611 37 Brno, Czech Republic

ARTICLE INFO

Article history:
Received 25 January 2008
Received in revised form 23 April 2008
Accepted 22 June 2008

Keywords:
Wine classification
Wine authentication
Artificial neural networks
Feature selection
ANOVA

ABSTRACT

This work demonstrates the possibility to use artificial neural networks (ANN) for the classification of white varietal wines. A multilayer perceptron technique using quick propagation and quasi-Newton propagation algorithms was the most successful. The developed methodology was applied to classify Slovak white wines of different variety, year of production and from different producers. The wine samples were analysed by the GC–MS technique taking into consideration mainly volatile species, which highly influence the wine aroma (terpenes, esters, alcohols). The analytical data were evaluated by means of the ANN and the classification results were compared with the analysis of variance (ANOVA). A good agreement amongst the applied computational methods has been observed and, in addition, further special information on the importance of the volatile compounds for the wine classification has been provided.

1. Introduction

Artificial neural networks (ANN) are used more and more frequently in chemistry (Gasteiger & Zupan, 1993). They have been applied in the authors' groups for various purposes, e.g. optimisation (Dohnal, Farková, & Havel, 1999; Farková, Peña-Mendez, & Havel, 1999; Havel, Madden, & Haddad, 1999; Havel, Peña, Rojas-Hernández, Doucet, & Panaye, 1998; Havliš, Madden, Revilla, & Havel, 2001; Pokorná, Revilla, Havel, & Patočka, 1999), quantification of unresolved peaks (Bocaz-Beneventi, Latorre, Farková, & Havel, 2002; Dohnal, Li, Farková, & Havel, 2002), estimation of peak parameters, estimation of model parameters in the equilibria studies (Havel, Lubal, & Farková, 2002), etc. Pattern recognition and object classification is also an important application area for the ANN (Ball et al., 2002; Fidencio, Ruisanchez, & Poppi, 2001; Sanni, Wagner, Briggs, et al., 2002).

A number of authors deal with the classification of wines (Almela, Javaloy, Fernandez-Lopez, & Lopez-Roca, 1996; Cliff & Dever, 1996; De la Presa-Owens & Noble, 1995; Gonzalez, Mendez, Sanchez, & Havel, 2000; Noble & Shannon, 1987). Varietal wines have been classified in terms of amino acid profiles (Etiévant, Schlich, Bouvier, Symonds, & Bertrand, 1988; Vasconcelos & Chaves das Neves, 1989) or organic acid profiles (Etiévant, Schlich, Cantagrel, Bertrand, & Bouvier, 1989), protein fractions (Almela et al., 1996; Larice, Archier, Rocheville-Divorne, Coen, & Roggero, 1989; Pueyo, Dizy, & Polo, 1993), traditional analytical parameters of wine (acids, ethanol, fructose, phenolic compounds, pH, etc.) (Almela et al., 1996; Cliff & Dever, 1996; Moret et al., 1980) and descriptive analysis (De la Presa-Owens & Noble, 1995; Dumont & Dulau, 1996; Noble & Shannon, 1987). However, the most common way to classify varietal wines is by monitoring the content of volatile aroma compounds mainly by employing gas chromatography and a subsequent application of various statistical methods (De la Calle Garcia, Reichenbaecher, & Danzer, 1998; Ferreira, Fernandez, & Cacho, 1996; Lozano, Santos, & Horrillo, 2005; Medina, 1996; Rapp & Guentert, 1985; Rapp, Guentert, & Heimann, 1985; Rapp, Suckrau, & Versini, 1993). Classification of wines by variety was described also in further papers (García et al., 2006; Lozano et al., 2006). However, the wine classification by other criteria, e.g. by geographical origin has also been described (Ballabio, Mauri, Todeschini, & Buratti, 2006; Capron, Smeyers-Verbeke, & Massart, 2006). In this work we have studied the possibility to use the ANN for the classification of Slovak white varietal wines with the aim to classify wines by different variety, producer/location and the year of production.

2. Materials and methods

2.1. Chemicals, samples and instrumentation

All used reagents were of analytical grade and all dilutions and sample preparations were made using deionized water (Milli-Q

* Corresponding author. Tel.: +421 33 5573360.
E-mail address: jan.mocak@stuba.sk (J. Mocak).

0308-8146/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2008.06.047

water purification system, Millipore, MA, USA). The following types of varietal wine samples were used: *Welsch Riesling*, *Gruener Veltliner* and *Mueller Thurgau*, all of vintage 1996, 1997 and 1998. The wine samples came from five producers located in West and South West Slovakia, namely in (a) Nitra, (b) Veľký Krtíš, (c) Pezinok, (d) Dvory nad Žitavou, and (e) Modra. The samples originated from the last mentioned producer were used only partly.

2.2. Isolation of volatile aroma compounds

Wine (10 mL) was pipetted into a conical screw-cap test tube, then 4.2 g of ammonium sulfate, 100 μL of 1,1,2-trichloro-1,2,2-trifluoroethane (Freon F113) and 10 μL of geranyl butanoate standard solution in F113 (c = 10 mg/mL) were added. The tube content was shaken intensively by a laboratory shaker for 1 h and then centrifuged at 1900 rpm for 15 min. The organic layer was then removed by a micro syringe and directly analysed by gas chromatography.

2.3. Gas chromatography with mass spectrometric detection

A Hewlett Packard HP 5890 II gas chromatograph with a Hewlett Packard HP 5971A mass selective detector was used for the GC–MS analysis. Sample injection was made using the splitless mode; the injected freon extract volume was 0.6 μL. A fused silica DB-WAX capillary column (30 m × 0.32 mm × 0.25 μm) was employed with the following temperature programme: 35 °C (0.5 min), 4 °C/min, 220 °C. Helium was used as the carrier gas (linear velocity 25 cm s^{-1}). The ionisation energy was 70 eV. The selected compounds were identified by comparing the spectral pattern with the NIST05 MS spectral database or using reference materials and published retention indices (Jennings & Shibamoto, 1980; Kondjoyan & Berdagué, 1996). Further instrumental details are described in (Petka, Farkas, Kovac, Balla, & Mocak, 2000; Petka et al., 1999).

Twenty-six volatile aroma compounds were originally analysed for two sets of wine samples containing three wine varieties from 4 or 5 different producers. Altogether 36 wine samples were analysed for three vintages in the first data set (four producers, one sample for the given variety, vintage and producer) and 87 samples (five producers, two samples for the given variety, vintage and producer, with three exceptions) were analysed in the second data set. The choice of the volatile aroma compounds is generally influenced by the grape type and not much by the fermentation process. Hence, the data tables should have contained 36 × 26 or 90 × 26 data (rows × columns) but due to several reasons, the obtained data were not fully complete. As for our way of data processing a complete data table is needed, only 19 (first data set) or 20 (second set) variables involving the concentrations of volatile aroma compounds in the table columns were used – those, which are surveyed in Table 1.

2.4. Artificial neural networks

Since the theory of the ANN is well described in monographs (Gasteiger & Zupan, 1993; Kvasnička et al., 1997; McClelland & Rumelhart, 1988) and scientific literature (Farková et al., 1999; Havel et al.,1998, 2002) only a short description of the ANN principles will be given. The use of the ANN for data processing can be characterised by analogy with biological neurons. The artificial neural network itself consists of interconnected neurons situated in an input layer, one or more hidden layer(s) and an output layer. The input neurons receive the input data characteristic for each observation, the output neurons provide predicted value or pattern of the studied objects. In most cases, the ANN architecture consists of two active layers – one hidden and one output layer. The neurons of two adjacent layers are mutually connected and the importance of each connection is expressed by weights.

The role of the ANN is to transform the input information into the output one. During the training process the weights are corrected so as to produce output values as close as possible to the desired (or target) values. The propagation of the signal through the network is determined by the weights associated to the connections between the neurons, which represent the synaptic strengths in biological neurons. The goal of the training step is to correct the weights w_{ij} so that they will give a correct output vector $\boldsymbol{y}$ (as close as possible to the known target vector $\boldsymbol{d}$) for the input vector $\boldsymbol{x}$ from the training set. After the training process has been completed successfully, it is hoped that the network will give a correct prediction for any new object x_n, not included in the training set.

The hidden x_i and the output y_i neuron activities are defined by the relations

$$x_i = t(\xi_i) \quad \text{or } y_i = t(\xi_i), \tag{1a, b}$$

$$\xi_i = \sum_{j=1}^{p} w_{ij} x_j + v_i, \tag{2}$$

where $j = 1,\ldots,p$ concern neurons x_j in the previous layer which precede the given neuron i. ξ_i is the net signal – the sum of the weighted inputs from the previous layer, v_i is the bias (offset), w_{ij} is the weight and $t(\xi_i)$ is *transfer function*, described by the threshold logic function, sigmoid function or hyperbolic tangent function (Otto, 1999, chap. 8.2). The most common sigmoid function is of the form

$$t(\xi_i) = \frac{1}{1 + e^{-k\xi_i}}, \tag{3}$$

where k is a constant.

The aim of the neural network training is to *minimise the error E* by changing the weights and offsets

$$E = \sum_{i=1}^{r} E_i = \sum_{i=1}^{r} (y_i - d_i)^2, \tag{4}$$

Table 1
Analysed volatile aroma compounds in Slovak white wines

Code	Type[a]	Compound	Code	Type[a]	Compound
v1	E	Ethylhexanoate	v12	T	α-Terpineol
v2	A	(E)-3-Hexen-1-ol	v16	T	Citronellol
v3	A	(Z)-3- Hexen-1-ol	v18	E	2-Ethylphenylacetate
v4	A	(E)-2-Hexen-1-ol	v19	T	Geraniol
v5	E	Ethyl 2-hydroxy-3-methylbutanoate	v20	T	3,7-Dimethyl-1,5-octadien-3,7-diol
v6	T	(E)-Furan linalool oxide	v22	A	1-Hexanol
v7[b]	T	Neroloxid	v23	E	Diethyl succinate
v8	T	(Z)-Furan linalool oxide	v24	C	Hexanoic acid
v9	E	Ethyl 3-hydroxy-butanoate	v25	A	Benzyl alcohol
v10	T	Linalool	v26	A	2-Phenylethanol

[a] Variable type: T, terpene; E, ester; A, alcohol; C, carboxylic acid.
[b] In the first part of the study this compound was not included so that only 19 variables were used.

where r is the number of the input–output vector pairs in the training set, d_i is the respective component of the required output vector and y_i is the response to the adequate component x_i of the input vector. The error E is minimised most often by the steepest descent method or another gradient method. The described theory is valid mainly for the multilayer perceptron algorithms like back propagation, quick propagation and quasi-Newton, with some differences in details.

The ANN calculations were made using Trajan 4.0 software package (Trajan Software, 1999). Usually the default settings of Trajan software were used but for the final network adjustment the Jog Weights procedure was applied, which adds a small random quantity to each weight to help the training algorithm go out of a local minimum. For comparison purposes a non-hierarchical one-way-ANOVA as well as three discriminant techniques was used – linear discriminant analysis, quadratic discriminant analysis and the *K*th nearest neighbour (Khattree & Naik, 2000). They should bring light on the classification process from a different angle.

3. Results and discussion

3.1. Building neural networks

Thirty-six wine samples of three varieties, produced by four producers in the course of three years (three vintages) characterised by 19 finally accepted variables (representing the concentrations of the volatile, aroma-creating compounds) have been used in the training process. The wine samples were classified using the ANN techniques according to the following criteria: (a) *variety* (three classes: Riesling, Veltliner and Mueller), (b) *producer* (four classes: *Nitra*, *Krtis*, *Pezinok* and *Dvory*), and (c) *vintage* (three classes: 1996, 1997 and 1998).

During the training step numerous networks with different architectures were examined. Since the number of input neurons as well as output neurons were set by the number of variables (19) and number of classes (three for variety as well as vintage, four for the producer), respectively, the network selection was based on the appropriate selection (a) of the number of hidden layers in multilayer perceptrons, (b) the number of hidden neurons in the network. The key decision on the number of hidden layers was made by Trajan's Intelligent Problem Solver. By the Solver a variety of algorithms for different network types are automatically tested and the best alternatives are determined. The Solver was applied separately for each of the three classification problems and it was found in all cases that the best network type is the three-layer multilayer perceptron (3-MLP), i.e. that with one hidden layer.

The number of hidden neurons, N_h, was found by examining several types of the 3-MLP with regard to the corresponding final root mean squared (RMS) error. The optimal number of hidden units was found according to the break on the RMS vs. N_h dependence for all studied problems. Decision about the network design was set for the corresponding training sets but not individually for each validation set in the leave-one-out procedure.

Amongst several applied perceptron learning algorithms the lowest RMS error was reached by a combination of quick propagation (QP) (Fahlman, 1988) and quasi-Newton (QN) (Bishop, 1995) methods. The learning process was initialised by the QP algorithm and then, after reaching the RMS error value of about 10^{-3}, it continued by using the QN algorithm. The number of epochs (iterative adjustment of optimum weights and thresholds for the entire training set) was ca. 300 when using QP and ca. 10,000 for QN. Neither QP or QN algorithms used alone nor the back propagation (BP) algorithm were as successful as the mentioned QP and QN combination. The main difference between the most common BP and more advanced QP or QN algorithms is in the weight updating procedure. Whereas BP calculates the local gradient of each weight with respect to each *case* (analysed object, e.g. a wine sample) and adjusts the network weights after each training case, QP or QN works out the average gradient of the error surface across all cases, before updating the weights once at the end of the epoch (the epoch is a single pass through the entire training set). The use of the fast converging QN algorithm after the QP one eliminates danger of its lesser numerical stability, e.g. a possible convergence to a local minimum.

3.2. Classification of wines of different variety, producer and the year of production

The samples of three white wine *varieties* were classified by means of the network and algorithms described in the previous paragraphs. The optimal network performance was found for the ANN architecture 19-2-3. Under the described conditions of the training session all 36 samples of the training set were successfully classified, i.e. a 100% classification success was recorded.

Validation of the applied neural network should be based on an independent test set of wine samples. However, due to a relative small number of available wine samples the leave-one-out (jack knife) validation technique was applied. Thirty-six couples of the training and test sets were created in such a way that 35 samples created the training set and the remaining single sample created the given test set, so that each wine sample was used just once as the test set. This sample arrangement enabled to perform 36 independent test sample evaluations by which a 69.4% success in the classification according to *variety* was recorded. This result is far above the 33.3% limit – the classification success corresponding to random classification using three classes.

Validation of the wine classification according to the producer and vintage was performed by the described leave-one-out technique in a similar way. The calculated success of classification was 69.4% for the *producer* and 80.6% for the *vintage* classification criteria. The classification according to *vintage* is relatively even more successful taking into consideration only a 25% success expected by a random classification into four classes.

3.3. Feature selection, detection of the best variables for the used classification criterion

The importance of an individual variable, i.e. the concentration of a selected volatile compound, is different and is dependent on the used classification criterion – variety, producer and vintage. Therefore the detection of the most useful variables for the wine classification according to the respective criterion brings new chemical information on the classified wines. Such a process is generally called feature selection. The main aim of such a selection procedure is the elimination of redundant variables thus avoiding problems with overfitting in the ANN application.

Feature selection can be implemented in the ANN (Trajan Software, 1999) in several ways: (1) forward feature selection, (2) backward feature selection, (3) the application of the genetic algorithm, (4) the application of the sensitivity based techniques (Sensitivity Analysis, Weigend Regularization) and (5) the reduction of variables by principal components analysis, usually combined with the use of the autoassociative networks (the last way is known as a feature reduction process). We have used the first two ways, which are frequently used in multivariate data analysis, e.g. in discriminant analysis or logistic regression; the forward and backward feature selection algorithms add or remove a variable one at a time, starting with zero and a full number of variables, respectively. Even though we have used both of them, a better results reproducibility was found for the backward selection. That is in accord with a generally known observation that the forward selection may miss key

variables if they are interdependent, as is true in our case (the results of correlation analysis have revealed five pairs of variables with the correlation coefficient r_{ij} above 0.9 and further nine pairs with $r_{ij} > 0.8$). Finally, it is also in accord with the rule that, generally, backward selection is to be preferred if there are a small number of variables, say, twenty or less (Trajan Software, 1999).

Table 2 shows the ranks of the most important variables detected by the backward selection process for all three kinds of the wine classification. The table shows the most influencing seven variables for each classification criterion. In addition, the achieved ranks of the best variables are then compared with the variable ranking obtained by ANOVA (ranks in parentheses). Some of the used variables are highly correlated (which was proved by large values of the pair correlation coefficients) and it influenced the way of the variable elimination in the backward selection process applied in ranking variables in the ANN but not in ANOVA. Therefore the same ranking cannot be expected, even though many concordant results were obtained. Discordant ranks, i.e. when some variables of the seven best ANN variables are not included amongst the seven best ranked by ANOVA, are highlighted in Table 2. Such behaviour was observed in six cases out of 21 and the respective ranks are mostly not very distant.

Very interesting from the chemical viewpoint is the type of the best variables for individual classification criteria. It is clear that the selected terpenes are most influential for wine classification by *variety*. On the contrary, terpenes are unimportant for the two remaining ways of classification for which the selected esters and alcohols are important. The wine classification by *producer* is mostly influenced by the selected esters and the selected alcohols are most influencing for the *vintage* classification.

3.4. Ranking of variables according to their importance by ANOVA

Importance of the individual volatile, aroma compounds in wine has been ranked also by analysis of variance (ANOVA). Ranking by ANOVA was made individually for each of the 19 selected variables so that the calculation by this approach regards especially the considered variable and is independent of other variables. The lower the *p*-value found by ANOVA, the larger the importance of this variable for the given type of classification. The same three classification criteria were used as in the previous ANN classification. The results are shown in Table 3.

Terpenes represent the most important variables for classifying white wine samples according to variety. Six of seven terpenes are

Table 2
ANN classification of white wines

Rank of the best variables	White wine classification according to								
	Variety			Producer			Vintage		
1	v12	T[a]	(1)[b]	v23	E	(1)[b]	v2	A	(3.5)[b]
2	v10	T	(3)[b]	v18	E	(6)[b]	v22	A	(1)[b]
3	v2	A	(7)[b]	v2	A	(15)[c]	v24	C	(6)[b]
4	v25	A	(9)[c]	v5	E	(8)[c]	v4	A	(2)[b]
5	v6	T	(2)[b]	v1	E	(2)[b]	v9	E	(9)[c]
6	v16	T	(10)[c]	v24	C	(7)[b]	v26	A	(3.5)[b]
7	v19	T	(6)[b]	v25	A	(10)[c]	v1	E	(7)[b]
Resume	5 Terpenes 2 Alcohols			4 Esters 2 Alcohols 1 Carboxylic acid			4 Alcohols 2 Esters 1 Carboxylic acid		

Best variables selected by the backward feature selection procedure and compared with the ANOVA results (ranks in parentheses).
[a] Variable type: T, terpene; E, ester; A, alcohol; C, carboxylic acid, further details are in Table 1.
[b] Agreement with the rank (in parenthesis) found by the one-way-ANOVA, crossed model (best seven variables in the ANN classification compared to the best seven variables found by ANOVA).
[c] Disagreement with the rank found by the one-way-ANOVA (highlighted). ANOVA ranking details are shown in Table 3.

Table 3
Non-hierarchical one-way-ANOVA: *p*-values and corresponding ranks of the volatile aroma compounds selected for wine classification according to *variety*, *producer* and *vintage*

Code	Type	ANOVA, *p*-value[a]			ANOVA, rank of the variable[b]		
		Variety	Producer	Vintage	Variety	Producer	Vintage
v1	E	0.105	[c]2.33×10^{-7}	[c]0.00066	14	2	7
v2	A	[c]0.00093	[c]0.288	[c]0.00001	7	15	3.5
v3	A	0.291	[c]0.00001	[c]0.0841	16	3.5	14
v4	A	0.0658	[c]0.00017	[c]1.68×10^{-7}	13	5	2
v5	E	0.415	[c]0.00175	[c]0.02797	17	8	11
v6	T	[c]1.65×10^{-6}	0.417	[c]0.00166	2	16	8
v8	T	[c]0.00013	0.446	[c]0.0688	4	17	13
v9	E	0.0513	0.218	[c]0.00169	12	14	9
v10	T	[c]1.72×10^{-6}	0.964	0.413	3	19	16
v12	T	[c]1.43×10^{-7}	0.17	[c]0.00926	1	12	10
v16	T	[c]0.0105	[c]0.00252	0.439	10	9	17
v18	E	[c]0.0196	[c]0.00022	0.504	11	6	18
v19	T	[c]0.00061	0.725	0.874	6	18	19
v20	T	[c]0.00052	0.216	0.135	5	13	15
v22	A	[c]0.00103	[c]0.00001	[c]3.04×10^{-8}	8	3.5	1
v23	E	0.537	[c]2.93×10^{-10}	[c]0.00018	18	1	5
v24	C	0.865	[c]0.00029	[c]0.00042	19	7	6
v25	A	[c]0.00167	[c]0.00261	[c]0.0449	9	10	12
v26	A	0.129	[c]0.00384	[c]0.00001	15	11	3.5

[a] The *p*-value expresses the probability that the variable is insignificant for the classification according to *variety*, *producer* and *vintage*, respectively. A low value means a large significance of the variable whose code is given in the first column.
[b] Rank of the variable according to the ANOVA *p*-value for three kinds of wine classification; the best rank is given by the lowest *p*-value.
[c] Significant with the the probability *P* equal or larger than 95%; $P = 100(1 - p)$.

ranked amongst the first seven selected variables, which confirm the ANN results. It is also clear that terpenes are not amongst the influential variables for wine classification according to producer or vintage. The first terpene compound in the mentioned classifications is ranked as the eighth and the ninth, respectively. Ester and alcohol compounds play a large role for the two remaining classification types. Considering best eight variables for the producer classification, they contain four esters (first two ranks), three alcohols and the only used acid. With regard to *vintage*, the most influential variables are alcohols (first four ranks), then two esters and the mentioned acid. Also these results are in very good accordance with the ANN results. The most important variable for all performed classifications is 1-hexanol (v22). The following four variables are very useful at the same time for two kinds of classification – by *producer* and *vintage*: 1-hexanol (v22), (E)-2-hexen-1-ol (v4), ethylhexanoate (v1), and diethyl succinate (v23).

3.5. Success in classification by ANN after feature selection

Based on a successfully performed feature selection, which was found to be in very good accordance with the ANOVA results, we have used the best seven variables for training the new ANN with the seven input and three output neurons. Calculations for the factors *variety* and *vintage*, followed by the leave-one-out validation were performed in the same way as described above. The optimum network architecture 7-3-3 was found according to the root mean squared error in both cases; the success in classification was much improved compared to the ANN results received with all 19 variables at the input (Table 4). However, no such improvement was found for the factor *producer* where the analogous 7-3-4 network was first examined (due to four classes at the output). With three added variables, a new network with 10 input neurons was trained and subsequently validated by the leave-one-out method. The achieved results were substantially better. Finally, the best classification was found for the 7-4-4 network. It is important to note that the selection of the optimum network architecture was made using all wine samples even though in the leave-one-out procedure one sample was always omitted. The results contained in Table 4 exhibit a 100% or a near 100% classification success for all three classification criteria (factors) when the best performing seven variables and the optimally created and trained neural network were used. About the same classification success was found for all training sets.

3.6. Classification of wine samples using enlarged set of data

When working with neural networks there exists a problem of overfitting, the consequence of it is a weak prediction ability of the used network even though the training set performance is excellent. The best solution of this problem is using a special validation set of data, in which several samples are taken from each class. Of course, these samples are not included in the training set.

Since the hitherto mentioned results dealt with a dataset containing very few samples the second, enlarged data set was also studied, in which two samples of each of three varieties, five producers and three vintages were taken. Due to absence of three samples altogether 87 samples were analysed and classified. This set was divided into the training set of 72 samples and the test

Table 4
Success in the classification of white wines using the ANN, best variables and optimally created networks

Factor[a]	AN network	Success[b] (%)	n_i[c]
Variety	19-2-3	69.4	11
	7-3-3	100.0	0
Producer	19-2-4	69.4	11
	10-2-4	94.4	2
	7-4-4	97.2	1
Vintage	19-2-3	80.6	7
	7-3-3	100.0	0

[a] Classification criterion.
[b] Success in classification by the leave-one-out algorithm; it is given by the number of correctly classified samples divided by the total 36 samples.
[c] n_i represents the number of the incorrectly classified samples in the leave-one-out validation.

Table 5
Selection of wine samples creating the data test set and classification performance by ANN for three different classification criteria

Variety	Year	Producer	Sample	Sample no.
RV	98	Nitra	a	60
RV	97	Velky Krtis	b	33
RV	96	Pezinok	a	5
RV	98	Dvory	b	67
RV	97	Modra	a	38
VZ	96	Nitra	b	12
VZ	98	Velky Krtis	a	72
VZ	97	Pezinok	b	45
VZ	96	Dvory	a	17
VZ	98	Modra	b	79
MT	97	Nitra	a	50
MT	96	Velky Krtis	b	23
MT	98	Pezinok	a	84
MT	97	Dvory	b	57
MT	96	Modra	a	28

Network	Correct/total	Performance (%)	Eliminated variables
Classification criterion: Vintage			
$20 \times 8 \times 3$	15/15	100.0	–
$14 \times 8 \times 3$	15/15	100.0	v3 v6 v8 v10 v25 v26
Classification criterion: Producer			
$19 \times 11 \times 5$	15/15	100.0	v19
$17 \times 11 \times 5$	15/15	100.0	v12 v19 v20
Classification criterion: Variety			
$15 \times 5 \times 3$	14/15	93.3	v1 v7 v8 v16 v24
$20 \times 11 \times 3$	14/15	93.3	–

Note: 20 variables (compound signals) were used when no one was eliminated.

Table 6
Classification performance by the sample category re-classification of the training data set (discrimination model calculation), in the leave-one-out validation and using the test data set

Way of classification	Software	KNN[a]	QDA	LDA
Year				
Model (72 samples)	SPSS	–	97.2	95.8
	SAS	94.4	100.0	95.8
Leave-1-out (87 samples)	SPSS	–	–	90.8
	SAS	88.5	89.7	90.8
Test set (15 samples)	SPSS	–	93.3	93.3
	SAS	86.7	93.3	86.7
Producers				
Model (72 samples)	SPSS	–	98.6	93.1
	SAS	90.3	100.0	93.1
Leave-1-out (87 samples)	SPSS	–	–	83.9
	SAS	87.4	87.4	78.2
Test set (15 samples)	SPSS	–	93.3	93.3
	SAS	86.7	93.3	93.3
Variety				
Model (72 samples)	SPSS	–	81.9	84.7
	SAS	81.9	100.0	84.7
Leave-1-out (87 samples)	SPSS	–	–	65.5
	SAS	62.1	89.7	65.5
Test set (15 samples)	SPSS	–	53.3	53.3
	SAS	66.7	73.3	53.3

[a] Five neighbours were used.

set of 15 samples, where each variety, producer and vintage were equally represented, as shown in the left part of Table 5.

The ANN performance was calculated for all three classification criteria, as shown in the right part of Table 5 together with the optimised network and the optimised set of variables selected by Intelligent Problem Solver in Trajan software. It is noteworthy that the use of a complete set of variables is less successful and elimination of insignificant variables is very useful. Under the mentioned conditions the classification of 15 test samples was very good; a 100% success was obtained for *vintage* and *producer* and 93.3% for *variety*.

Confrontation of the ANN results with three methods of discriminant analysis is possible by comparing the results shown in Tables 5 and 6. Two parametric methods – linear (LDA) and quadratic (QDA) discriminant analyses were used there together with the nonparametric *K*th nearest neighbour method (KNN). It is obvious that the ANN results are the best even though also LDA, QDA and KNN results are very good in particular cases except classification by *variety*.

4. Conclusions

It has been demonstrated that artificial neural network can be a very useful tool for the classification of wines. It has been found that the success of classification increases when instead of original 19 or 20 variables a smaller set of optimally selected variables is used. This fact is explainable by improving the signal-to-noise ratio and reducing the effect of overfitting in the ANN model. The optimal variable selection was performed by the backward selection procedure. The selected white wines of different *vintage* can be classified with the prediction success of 100%. For the criterion *producer* the classification performance of 97.2% was found for the first data set by leave-one-out validation and 100% for the second data set by using 15 test data. For the criterion *variety* the classification performance of 100% was found for the first data set by leave-one-out validation and 93.3% for the second data set by using 15 test data. The ANOVA results, independent of the variable correlations, can be complementarily used with the ANN results. The ranks of the most influential variables for all performed kinds of classification by both techniques were in very good accordance. The most important compounds for white wine classes differentiation are (1) α-terpineol for the variety classification (first rank in the ANN as well as ANOVA), (2) diethyl succinate (first rank in the ANN and ANOVA) for the classification by producer, (3) 1-hexanol (the second in the ANN and the first in ANOVA) and (E)-3-hexen-1-ol (the first in the ANN and the third in ANOVA) for the classification by vintage.

Acknowledgements

This work was supported by the Slovak Grant Agency – Project VEGA No. 1/3584/06 and Project APVV-0057-06.

References

Almela, L., Javaloy, S., Fernandez-Lopez, J. A., & Lopez-Roca, J. M. (1996). Varietal classification of young red wines in terms of chemical and colour parameters. *Journal of the Science of Food and Agriculture, 70*, 173–180.

Ball, G., Mian, S., Holding, F., Allibone, R. O., Lowe, J., & Ali, S. , et al. (2002). An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics, 18*, 395–404.

Ballabio, D., Mauri, A., Todeschini, R., & Buratti, S. (2006). Geographical classification of wine and olive oil by means of classification and influence matrix analysis (CAIMAN). *Analytica Chimica Acta, 570*, 249–258.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford: University Press.

Bocaz-Beneventi, G., Latorre, R., Farková, M., & Havel, J. (2002). Artificial neural networks for quantification in unresolved capillary electrophoresis peaks. *Analytica Chimica Acta, 452*, 47–63.

Capron, X., Smeyers-Verbeke, J., & Massart, D. L. (2006). Multivariate determination of the geographical origin of wines from four different countries. *Food Chemistry, 225*, 559–568.

Cliff, M. A., & Dever, M. C. (1996). Sensory and compositional profiles of British Columbia Chardonnay and Pinot noir wines. *Food Research International, 29*, 317–323.

De la Calle Garcia, D., Reichenbaecher, M., & Danzer, K. (1998). Klassifizierung von Weinen mittels multivariater Datenanalyse anhand der SPME/CGC-Chromatogramme von Aromastoffen. *Vitis, 37*, 181–188.

De la Presa-Owens, C., & Noble, A. C. (1995). Descriptive analysis of 3 white wine varieties from penedes. *American Journal of Enology and Viticulture, 46*, 5–9.

Dohnal, V., Farková, M., & Havel, J. (1999). Prediction of chiral separations using combination of experimental designs and artificial neural networks. *Chirality, 11*, 616–621.

Dohnal, V., Li, H., Farková, M., & Havel, J. (2002). Quantification of unresolved peaks of chiral compounds using artificial neural networks. *Chirality, 14*, 509–518.

Dumont, A., & Dulau, L. (1996). The role of yeasts on the formation of wine flavors. In T. Henick-Kling, T. E. Wolf, & E. M. Harkness (Eds.), *Proceedings of the 4th international symposium on cool climate viticulture and enology* (pp. VI-24–VI-28). Rochester, NY: State Agricultural Experimental Station.

Etiévant, P., Schlich, P., Bouvier, J. C., Symonds, P., & Bertrand, A. (1988). Varietal and geographic classification of French red wines in terms of elements, amino acids and aromatic alcohols. *Journal of the Science of Food and Agriculture, 45*, 25–41.

Etiévant, P., Schlich, P., Cantagrel, R., Bertrand, A., & Bouvier, J. C. (1989). Varietal and geographic classification of French red wines in terms of major acids. *Journal of the Science of Food and Agriculture, 46*, 421–438.

Fahlman, S. E. (1988). Faster-learning variations on back-propagation: an empirical study. In D. Touretzky, G. E. Hilton, & T. J. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 38–51). San Mateo, CA: Morgan Kaufman.

Farková, M., Peña-Mendez, E. M., & Havel, J. (1999). The use of artificial neural networks in capillary zone electrophoresis. *Journal of Chromatography A, 848*, 365–374.

Ferreira, V., Fernandez, P., & Cacho, J. F. (1996). A study of factors affecting wine volatile composition and its application in discriminant analysis. *Food Science and Technology, 29*, 251–259.

Fidencio, P. H., Ruisanchez, I., & Poppi, R. J. (2001). Application of artificial neural networks to the classification of soils from Sao Paulo state using near-infrared spectroscopy. *Analyst, 126*, 2194–2200.

García, M., Fernández, M. J., Fontecha, J. L., Lozano, J., Santos, J. P., Aleixandre, M., et al. (2006). Differentiation of red wines using an electronic nose based on surface acoustic wave devices. *Talanta, 68*, 1162–1165.

Gasteiger, J., & Zupan, J. (1993). Neural networks in chemistry. *Angewandte Chemie, 32*, 503–527.

Gonzalez, G., Mendez, E. M. P., Sanchez, M. J. S., & Havel, J. (2000). Data evaluation for soft drink quality control using principal component analysis and back-propagation neural networks. *Journal of Food Protection, 63*, 1719–1724.

Havel, J., Lubal, P., & Farková, M. (2002). Evaluation of chemical equilibria with the use of artificial neural networks. *Polyhedron, 21*, 1375–1384.

Havel, J., Madden, J. E., & Haddad, P. R. (1999). Prediction of retention times for anions in ion chromatography using artificial neural networks. *Chromatographia, 49*, 481–488.

Havel, J., Peña, E. M., Rojas-Hernández, A., Doucet, J.-P., & Panaye, A. (1998). Neural networks for optimization of high-performance capillary zone electrophoresis methods. A new method using a combination of experimental design and artificial neural networks. *Journal of Chromatography, 793*, 317–329.

Havliš, J., Madden, J. E., Revilla, A. L., & Havel, J. (2001). High-performance liquid chromatographic determination of deoxycytidine monophosphate and methyldeoxy-cytidine monophosphate for DNA demethylation monitoring: Experimental design and artificial neural networks optimisation. *Journal of Chromatography B, 755*, 185–194.

Jennings, W., & Shibamoto, T. (1980). *Qualitative analysis of flavour and fragrance volatile by glass capillary gas chromatography*. New York: Academic Press.

Khattree, R., & Naik, D. N. (2000). *Multivariate data reduction and discrimination*. Cary, North Carolina: SAS Institute.

Kondjoyan, N., & Berdagué, J. L. (1996). A compilation of relative retention indices for the analysis of aromatic compounds. INRA de THEIX, Genes Campanelle.

Kvasnička, V., Beňušková, Ľ., Pospíchal, J., Farkaš, I., Tiňo, P., & Kráľ, A. (1997). *Introduction to theory of neural networks*. Bratislava: IRIS Publisher [in Slovak].

Larice, J.-L., Archier, P., Rocheville-Divorne, C., Coen, S., & Roggero, J.-P. (1989). anthocyanique des cépages. II. Essai de classification sur trois ans par analyse en composantes principales et étude des variations annuelles de cépages de meme provenance. *Revue Française d'Oenologie, 29*, 7–12.

Lozano, J., Fernández, M. J., Fontecha, J. L., Aleixandre, M., Santos, J. P., Sayago, I., et al. (2006). Wine classification with a zinc oxide SAW sensor array. *Sensors and Actuators B, 120*, 166–171.

Lozano, J., Santos, J. P., & Horrillo, M. C. (2005). Classification of white wine aromas with an electronic nose. *Talanta, 67*, 610–616.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing*. Cambridge, England: A Bradford Book, The MIT Press.

Medina, B. (1996). Wine authenticity. In P. R. Ashurst & M. J. Dennis (Eds.), *Food authentication* (pp. 60–107). London: Chapman & Hall.

Moret, I., Scarponi, G., Capodaglio, G., Zanin, S., Camaiani, G., & Toniolo, A. (1980). Analytical parameters in the characterization of three Venetian wines. Application of the statistical linear discriminant analysis. *American Journal of Enology and Viticulture, 31*, 245–249.

Noble, A. C., & Shannon, M. (1987). Profiling Zinfandel wines by sensory and chemical analyses. *American Journal of Enology and Viticulture, 38*, 1–5.

Otto, M. (1999). *Chemometrics*. Weinheim: Wiley.

Petka, J., Farkas, P., Kovac, M., Balla, B., & Mocak, J. (2000). Classification of selected Slovak varietal wines. *Kvasny Prumysl, 46*, 243–246 (in Slovak).

Petka, J., Mocak, J., Farkas, P., Balla, B., Sadecka, J., & Kovac, M. (1999). Utilisation of multi-variate statistical methods for authentication of Slovak varietal wines. In R. Lásztity & W. Pfannhauser (Eds.). *Proceedings EURO FOOD CHEM X – Functional foods* (Vol. 3, pp. 956–964). Budapest: FECS Food Chemistry Division.

Pokorná, L., Revilla, A. L., Havel, J., & Patočka, J. (1999). Capillary zone electrophoresis determination of galanthamine in biological fluids and pharmaceutical preparatives: Experimental design and artificial neural network optimization. *Electrophoresis, 20*, 1993–1997.

Pueyo, E., Dizy, M., & Polo, M. C. (1993). Varietal differentiation of must and wines by means of protein fraction. *American Journal of Enology and Viticulture, 44*, 255–260.

Rapp, A., Guentert, M., & Heimann, W. (1985). Beitrag zur Sortencharakterisierung der Rebsorte Weisser Riesling. I. Untersuchung der Aromastoffzusammensetzung von ausländischen Weissweinen mit der Sortenbezeichnung ‘‘Riesling”. *Zeitschrift Lebensmittel Untersuchung Forschung, 181*, 357–361.

Rapp, A., & Guentert, M. (1985). Beitrag zur Charakterisierung des Weines der Rebsorte Weisser Riesling. II. Untersuchung der Aromastoffzusammensetzung deutscher Weissweine der Rebsorten Weisser Riesling, Müller-Thurgau und Silvaner. *Vitis, 24*, 139–150.

Rapp, A., Suckrau, I., & Versini, G. (1993). Untersuchungen des Trauben- und Weinaromas. Beitrag zur Sortencharakterisierung neutraler Rebsorten (Silvaner, Weissburgunder, Rulaender). *Zeitschrift Lebensmittel Untersuchung Forschung, 197*, 249–254.

Sanni, O. D., Wagner, M. S., Briggs, D., Castner, D. G., & Vickerman, J. C. (2002). Classification of adsorbed protein static ToF-SIMS spectra by principal component analysis and neural networks. *Surface and Interface Analysis, 33*, 715–728.

Trajan Software (1999). *Trajan neural network simulator, Release 4.0 D*. Durham, UK: Trajan Software Ltd..

Vasconcelos, A. M. P., & Chaves das Neves, H. J. (1989). Characterization of elementary wines of *Vitis vinifera* variety by pattern recognition of free amino acid profiles. *Journal of Agricultural and Food Chemistry, 37*, 931–937.