Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Talanta 72 (2007) 263–268
Talanta
www.elsevier.com/locate/talanta

Differentiation of two Canary DO red wines according to their metal content from inductively coupled plasma optical emission spectrometry and graphite furnace atomic absorption spectrometry by using Probabilistic Neural Networks

Isabel M. Moreno [a], Dailos González-Weller [b], Valerio Gutierrez [c], Marino Marino [d], Ana M. Cameán [a], A. Gustavo González [e,*], Arturo Hardisson [b]

[a] *Nutrition, Food Chemistry and Toxicology, Faculty of Pharmacy, University of Sevilla, C/Profesor García González 2, 41012 Sevilla, Spain*
[b] *Obstetrics, Gynaecology, Paediatrics, Preventive Medicine and Public Health, Toxicology and Legal Medicine, Faculty of Medicine, University of La Laguna, Campus de Ofra s/n, 38071 La Laguna, Tenerife, Spain*
[c] *Chemistry Engineering and Pharmacist Technology, Faculty of Chemistry, University of La Laguna, C/Astrofísico Francisco Sánchez s/n, 38071 La Laguna, Tenerife, Spain*
[d] *Laboratory of Public Health, C/Rambla General Franco 53, 38006 Santa Cruz de Tenerife, Tenerife, Spain*
[e] *Analytical Chemistry, Faculty of Chemistry, University of Sevilla, C/Profesor García González no. 1, 41012 Sevilla, Spain*

Received 3 July 2006; received in revised form 9 October 2006; accepted 19 October 2006
Available online 28 November 2006

Abstract

The metal content of 54 commercialized wines (30 samples from Tacoronte-Acentejo DO (class T) and 24 Valle de la Orotava DO (class O) wines) was performed by ICP-OES (Al, Ba, Cu, Fe, Mn, Sr, Zn, Ca, K, Na and Mg) and GF-AAS (Ni and Pb). Wine samples were processed by dry ashing followed by solution with 5% nitric acid. Metals were considered as suitable descriptors to differentiate between T and O classes. Supervised learning pattern recognition procedures were applied. Linear discriminant analysis (LDA) led to good results up to about 90% of correct classification. In order to improve the results, another kind of algorithms able to model non-linear separation between classes was considered: Probabilistic Neural Networks. Accordingly, excellent results were obtained, leading to sensitivities and specificities higher than 95% for the two classes.

Keywords: Wine differentiation; Linear discriminant analysis; Probabilistic Neural Networks; Canary island

1. Introduction

Wine is a widely consumed beverage in the world and has an obvious commercial value and social importance. Therefore, the evaluation of the quality of wines is important for manufacturers, merchants and consumers. Denomination of origin (DO) have been created in Spain, the main purpose of whose Regulating Councils is to guarantee the source and quality of the wines to prevent fraud and to guarantee origin [1]. To attain it, a by far useful tool is the knowledge of the values of several chemical descriptors that allow a link to be established between the wine samples and their geographical origin.

The quantity of inorganic ions in wine is of great interest, because of their influence on wine technology as well as their toxic effects [2]. However, one of the main interests is to use the mineral content to characterise the wines by their geographical origin taking into account the relationship between the metallic content in samples and soil composition [3]. This differentiation can be carried out by using major, trace and ultra-trace elements [4]. Climatic conditions influence on necessary fungicide treatments of wines that determine the level of copper in grapes, and edaphical features also have influence on metal level such as Na, Ca, K and Fe of grapes [5]. The two major sources of Pb in wine are soil-related and anthropogenic: winery equipment, vine treatment, atmospheric aerosols rich in industrial fumes

* Corresponding author. Tel.: +34 954557173; fax: +34 954557168.
E-mail address: agonzale@us.es (A.G. González).

0039-9140/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2006.10.029

and exhaust gases. Because of government polices to reduce total Pb emissions and replace old winery equipment, the Pb content in wine is gradually decreasing. The Organisation Internationale de la Vigne et du Vin (OIV) limits Pb amounts in wine to 200 μg L^{-1} [6].

Metal elements have been widely used for differentiating Spanish wines according to geographical origin [5,7–13]. In the present work, two Canary DO were studied: Tacoronte-Acentejo and Valle de la Orotava. Particularities of wines from the Canary Island are that they are elaborated with autochthonous varieties that cannot be found elsewhere in the world, the vines are not contaminated by phylloxera and soils are volcanic [14]. With respect to these two growing areas, the production region of wines protected by the Tacoronte-Acentejo DO is located at the north side of the Tenerife Island and is run through by steep mountain-clefts. The district contains nine communities: La Laguna, Tegueste, Tacoronte, El Sauzal, La Matanza de Acentejo, La Victoria de Acentejo, Santa Ursula, El Rosario and Santa Cruz de Tenerife (Anaga). The cultivated area is approximately 2422 ha, distributed among the different communities. The grape and wine-growing region which is currently protected under the DO of Valle de la Orotava comprises a region located at the north side of the Tenerife Island. At present, the planted area inscribed of this DO is 671 ha and includes the cities of La Orotava, Los Realejos and El Puerto de la Cruz.

Tacoronte-Acentejo DO uses grapes belonging to Canary variety Listán Negro and Negramoll. The hills and valleys of this district are covered by grapes, which reach from the sea until almost 1000 m. This area is the biggest wine-district and the one with the most intensive cultivation of the Canary Islands. It represents 40% of the wine cultivation area of the island of Tenerife and 20% of the area of the Canary archipelago. The cultivation is situated at hillsides, which are orientated to the north and the sea. The grounds have a red colour and are rich of organic earth, deficient in lime, but rich in nitrogen, phosphorus and potassium.

El Valle de la Orotava DO employs grapes belonging to the one of the autochthonous varieties above mentioned (Listán Negro). The vines of this district reach from the sea since 250–700 m. The cultivation is situated in a vast valley which is making one's way at the foot of the mountain Teide and spreading out until the sea, showing very different hues according to the distinct terrains and unevenness, vegetation and farming.

Typical vineyard soil is permeable, rich in minerals nutrients and with a slightly acid pH because of its volcanic origin.

The north coast of the island is protected from the African winds and influenced by the September winds and the influence of the altitude causes a mild and pleasant climate which is steady and has almost no notable thermal differences. The yearly temperature is always above 7 °C. Although the sky is often cloudy (with an atmospheric humidity of 60%) the rainfalls are few, because the low temperature holds up the condensation of the humidity.

Accordingly, soil differences and, in minor extent, differences in grape varieties could be reflected in the metal content of the two DO wines. Consequently, the aim of the present paper is the determination of the content of Al, Ba, Cu, Fe, Mn, Sr, Zn, Ca, K, Na, and Mg by using inductively coupled plasma atomic emission spectrometry (ICP-OES) and Ni and Pb measured by graphite furnace atomic absorption spectrometry (GFAAS) in red wines to differentiate between the two DO mentioned (Tacoronte-Acentejo and Valle de la Orotava).

For classification and discrimination between wine samples coming from these two producing regions, supervised pattern recognition techniques such as linear discriminant analysis (LDA) [15] and Probabilistic Neural Networks (PNN) [16] were applied.

Table 1
Analytical lines used for each element and the instrumental conditions in the ICP-OES apparatus

Parameter	
RF frequency (MHz)	27.12
Operating power (W)	1350
Coolant Ar flow (L min^{-1})	20
Plasma Ar flow (L min^{-1})	0.6
Carrier Ar flow (L min^{-1})	0.5
Nebulizer type	Babington V
Detection wavelengths (nm^{-1})	
Al	396.152
Ba	233.527
Ca	393.366
Fe	259.940
K	766.490
Mg	280.270
Mn	257.610
Na	589.592
Sr	346.446
Zn	213.856
Cu	324.754

2. Experimental

2.1. Apparatus

Elemental analyses were carried out on a Thermo Jarrell Ash Atom Scan 25 inductively coupled plasma atomic emission spectrometer (Genesis Laboratory Systems Inc., CO, USA). Table 1 shows the analytical lines used for each element, as well as the instrumental conditions. Moreover, a Perkin-Elmer 4100 ZL Zeeman atomic absorption spectrometer attached to a PE HG 4100 graphite furnace (with a PE AS-70 automatic inyector, all from Perkin-Elmer, Norwalk, CT, USA) was employed for quantitation of Pb and Ni by the GFAAS technique. The graphite furnace temperature program is gathered in Table 2.

2.2. Chemical and reagents

Merck (Darmstadt, Germany) CertiPUR®. ICP-multielement standard solutions of about 1000 mg L^{-1} were used as stock solution for calibration. Other reagents where of analytical grade. Milli-Q treated water was used throughout.

Table 2
HG 4100 Graphite furnace parameters used for analysis of Ni and Pb in red wines from Tacoronte and Orotava DO

Step	Temperature (°C)		Ramp time (s)	Hold time (s)	
	Ni	Pb		Ni	Pb
1	100	100	10	10	20
2	130	130	25	20	20
3	450	450	15	20	20
4	1350	800	5	20	10
5	2350	1250	0	5[a]	3[a]
6	2550	2500	2	4	3

Argon flow rate, 250 mL min^{-1}; injection volume, 20 μL.
[a] Gas stop for reading.

2.3. Samples

Thirty samples of red wines of different brands with D.O. trademark "Tacoronte-Acentejo" (code T) and 24 samples of "El Valle de la Orotava" (code O) were purchased in liquor retails and markets. All the samples were from 2004-vintage. The alcoholic content ranged from 12 to 14% (v/v) ethanol for all wine samples.

The plastic containers used for storing and treating the samples were cleaned to avoid contamination of the samples with traces of any metal. Containers were treated with 5% nitric acid during 24 h followed with two washes with Milli-Q water.

Once opened, wines samples were digested according to the following procedure:

The 25 mL of each sample were placed in porcelain crucibles. To avoid contamination and cross-talking between the samples, single used plastic tools were utilised to transfer the material. Each sample was then dried in an oven at a temperature of 50 °C for at least 12 h. The crucibles with the samples were then introduced in muffle ovens and burned to ash at 450 °C. The temperature in the muffle oven was increased at a rate of approximately 50 °C h^{-1} and maintained at 450 °C during 18–24 h. The white ashes obtained with this procedure were then dissolved in 5% nitric acid to a volume of 25 mL.

2.4. Data analysis

The content of each mineral element was considered as chemical descriptor. Pattern recognition methods were applied to the data matrix, composed of 13 columns (the analysed elements) and 54 rows (red wines). LDA and PNN were applied for differentiation between class T and class O wines. The statistical package, STATISTICA 7 from Statsoft [17] was used for all the chemometric calculations.

3. Results and discussion

3.1. Mineral content in red wine samples

The metal content of the two different DO red wines was determined and carefully scrutinized. The results, expressed in milligram per litre (except Pb and Ni values, in microgram per litre), were obtained from triplicate measurements and rounded up to the last significant figure associated with random error. Table 3 shows these results. The corresponding descriptive basic statistic for both DO wines samples can be seen in Table 4. Looking at these values, K was the element with a major content in all samples. The mean concentration of K was similar in Tacoronte and Orotava wines samples, with average concentrations of 1363.6 and 1307.2 mg L^{-1}, respectively. Ca, Mg and Na presented lower and similar contents, their average values being 111.6, 107 and 106.6 mg L^{-1}, respectively, in Tacoronte samples and 124.9, 110.7 and 84 mg L^{-1} in Orotava ones. Nickel was also present with values higher than 10 mg L^{-1} (37.1 and 31.5 mg L^{-1} in Tacoronte and Orotava, respectively) and Al was present with values near 10 mg L^{-1}. The other analysed metals mostly appeared with values close to 1mg L^{-1}, such as Fe, Mn, Sr and Zn, and even lower, being Ba and Cu the metals with the lowest concentration in both class (T and O). All the values obtain for both class of red wines were similar with the exception of Pb which was not detected in the Orotava samples and close to 1 mg L^{-1} in Tacoronte ones.

3.2. Statistical procedures for classification

Using the mineral content found in the analysed red wines samples as chemical descriptors, statistical methods were applied in order to establish differences between both DO wines samples for classification and authentication purposes. There are a number of display methods for visualizing data trends, such as principal component analysis (PCA) [18] or cluster analysis (CA) [19]. However, our research starts from the a priori knowledge of class membership of the samples to be processed and hence, typical supervised learning pattern recognition methods have to be applied. In our case, two classes were considered, i.e. Tacoronte (T) and Orotava (O) red wines samples. First of all, the method of LDA was applied for building linear frontiers between the two classes. Because of possible non-linear nature of class distribution, Bayesian artificial neural networks such as PNN were also applied.

3.2.1. LDA

Linear discriminant analysis is a typical discriminating method, belonging to the first level of PR, where objects are classified into either of a number of defined classes [20]. Discriminant functions are obtained as linear combination of metal descriptors to maximise the F-ratio of between class sum of squares and within class sum of squares. If we have p descriptors and g classes, the number of uncorrelated discriminant functions are p or $g - 1$ whichever is smaller; and so, in our case only one discriminant function can be obtained. The parameter called Wilks' lambda is the ratio between the within class sum of squares and the total sum of squares, calculated for each descriptor. The discrimination power of a given descriptor is better when its Wilks' lambda is lesser [21,22]. Thus, after applying standard LDA and according to the Wilks' lambda, six descriptors, namely Zn, Ba, Pb, Na, Mg and Al, seem to be optimal ones for distinguishing between class

Table 3
Experimental results[a] for determination of elements ($mg\,L^{-1}$) in T and O red wines samples by ICP-OES and GFAAS[b]

Sample	Code	Pb[b]	Ni[b]	Al	Ba	Cu	Fe	Mn	Sr	Zn	Ca	K	Na	Mg
1	T	2.64	78.66	6.68	0.05	1.30	4.57	2.15	0.73	0.86	94.53	1110.25	145.75	107.88
2	T	1.00	39.94	8.63	0.06	0.12	3.34	1.77	0.48	0.64	76.45	1248.50	79.10	127.33
3	T	0.85	46.16	11.04	0.15	0.03	3.50	2.26	0.71	0.80	103.70	1335.00	84.28	119.05
4	T	1.07	33.22	12.33	0.10	0.15	3.25	1.31	0.67	0.48	102.80	1392.75	81.70	118.88
5	T	0.43	22.47	7.96	0.17	0.31	2.71	0.88	1.24	0.51	112.13	1238.00	108.63	123.75
6	T	2.23	15.24	5.19	0.11	0.53	1.14	2.58	0.81	1.04	95.65	1325.50	127.80	132.23
7	T	ND	42.12	7.31	0.14	0.90	2.99	2.79	0.87	0.78	130.73	980.75	110.38	111.48
8	T	ND	24.08	6.68	0.09	0.12	2.62	3.29	0.68	0.52	118.05	1193.25	102.00	114.05
9	T	ND	40.41	6.86	0.11	0.05	2.83	1.17	0.82	0.52	110.08	1311.50	78.78	113.53
10	T	0.17	44.60	7.02	0.14	0.10	3.60	2.74	1.09	1.01	132.83	1309.25	160.80	134.85
11	T	0.00	29.59	6.94	0.12	0.14	4.00	2.87	0.98	0.97	111.73	1430.50	96.93	123.20
12	T	1.32	34.41	7.28	0.13	1.20	3.19	1.13	0.96	0.65	100.75	1438.50	67.45	99.45
13	T	1.17	58.86	6.70	0.12	0.07	6.70	2.24	0.57	0.66	114.38	1369.50	120.15	136.33
14	T	ND	26.43	7.27	0.15	0.67	2.47	2.31	0.95	0.75	213.28	1826.50	256.03	236.15
15	T	1.07	24.95	5.60	0.14	0.13	2.32	1.67	0.94	0.78	98.85	1408.75	114.18	140.83
16	T	0.81	22.52	5.53	0.12	1.27	2.45	2.02	1.05	0.49	91.10	1474.50	96.43	121.08
17	T	1.15	27.65	5.27	0.10	0.16	2.66	2.57	0.69	0.62	98.75	1205.50	106.80	135.05
18	T	0.62	29.69	12.45	0.14	0.11	2.68	2.67	1.04	0.89	107.05	1395.75	69.28	137.90
19	T	0.83	24.75	5.87	0.10	0.25	2.43	1.63	0.74	0.66	110.45	1169.00	67.75	98.35
20	T	1.42	73.05	6.85	0.10	0.07	2.68	0.87	1.06	0.68	112.55	1322.00	89.80	122.95
21	T	1.22	71.63	5.94	0.08	0.14	2.66	4.08	1.01	0.67	114.40	1356.00	111.30	110.40
22	T	0.88	30.08	8.99	0.12	0.34	4.30	2.51	1.08	0.78	133.70	1595.00	95.45	154.55
23	T	ND	24.92	5.12	0.11	0.17	1.91	2.14	0.75	1.01	114.40	1494.00	124.25	128.10
24	T	7.36	39.76	8.23	0.16	0.22	2.55	0.82	1.12	0.78	115.90	1613.50	174.25	116.45
25	T	1.01	28.57	8.16	0.11	0.14	2.74	1.84	0.80	0.54	114.80	1377.00	75.15	109.55
26	T	0.87	32.97	9.29	0.14	0.30	3.08	0.94	0.97	0.61	98.20	1535.50	114.80	116.70
27	T	ND	39.66	7.85	0.12	0.06	3.16	1.16	1.02	0.67	150.45	1231.00	87.70	108.45
28	T	0.96	31.09	6.23	0.16	0.10	2.57	1.86	0.86	0.64	119.00	1232.50	94.70	129.25
29	T	2.16	44.06	7.66	0.18	0.07	3.29	1.42	0.73	1.03	65.40	1590.00	74.30	117.75
30	T	2.11	40.37	5.10	0.11	0.08	2.03	1.04	0.75	0.69	105.05	1451.00	93.30	112.65
31	O	ND	20.89	9.35	0.11	0.28	2.39	1.85	1.11	0.92	118.28	1224.33	68.90	115.12
32	O	ND	29.06	9.70	0.16	0.22	3.12	0.95	0.97	0.76	91.77	933.17	76.88	102.42
33	O	ND	22.95	8.73	0.15	0.19	3.00	0.99	0.97	0.81	85.78	1607.00	75.38	103.28
34	O	ND	20.42	8.69	0.12	0.06	2.57	1.61	0.82	0.45	99.65	1922.17	63.20	102.45
35	O	ND	31.90	9.00	0.10	0.13	1.93	1.65	0.75	0.69	118.23	1824.25	58.50	100.88
36	O	ND	24.58	6.45	0.09	1.38	2.66	1.89	0.58	0.85	88.85	1259.75	71.10	87.55
37	O	ND	17.86	6.53	0.12	0.10	2.63	1.08	0.50	0.73	100.95	1077.25	83.65	87.33
38	O	ND	15.92	5.15	0.08	0.05	2.73	1.18	0.45	0.91	89.40	1127.50	57.30	88.73
39	O	ND	22.76	5.94	0.05	0.04	2.50	1.95	0.28	0.10	79.73	1371.75	65.85	98.23
40	O	ND	50.87	12.73	0.11	0.37	3.49	3.43	0.88	0.75	118.73	1564.75	78.13	119.30
41	O	ND	22.06	8.63	0.10	0.63	3.00	1.63	0.85	1.01	108.10	1199.50	75.20	120.00
42	O	ND	18.19	11.27	0.11	0.45	2.36	2.10	1.10	0.71	122.23	1385.00	71.33	120.85
43	O	ND	31.09	8.40	0.12	0.04	3.68	1.82	1.22	0.67	127.40	1326.50	106.05	110.63
44	O	ND	27.48	6.09	0.08	0.17	2.93	2.18	0.88	0.77	92.50	1048.25	99.45	96.35
45	O	ND	27.20	7.25	0.11	0.14	3.65	1.72	1.05	0.87	116.25	1113.50	104.80	103.85
46	O	ND	19.69	7.26	0.09	0.21	2.80	2.53	0.83	1.07	102.30	1229.00	71.85	122.33
47	O	ND	36.35	10.30	0.24	3.09	1.82	2.82	1.44	0.00	107.65	1281.50	85.00	127.85
48	O	ND	32.41	10.22	0.20	0.18	5.22	1.61	1.80	0.50	134.40	1577.50	84.40	138.85
49	O	ND	53.51	8.98	0.19	0.14	5.38	2.33	1.31	0.78	119.25	1032.50	86.35	110.45
50	O	ND	41.13	14.34	0.26	1.15	4.69	2.85	1.43	0.13	136.00	1491.00	138.95	136.35
51	O	ND	46.52	12.77	0.28	2.73	2.29	2.36	1.49	ND	111.20	1313.50	105.90	127.20
52	O	ND	64.44	9.99	0.23	0.22	4.12	1.88	1.23	ND	109.45	1350.50	112.55	125.90
53	O	ND	33.76	7.21	0.13	0.05	3.33	1.73	0.79	0.71	105.35	1111.00	123.90	129.60
54	O	ND	35.31	6.45	0.08	0.03	3.02	0.91	0.58	0.83	97.85	732.00	96.45	92.45

ND, not detected.
[a] Mean of triplicate determinations.
[b] Pb and Ni values were expressed in $\mu g\,L^{-1}$.

T and O samples. Evaluation samples are classified from the estimation of an a posteriori probability of class membership using the Bayes' theorem. In order to evaluate the classification performance, the leave-one-out method was used as a validation procedure, obtaining a prediction ability of 90.7%. This result, albeit fairly good, could be improved by using artificial neural networks that are able to model non-linear frontiers between classes.

Table 4
Metal concentration in Tacoronte red wine samples ($n = 30$) and Villa de la Orotava samples ($n = 24$)

Element	"Tacoronte samples"			"Orotava samples"		
	Mean ± S.D. (mg L^{-1})	Median (mg L^{-1})	Range of quantified values (mg L^{-1})	Mean ± S.D. (mg L^{-1})	Median (mg L^{-1})	Range of quantified values (mg L^{-1})
Al	7.32 ± 1.91	6.94	4.37–13.31	8.79 ± 2.29	8.54	5.15–14.34
Ba	0.12 ± 0.03	0.12	0.042–0.18	0.14 ± 0.06	0.12	0.05–0.29
Ca	111.55 ± 26.79	107.83	64.00–252.45	106.99 ± 17.99	104.45	82.40–172.85
Cu	0.31 ± .37	0.14	0.06–1.34	0.47 ± 0.76	0.18	0.03–2.78
Fe	3.01 ± 0.96	2.72	1.12–6.85	3.10 ± 0.88	2.96	1.77–5.38
K	1363.47 ± 170.62	1354.50	968.00–1887.50	1307.16 ± 290.02	1297.50	732.00–2016.50
Mg	124.94 ± 25.60	120.55	96.55–280.85	110.66 ± 15.52	109.53	70.65–139.80
Mn	2.36 ± 3.29	1.99	0.81–4.01	1.82 ± 0.62	1.76	0.91–3.46
Na	106.58 ± 39.42	96.50	62.05–220.75	83.99 ± 20.29	79.00	54.45–138.95
Ni[a]	37.11 ± 16.21	33.08	14.39–103.4	31.55 ± 13.62	27.57	14.5–69.18
Pb[a]	1.16 ± 1.46	0.95	0.17–7.36	ND	ND	ND
Sr	0.88 ± .18	0.86	0.44–1.24	0.97 ± 0.36	0.95	0.27–1.84
Zn	0.74 ± .17	0.72	0.44–1.05	0.63 ± 0.33	0.71	0.04–1.08

[a] Ni and Pb values were expressed in μg L^{-1}.

3.2.2. PNN

In this paper, we propose to use Probabilistic Neural Networks (PNNs) for the classification of the two classes of wine. The PNN approach combines both Bayes theorem of conditional probability and Parzen's method for estimating the probability density functions of the random variables. Unlike other neural network training paradigms, PNNs are characterised by high training speed and their ability to produce confidence levels for their classification decision. The probabilistic neural network was developed by Donald Specht. His network architecture was first presented in two papers, *Probabilistic Neural Networks for Classification*, *Mapping or Associative Memory and Probabilistic Neural Networks*, released in 1988 and 1990, respectively. This network provides a general solution to pattern classification problems by following an approach developed in statistics, called Bayesian classifiers. Bayes theory, developed in the 1950s, takes into account the relative likelihood of events and uses a priori information to improve prediction. We will outline the PNN performance in the following.

Consider Q classes in the training set, $C_1, C_2, \ldots, C_Q$; and $n_1, n_2, \ldots, n_Q$ the corresponding number of pattern vectors that belongs to each class. The prior probability for that a given pattern vector $\mathbf{x}$ belongs to class C_K can be estimated from Laplace's rule:

$$P(C_K) = \frac{n_K}{\sum_{j=1}^{Q} n_j}$$

but if we know the components $x_1, x_2, \ldots, x_p$, this information can be added to the network and calculate the posterior probability by using the Bayes' theorem:

$$P\left(\frac{C_K}{\mathbf{x}}\right) = \frac{p(\mathbf{x}/C_K)P(C_K)}{p(\mathbf{x})}$$

P refers to the probability and p to the probability density function (PDF). Thus, $p(\mathbf{x}/C_K)$ is the conditional PDF of the pattern vector once it belongs to class C_K. $p(\mathbf{x})$ is the PDF of pattern vectors and plays the role of a normalization factor. The posterior probability $P(C_K/\mathbf{x})$ is the probability that the pattern belongs to class C_K once their components are known. This probability can be evaluated by choosing the class having the high posterior value:

$$P\left(\frac{C_K}{\mathbf{x}}\right) = \max_{j \in \{1,2,\ldots,Q\}} P\left(\frac{C_j}{\mathbf{x}}\right) \quad (1)$$

Taking into account that $p(\mathbf{x})$ is a normalization factor, we can write:

$$p\left(\frac{\mathbf{x}}{C_K}\right) P(C_K) = \max_{j \in \{1,2,\ldots,Q\}} p\left(\frac{\mathbf{x}}{C_j}\right) P(C_j) \quad (2)$$

The estimation of conditional PDF for every class can be easily done by applying the modified Parzen's estimator that is the activation function of PNN and acts as a multivariate Gaussian operator:

$$p\left(\frac{\mathbf{x}}{C_K}\right) = \frac{1}{(2\pi)^{p/2}\sigma_K^p n_K} \sum_{i=1}^{n_K} \exp\left(-\sum_{j=1}^{p}\left(\frac{x_j - x_{ij}^{(K)}}{\sigma_K}\right)^2\right) \quad (3)$$

x_j and x_{ij} are the components of any pattern vector $\mathbf{x}$ and of the pattern vector $\mathbf{x}_i^{(K)}$ belonging to class K. σ_K is the called smoothing factor, which is optimised during training.

PNN are arranged into four layers: input layer, pattern layer, summation layer and decision layer. Fig. 1 shows the corresponding architecture. The input layer is used to store the new samples of the validation set. Pattern vectors of the training set are used to optimise the smoothing factor in the training step. The pattern layer contains as many neurons as pattern vectors of the training set grouped by classes. PNN training is accomplished by simply copying each pattern in the training set to the neurons of the pattern layer. The summation layer consists of one neuron for each class and sums the outputs from all pattern neurons. This gives a measure of the posterior probability density function for each class when an input vector is processed. The decision layer consists of one neuron that search for the maximum posterior probability density function, and assign to the input vector the class with highest probability.

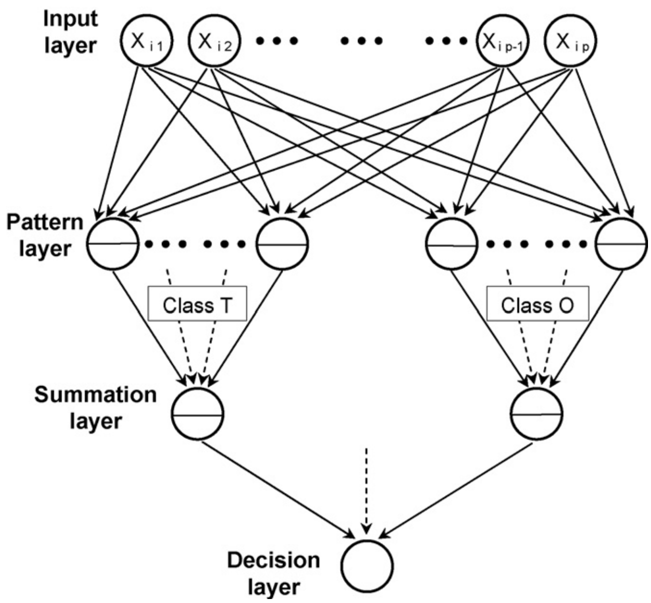

Fig. 1. Architecture of a typical PNN for two classes O and T.

For validation purposes the whole data set is then split randomly into two sets, the training and the evaluation set, each containing about 50% samples of every class. Once the classification rule is developed, some workers consider as validation parameters the recalling efficiency (rate of training samples correctly classified by the rule) and, specially, the prediction ability (rate of evaluation samples correctly classified by the rule). However, these parameters could be misleading because they do not consider the number of false positive and false negative for each class. These two concepts provide a deep knowledge of the classes' space. Accordingly, it seems to be more advisable the use of terms sensitivity (SENS) and specificity (SPEC) for validating the decision rule [23]. The SENS of a class corresponds to the rate of evaluation objects belonging to the class that are correctly classified, and the SPEC of a class corresponds to the rate of evaluation objects not belonging to the class that are correctly considered as belonging to the other classes. This may be explained in terms of the first and second kind of risks associated with prediction. The first kind of errors (α) corresponds to the probability of erroneously reject a member of the class as a non-member (rate of false negative, FN). The second kind of errors (β) corresponds to the probability of erroneously classify a non-member of the class as a member (rate of false positive, FP). Accordingly, for a given class A, and setting n_A as the number of members of class A, $\bar{n}_\mathrm{A}$ as the number of non-members of class A, $\langle n_\mathrm{A}\rangle$ as the number of members of class A correctly classified as "belonging to class A" and $\langle \bar{n}_A\rangle$ as the number of non-members of class A classified as "not belonging to class A", we have:

$$\mathrm{SENS} = \frac{\langle n_\mathrm{A}\rangle}{n_\mathrm{A}} = 1 - \alpha = 1 - \frac{\mathrm{FN_A}}{n_\mathrm{A}},$$
$$\mathrm{SPEC} = \frac{\langle \bar{n}_\mathrm{A}\rangle}{\bar{n}_\mathrm{A}} = 1 - \beta = 1 - \frac{\mathrm{FP_A}}{\bar{n}_\mathrm{A}}$$

In our case, we have 13 input neurons (one for any metal descriptor), 54 pattern neurons (30 for T class and 24 for O class), 2 summation neurons (the two classes) and 1 decision neuron (the class winner). In this case, and dealing with two classes, it is clear that $\mathrm{FP_O} = \mathrm{FN_T}$ and $\mathrm{FP_T} = \mathrm{FN_O}$.

By suitably adjusting the smoothing factor minimising the training error, we obtain in average FP = FN = 1 for the two classes. Thus, class O has SENS = 95.8% and SPEC = 96.7% and class T, conversely, SENS = 96.7% and SPEC = 95.8%. These findings fit very well with the assumption of two classes corresponding to the investigated DO.

4. Conclusion

For differentiation of the considered samples of the Tacoronte and Orotava DO wines, PNN algorithms give excellent results, leading to specificities and sensitivities higher than 95%.

References

[1] S. Pérez-Magariño, M. Ortega-Heras, M.L. González-San José, Z. Boger, Talanta 62 (2004) 983.
[2] G. González Hernández, A. Hardisson de la Torre, Arias LeónF J.J., Z Lebensm Unters Forsch 203 (1996) 517.
[3] S. Frías, J.E. Conde, M.A. Rodríguez, V. Dohnal, J.P. Pérez-Trujillo, Nahrung/Food 46 (2002) 370.
[4] B. Medina, in: P.R. Ashurst, M.J. Dennis (Eds.), Wine Authenticity, Food Authentication, Lonndon, UK, 1996, pp. 60–107.
[5] M. Ortega, M.L. Gonzalez-San Jose, S. Beltran, Quim. Anal. 18 (1999) 127.
[6] M. Barbaste, B. Medina, J.P. Pérez-Trujillo, Food Addit. Contam. 20 (2003) 141.
[7] M.S. Larrechi, J. Guasch, F.X. Rius, A. Solé, Vitis 26 (1987) 172.
[8] M. Gonzales-Larraina, A. Gonzales, B. Medina, Connai. Vigne. Vin. 21 (1987) 127.
[9] C. Herrero-Latorre, B. Medina, Connais. Vigne. Vin. 24 (1990) 147.
[10] M.J. Latorre, C. Garcia-Jares, B. Medina, C. Herrero, J. Agric. Food Chem. 42 (1994) 1451.
[11] V. Lizama, J.L. Aleixandre, I. Alvarez, M.J. García, Riv. Vitic. Enol. 50 (1997) 29.
[12] S. Frías, J.P. Perez -Trujillo, E.M. Peña, J.E. Conde, Eur. Food Res. Technol. 213 (2001) 145.
[13] A. Jos, I. Moreno, A.G. González, G. Repetto, A.M. Cameán, Talanta 63 (2004) 382.
[14] J.P. Pérez-Trujillo, M. Barbaste, B. Medina, Anal. Lett. 36 (2003) 679.
[15] D. Coomans, D.L. Massart, L. Kaufman, O. Anal, Chim. Acta 112 (1979) 97.
[16] D. Sprecht, Neural Netw. 3 (1990) 109.
[17] StatSoft Inc., STATISTICA for Windows (Computer Program Manual), Tulsa, OK, 2005.
[18] I.T. Jollife, Principal Component Analysis, Springer-Verlag, Berlín, 1986.
[19] N. Bratchell, Chemom. Int. Lab. Sys. 6 (1989) 105.
[20] C. Albano, W. Dunn, U. Edlund, E. Johansson, B. Nordén, M. Sjöström, S. Wold, Anal. Chim. Acta 103 (1978) 429.
[21] D. González-Arjona, A.G. González, Anal. Chim. Acta 363 (1998) 89.
[22] A. Alcázar, F. Pablos, M.J. Martín, A.G. González, Talanta 57 (2002) 45.
[23] D. González-Arjona, G. López-Pérez, A.G. González, Chemom. Intell. Lab Sys. 57 (2001) 133.