Vitis **48** (2), 97–100 (2009)

Creating an artificial wine taster: Inferring the influence of must and yeast from the aroma profile of wines using artificial intelligence

M. Tiefenbrunner[1)], H. Gangl[2)], G. Tscheik[2)] and W. Tiefenbrunner[2)]

[1)] Logistic Management Service, München, Germany
[2)] Bundesamt für Weinbau, Eisenstadt, Austria

Summary

The human brain is able to compute information from very complex olfactorical impressions. The special pattern of the concentrations of hundreds of aroma constituents allows an experienced wine taster to determine special features of the wine, for instance grape variety or vintage.

Artificial Neural Networks are often used to recognize shapes and patterns like faces or finger prints. Here we use Artificial Neural Networks to mimic the abilities of a wine taster to deal with very complex olfactorical patterns. We produced 120 unique wines combining twelve different grape musts and ten yeast strains and determined the aroma profile (83 aroma constituents) of all wines. We analyzed the ability of a well trained neural network to recognize the used must variety and the fermenting yeast strain from unknown aroma profiles. Furthermore we investigated the capability to predict the aroma profile of a wine with a must variety/yeast strain combination that is new to the neural network.

In 96 % of all trials the neural network identified the must that was used for wine production correctly (expected random propability: 8 %). An accurate identification of the yeast strain, used for fermentation, occurred in 67 % of all trials (propability by chance: 10 %).

The aroma profiles of the must/yeast combinations new to the neural network were forecasted with a divergence of only 2.1 % compared to the actual wine of this production characterization. Thus we conclude that a comprehensive description of wines using neural networks is possible.

K e y w o r d s : artificial neural network, artificial neuronal network, wine, wine taster, grape must, yeast, aroma profile.

Introduction

Artificial Neural Networks (ANNs) were invented to mimic the extraordinary ability of the brain to recognize shapes and patterns. They are already widely used e.g. for identification of finger prints, faces, signatures (Marinai *et al.* 2005), forecasting in medical diagnosis (Stephan *et al*. 2005) and economy (Pino *et al.* 2008) and to predict human behaviour (Marchiori *et al.* 2008). An ANN has the ability to learn and to abstract (Rumelhart 1986), which are very important skills if the abstraction rules of the process of recognition are unknown to the user.

In this study ANNs are used to mimic the abilities of the human brain to compute information from very complex olfactorical impressions. The pattern of the concentrations of hundreds of aroma constituents allows an experienced wine taster to determine special features of the wine, e.g. grape variety or vintage. We analyzed whether a well trained neural network has a similar or even superior ability (1) to recognize the used grape must variety. The grape gives the basic constituents (Cole and Noble 1997, Lambrechts and Pretorius 2000, Fleet 2003) for the resulting wine and hence it certainly influences the taste in a characteristic way; (2) to perceive the fermenting yeast strain from wine aroma profiles. This implies that each yeast strain has a typical mode to change the aroma composition of the fermentation product. During alcoholic fermentation, the yeast uses grape juice components to create a lot of aroma constituents (acids, alcohols, esters, terpenes and others) and hence the yeast is by no means less important to the flavour and taste of the wine than the grape, although its influence is not a primary one; (3) to predict the aroma profile of a wine with an unknown must variety/yeast strain combination. This presupposes that the production of a wine out of a specifical must using a special yeast strain is a repeatable process.

Material and Methods

Beside the used must and the fermenting yeast, the technique of wine making and the way, how and how long the wine is stored, determines its aroma. Therefore the production of wines took place under laboratory conditions. 120 unique wines combining twelve different grape musts and ten yeast strains were created, furthermore ten wines that were replicas from the ones that were produced with the grape variety 'Grüner Veltliner' and all yeast strains were made. The aroma profile (83 aroma constituents) of all wines was determined using Headspace-SPME-GCMS (Pawliszyn 2000).

Twelve varieties of grape juice, harvested in 2005 and 2006, were used cleared and pasteurized for vinification. All juices - with one exception - were from *Vitis vinifera*. One must originated from a red hybrid of different *Vitis*

Correspondence to: Dr. W. Tiefenbrunner, Bundesamt für Weinbau, Gölbeszeile 1, 7000 Eisenstadt, Austria. E-mail: w.tiefenbrunner@bawb.at

species ('Ripatella'). 25 aroma constituents of all musts were analyzed, not the same as for wines, because different components are frequent in musts and wines. Ten different yeast strains were used in this experiment, seven strains of *Saccharomyces cerevisiae*, one of *S. bayanus* var. *uvarum*, one hybrid with the parental species *S. cerevisiae* and *S. kudriavzevii*, and one where the species is unknown. With one exception the yeasts are filed and deposited at the ACBR culture collection of the University of Natural Resources and Applied Life Sciences in Vienna. Isolation was done during a monitoring in 2003 and 2004.

A three layer Feedforeward-Backpropagation-ANN was developed using Borland Developer Studio 2006, written in Delphi Pascal (Borland International, Scott's Valley CA, USA).

Recognition of must variety: The identification of the must that had been used to produce a wine is one of the abilities an experienced taster should have to a certain extent. To teach must recognition out of the aroma composition of the wines to the ANN, the maximum possible learn set of 129 wines was used and the last wine remained unknown to the neural network and represented the singleton test set. Starting always with untold ANNs, this procedure was repeated 130 times, so that each wine was utilized once to find out, whether a correct assignement occurred (leaving-one-out or jackknife method).

In 93 % the assignement was correct (121 out of 130 cases). By pure chance, one would appreciate a mean success rate of 8 %. Thus we can say that the ANN is a very good taster under ideal conditions and is able to learn the general abstraction rules of the influence of must on wine aroma.

Most erroneous identifications (Tab. 1) occurred concerning only two wines, Weissburgunder (column 09), that was confused with 'Grüner Veltliner II' (row 04), 'Welschriesling II' (row 06) or 'Rheinriesling' (row 10); and 'Zweigelt' (column 11), that was (surprisingly) confused with 'Welschriesling II' or 'Rheinriesling'. 'Rheinriesling' was the most frequent incorrect output, but on the other hand all wines that were actually 'Rheinriesling' were correctly identified.

Recognition of the yeast strain: The identification of the yeast strain being used for fermentation out of the aroma profile of a wine is certainly more difficult than must recognition. Not the basic constituents but fundamental metabolic pathways must be detected. The conditions and analysis methods (jackknife method) of this experiment were similar to the one for recognizing the must. In this case we appreciate an identification rate of 10 % by pure chance.

In 87 out of 130 cases the assignement was correct, which gives an accurate identification of 67 %. The identification ability of the ANN differed depending on the yeast strain (Tab. 2):

HA 2245, a *S. cerevisiae* strain isolated in South Styria and HA 2139, a *S. bayanus* var. *uvarum* strain from the north of Burgenland (index 9 and 4 from Tab. 2) were always correctly identified. HA 2198 (index 7), *S. cerevisiae* from South Styria and HA 1836 a hybrid (*S. cerevisiae* x *S. kudriavzevii*, index 1), isolated in Lower Austria were only once erroneously identified and HA 1919 (*S. cerevisiae* strain from the north of Burgenland, index 2) only three times. Most confusions occured within the yeasts of the vinegrowing region Neusiedlersee-Hügelland (indexes 2-6), especially the ones isolated in St. Georgen (indexes 4-6). In contrast no within confusion occured in yeasts isolated in the Thermenregion (indexes 0 and 1) and only one in the yeasts of South Styria (indexes 7-9). Yeasts of different winegrowing regions were frequently confused too.

Forecasting of the wine aroma: The assumption of forecasting - that the production of a wine out of a specific must using a special yeast strain is a repeatable process - was checked using the 10 'Grüner Veltliner I' (GV I) wines that were produced in two replicas. Comparison of the euklidean distances (all 83 aroma constituents were considered) of all wines were computed. As expected the two replicas of GV I produced by the same yeast strain have the most similar aroma composition on average. Pairs of GV I with other wines created by the same yeast in mean have twice the distance. 'Welschriesling' is most similar to GV I, 'Ripatella' and especially 'Muskat Ottonel' is least similar.

Table 1

Recognition matrix concerning the must identification experiment. Grape must recognition of the ANN (rows) versus actual must identity (columns). The diagonal represents correct recognitions (93 %)

Grape must		00	01	02	03	04	05	06	07	08	09	10	11		%
Grüner Veltliner I	00	20	1											21	105
Welschriesling I	01		9											9	90
Müller Thurgau	02			10										10	100
Blauburger	03				10									10	100
Grüner Veltliner II	04					10					2			12	120
Ripatella	05						10							10	100
Welschriesling II	06							10			1		1	12	120
Muskat Ottonel	07								10					10	100
Bouvier	08									9				9	90
Weissburgunder	09										6			6	60
Rheinriesling	10									1	1	10	2	14	140
Zweigelt	11												7	7	70
		20	10	10	10	10	10	10	10	10	10	10	10		

Table 2

Recognition matrix concerning the yeast strain identification experiment. Yeast recognition of the ANN (rows) versus actual yeast identity (columns). The columns in the diagonal represent correct identifications (67 %)

Yeast ID	Species		00	01	02	03	04	05	06	07	08	09		%
HA 1834	*S. cerevisiae*	00	4			1			2		1		8	62
HA 1836	*S. cerevisiae* x *S. kudriavzevii*	01		12	1	1		2		1	1		18	138
HA 1919	*S. cerevisiae*	02	2		10	2							14	108
HA 1863 - HA 1864	*S. cerevisiae*	03	2		1	6		1			1		11	85
HA 2139	*S. bayanus* var. *uvarum*	04	1	1		1	13	1	3		3		23	177
HA 2170	*S. cerevisiae*	05	1		1	2		8	2				14	108
n. i.	Not identified	06	1						3				4	31
HA 2198	*S. cerevisiae*	07								12			12	92
HA 2195	*S. cerevisiae*	08						1			6		7	54
HA 2245	*S. cerevisiae*	09	2						3		1	13	19	146
			13	13	13	13	13	13	13	13	13	13		

For prediction of wine aroma a learn set of 129 wines was used. The remaining wine and hence a combination of must and fermenting yeast strain unknown to the neural network was utilized as singleton test set. The jackknife method was performed.

Compared with the whole range of the aroma constituents within the 130 wines, the difference between original and forcasted wine was only 2.1 % (Figure, a) on average (per wine and aroma compound). The wines that were predicted worst are the ones originating from 'Muskat Ottonell' must. Here the mean error was 4.1 %. Especially the forecasting of terpenes was erroneous in these wines. This is of course understandable, because the concentration of terpenes in these wines are extremely high compared to the others. In 'Blauburger' wines bad forecasting concerned mainly alcohols, especially methanol and benzylalcohol. These two alcohols were found in relatively high concentrations in all 'Blauburger' wines created in this experiment. In 'Weissburgunder' the wine produced of HA 2245, *S. cerevisiae* of South Styria, was predicted with low quality. The same is true for the GV II wines produced by HA 1836, the hybrid strain from the vinegrowing region Thermenregion and, to a minor extent, HA 1919, *S. cerevisiae* from Neusiedlersee-Hügelland.

In general, the aroma composition of actual and predicted wine (exemplarily shown for the 10 'Zweigelt' wines in Figure, b) is more similar to one another than to most of the other wines.

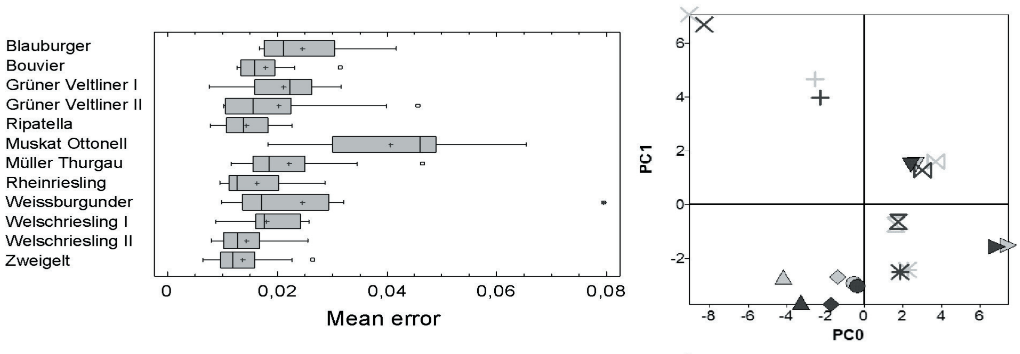
Furthermore it was tried to predict the wine aroma out of the must aroma. Contrary to the previous conditions the ANN didn't know any wine created from the test grape must, the learn set contained only the aromas of the wines produced with the aid of the other musts. Under these circumstances the ANN was not able to forecast aroma profiles better than in the case of random input data for aroma constituents. It may be that the 25 aroma components of the must did not comprise enough information to forecast the 83 constituents of the wines.

Possible future applications for an ANN concerning advances in wine consulting are a better way to make decisions about choosing combinations of vine variety and yeast strain to gain a high quality wine. Further factors that could be recognized by the ANN are terroir influences and - most important - the wine tasters judgements.

References

Cole, V. C.; Noble, A. C.; 1997: Flavour chemistry and assessment. In: A. G. H. Law, J. R. Piggot (Eds): Fermented beverage production, 361-385. Acad. Prof., London, UK.

Fleet, G. H.; 2003: Yeast interactions and wine flavour, Int. J. Food Microbiol. **86**, 11-22.



a b

Figure: (**a**) Average forecasting error for the wines, sorted after grape must; (**b**) Comparison of predicted (dark) and actual (lightgrey) 'Zweigelt' wines using adaptive PCA. Corresponding wines (same symbol) are often "paired", indicating a high aroma profile similarity of actual and predicted wines.

Lambrechts, M. G.; Pretorius, I. S.; 2000: Yeast and its importance to wine aroma, S. Afr. J. Enol. Vitic. **21**, 97-129.

Marchiori, D.; Warglien, M.; 2008: Predicting human interactive learning by regret driven neural networks, Science **319**, 1111-1113.

Marinai, S.; Gori, M.; Soda, G.; 2005: Artificial neural networks for document analysis and recognition. IEEE Transact. Pattern Anal. Machine Intelligence **27**, 23-35.

Pawliszyn, J.; 2000: Theory of solid-phase microextraction. J. Chr. Sci. **38**, 270-278

Pino, R.; Parreno, J.; Gomez, A.; Priore, P.; 2008: Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. Engin. Applic. Artificial Intelligence **21**, 53-62.

Rumelhart, D. E.; Hinton, G. E.; Williams, R. J.; 1986: Learning representations by back-propagating errors. Nature **323**, 533-536.

Stephan, C.; Cammann, H.; Jung, K.; 2005: Artificial neural networks: Has the time come for their use in prostate cancer patients? Nat. Clinic. Pract. Urol. **2**, 262-263.

Received June 16, 2008